MICROBIAL BIOTECHNOLOGY

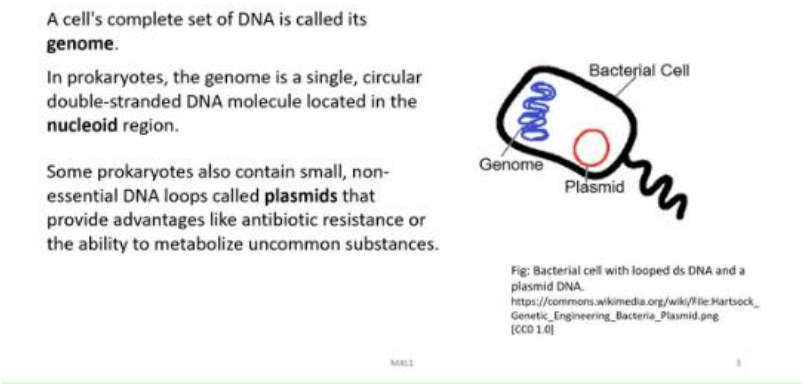
Prof. Utpal Bora
Department of Biosciences and Bioengineering
Indian Institute of Technology Guwahati

Lecture-13

Lec 13: Genome structure, transcription, genetic code, translation

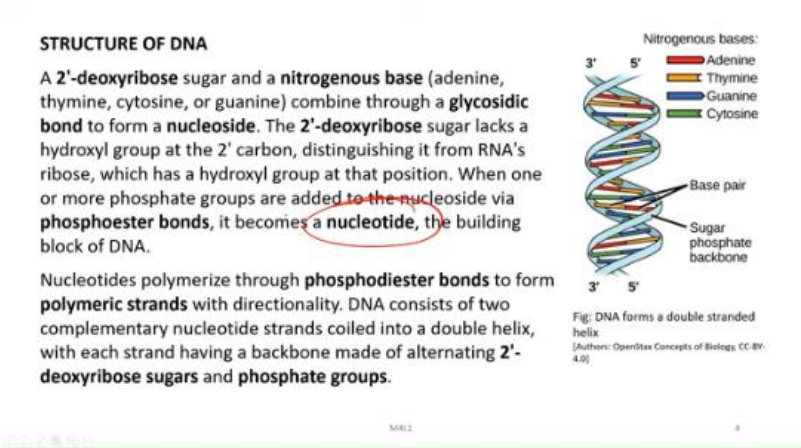
Hello friends, welcome to my course on microbial biotechnology. Today we will be discussing module 4 where the first lecture is on genome structure, transcription, genetic code and translation. We will also of course start with the discussion on the structure of DNA and RNA and also replication which is an important part of this lecture. So let us first once again revisit the concept of genome. You already know that a cell's complete set of DNA is called as its genome.

In prokaryotes, the genome is a single, circular, double-stranded DNA and this is located in a region which is known as the nucleoid because prokaryotes do not have a well-defined nucleus. Some prokaryotes also contain a small non-essential DNA loop called as plasmid but it is not exactly non-essential it gives the bacteria a lot of advantage in coping to environmental stress like antibiotic resistance etc. which we have discussed in some of the earlier lectures and this also helps in metabolizing certain uncommon or toxic substances. Now let us start looking into the structure of DNA.



So some of the simple molecules like 2-deoxyribose sugar, they combine with nitrogenous bases like adenine, thiamine, cytosine and guanine through a glycosidic bond to form a

compound which we call as a nucleoside. The 2-prime deoxyribose sugar that takes part in this reaction lacks a hydroxyl group at the two prime carbon and thereby distinguishing it from RNA's ribose sugar which is a hydroxyl group at that position. When one or more phosphate groups are added to these nucleoside via phosphoester bonds, it becomes what we call a nucleotide. This nucleotide are the building blocks of DNA.



So if we want to know the structure of DNA, it is essential to understand what a nucleoside is and what a nucleotide is. These nucleotides further polymerize through phosphodiester bonds to form polymeric strands, which have directionality, for example, 3' to 5' or 5' to 3'. DNA consists of two complementary nucleotide strands, coiled into a double helix, with each strand having a backbone made of alternating 2' deoxyribose sugars and phosphate groups. This is the simplest way to understand the structure of a DNA molecule, with simple molecules like deoxyribose sugar joining nitrogenous bases to form nucleosides, which then add phosphate groups to form nucleotides. These nucleotides polymerize to form single strands with complementarity. In the DNA double structure, they exist in a double helix.

These nucleotides are linked in a polynucleotide chain, as mentioned, by phosphodiester bonds. These bonds form between the 3' hydroxyl group of one nucleotide's sugar and the phosphate group of another nucleotide's 5' hydroxyl group. This linkage creates the sugar-phosphate backbone, a defining feature of the DNA double helix. You can see here the sugar residues and the sugar-phosphate backbone with the phosphate groups here. Then you see thymine, adenine, guanine, and cytosine here.

Nucleotides are linked in a polynucleotide chain by phosphodiester bonds, which form between the 3'-hydroxyl group of one nucleotide's sugar and the phosphate group of another's 5'-hydroxyl.

This linkage creates the sugar-phosphate backbone, a defining feature of the DNA double helix.

One end of the chain has a free 5' phosphate or hydroxyl group, while the other end has a free 3' phosphate or hydroxyl group, referred to as the 5' and 3' ends, respectively.

This directionality is crucial for DNA replication and transcription.

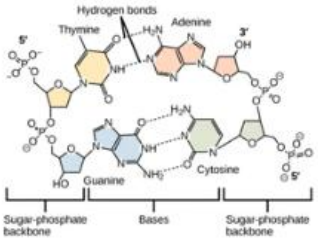


Fig: Adenine pairs with thymine and cytosine pairs with guanine.
(Authors: OpenStax Concepts of Biology, CC-BY-4.0)

Now, let us look at one end of this chain. Here, you can see a 5' end, and alternatively, in the other strand, a 3' end. This is one polymer. This is another polymer of the polynucleotide chains. One end has a free 5' phosphate or hydroxyl group, while the other end has a free 3' phosphate or hydroxyl group. These are referred to as the 3' and 5' ends.

This directionality is very, very crucial for DNA replication and transcription. So these DNA bases, these are flat and then heterocyclic structures made of carbon and nitrogen and they are divided into two categories. The first one are the purines where you have the two members adenine and guanine. They have two rings as you can see over here two rings each then you have the purines pyrimidines cytosine and thymine which have only one single ring each and now adenine will pair up with thymine and guanine will pair up with cytosine and these pairing is stabilized by hydrogen bonds as you can see here by these dotted lines over here and from the diagram itself you can know that thymine and adenine forms two hydrogen bonds while guanine and cytosine will be forming three hydrogen bonds.

DNA bases are flat, heterocyclic structures made of carbon and nitrogen.

They are divided into two categories:

purines (adenine and guanine) with two rings, and

pyrimidines (cytosine and thymine) with one ring.

In the double helix, bases pair via hydrogen bonds:

adenine pairs with thymine (A-T), and

guanine pairs with cytosine (G-C).

The two DNA strands are antiparallel, running in opposite directions (one 5' to 3', the other 3' to 5'), with base pairing occurring along the strands' entire length.

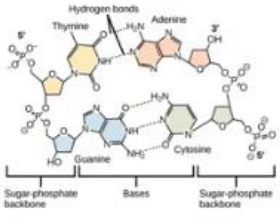
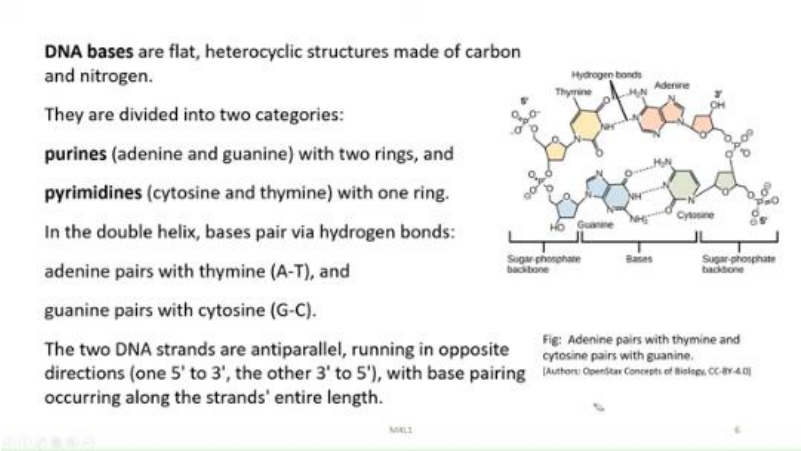


Fig: Adenine pairs with thymine and cytosine pairs with guanine.
(Authors: OpenStax Concepts of Biology, CC-BY-4.0)

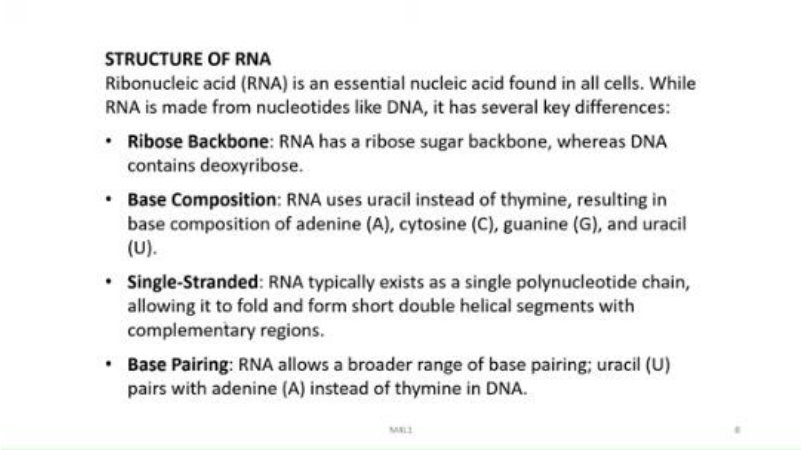
So, in a double helix, in brief terms, a base pair via hydrogen bonds, adenine pairs with thymine by two hydrogen bonds and guanine pairs with cytosine by three hydrogen bonds. The two DNA strains are antiparallel running in opposite directions, one 5' to 3' and the other 3' to 5'. With base pairing occurring along the strength and tail length and these type of ATGC pairing is what we call as complementary base pairing. And so, the two opposite strands are complementary to one another, only then a DNA double helix can be stable.



Most prokaryotic cells contain a single circular chromosome located in the cytoplasm's nucleate region, is already known to you. And E. coli genome has about 4.6 million base pairs. This DNA undergoes supercoiling. Which can be under wound or over wound. Proteins and enzymes such as DNA gyres help maintain and regulate the supercoil structure which is crucial for DNA replication, transcription and fitting the DNA within the small cell.

Now, let us look into the structure of RNA or ribonucleic acid, which is an essential nucleic acid found in all cells. While RNA is made from nucleotides like DNA through a process called transcription, it has several key differences. RNA has a ribose sugar backbone, whereas DNA contains deoxyribose. Then, base composition also has certain differences. RNA uses uracil instead of thymine, resulting in a base composition of adenine, cytosine, guanine, and uracil.

RNA typically exists in a single-stranded form. This allows it to fold and form short double-helical structures or segments with complementary regions. RNA allows a broader range of base pairing. Uracil pairs with adenine instead of thymine, as in the case of DNA. Now, let us look into one of the most important fundamental processes, replication, which is basically the doubling of a DNA molecule.

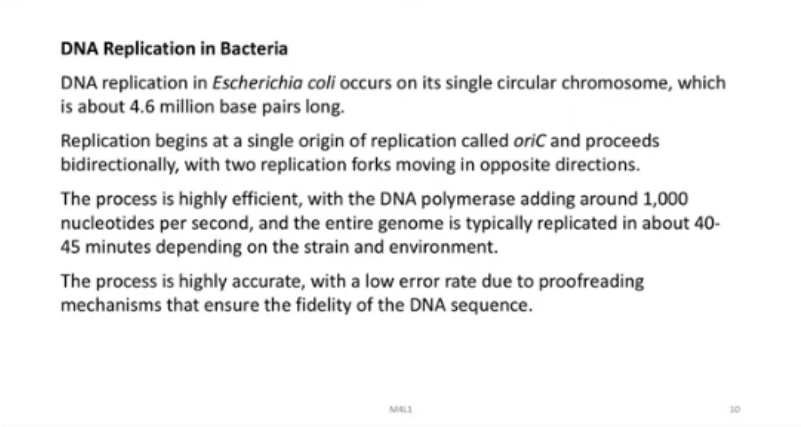


STRUCTURE OF RNA

Ribonucleic acid (RNA) is an essential nucleic acid found in all cells. While RNA is made from nucleotides like DNA, it has several key differences:

- **Ribose Backbone**: RNA has a ribose sugar backbone, whereas DNA contains deoxyribose.
- **Base Composition**: RNA uses uracil instead of thymine, resulting in base composition of adenine (A), cytosine (C), guanine (G), and uracil (U).
- **Single-Stranded**: RNA typically exists as a single polynucleotide chain, allowing it to fold and form short double helical segments with complementary regions.
- **Base Pairing**: RNA allows a broader range of base pairing; uracil (U) pairs with adenine (A) instead of thymine in DNA.

So, DNA replication in bacterial E. coli occurs on its single circular chromosome, which is about 4.6 million base pairs long, as already mentioned. The replication begins at a single origin of replication called OriC and proceeds bidirectionally with two replication forks moving in opposite directions. The process is highly efficient, with DNA polymerizing around 1000 nucleotides per second, which is very fast, and the entire genome is typically replicated in about 40 to 45 minutes. This depends on the strain and the environment in which it exists. The process is highly accurate, with a low error rate due to proofreading mechanisms that ensure the fidelity of the DNA sequences.



DNA Replication in Bacteria

DNA replication in *Escherichia coli* occurs on its single circular chromosome, which is about 4.6 million base pairs long.

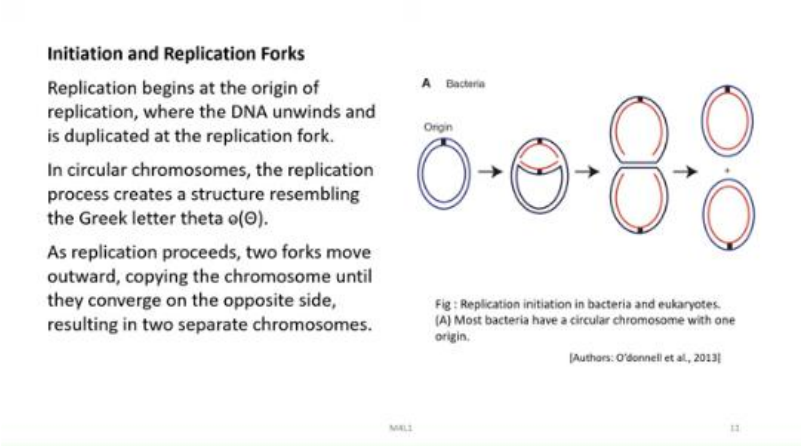
Replication begins at a single origin of replication called *oriC* and proceeds bidirectionally, with two replication forks moving in opposite directions.

The process is highly efficient, with the DNA polymerase adding around 1,000 nucleotides per second, and the entire genome is typically replicated in about 40-45 minutes depending on the strain and environment.

The process is highly accurate, with a low error rate due to proofreading mechanisms that ensure the fidelity of the DNA sequence.

So you can see here one DNA molecule of a bacterium and it has a origin, the point in the DNA molecule, double-stranded molecule from where the replication starts and you can see the replication is bidirectional, it is proceeding in both the directions starting from the origin. So replication begins at the origin of replication. Where the DNA unweans and is duplicated at the replication fork, in cyclic chromosomes, the replication process creates a structure resembling the Greek letter theta. So, this structure, as you can see, looks like the Greek letter theta. That is why it is also known as theta replication.

As replication will proceed, the two forks move outward copying the chromosome until they converge on the opposite sides resulting in two separate DNA molecules or chromosomes. What are the components of the replication machinery which helps in converting a single chromosome or DNA molecule into two DNA molecules or two chromosomes. So some of the key components are proteins like helicases.



They unwind DNA ahead of the So, here is the replication fork as you can see over here and the helicase is unwinding the DNA ahead of this replication fork. So, since this is trying to open up DNA, it is energy consuming process that energy is supplied by ATP. Then there are certain proteins known as single-stranded DNA binding proteins or SSBs. So, since these are the bases as depicted in this diagram, once the helicase opens these DNA double-strand

So, because of complementarity, they will again form a double-stranded structure to prevent the single-strand binding proteins from stabilizing the single-stranded DNA and keep these strands open. So, basically, they prevent separated strands from reannealing. Then we have certain enzymes known as topoisomerases. These relieve tension from unwinding because of supercoiling. The DNA molecule is very difficult to open, and since we are opening it up, tension is building up.

Replication Machinery
Key proteins in DNA replication include:

- **Helicases**: Unwind DNA ahead of the replication fork using ATP.
- **Single-Stranded DNA Binding Proteins (SSBs)**: Prevent separated strands from re-annealing.
- **Topoisomerases**: Relieve tension from unwinding DNA; DNA gyrase is a key enzyme in E. coli.
- **DNA Polymerase**: Adds nucleotides in the 5' to 3' direction, requiring a 3'-OH group, and uses nucleoside triphosphates (dATP, dGTP, dTTP, dCTP) as building blocks and energy sources.

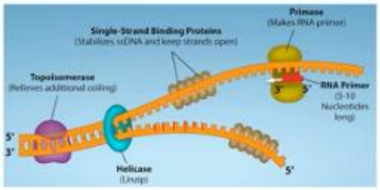


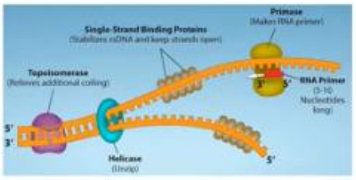
Fig: Components of DNA Replication. [Authors: OpenStax Biology 2e, CC-BY-4.0]

E. coli DNA polymerase III is the primary enzyme responsible for DNA synthesis, while DNA polymerase I plays a key role in removing RNA primers and filling in gaps.

So, to relieve the tension, the topoisomerase will unwind the DNA molecule. DNA gyrase is a key enzyme in E. coli in this case. Then we have other molecules like DNA polymerases, which do the polymerization work. So here, they add nucleotides in the 5' to 3' direction, which requires a 3' OH group and uses nucleoside triphosphates. Like dATP, dGTP, dTTP, and dCTP as building blocks and the energy source.

In E. coli, DNA polymerase 3 is the primary enzyme responsible for DNA synthesis, while DNA polymerase 1 plays a key role in removing RNA primers and filling in gaps. So, these are some of the important components of the DNA replication machinery. Other important ones are the primers, and we also need an RNA primer, which is around 5 to 10 nucleotides long. A DNA polymerase cannot act on a DNA template on its own.

It requires a RNA primer from where it will start the addition of the nucleotides one by one. So these RNA primer is synthesized by this enzyme called primase. So, primase synthesizes a salt RNA primer, which is about 10 nucleotides, which is complementary to the DNA portion to which it binds, enamel DNA polymerase III to begin synthesis. Due to DNA's anti-parallel nature, replication differs for the two strands. The leading strand is synthesized continuously in the 5' to 3' direction towards the replication fork, allowing smooth uninterrupted synthesis.



Primer and Strand Synthesis

Primase synthesizes a short RNA primer (about 10 nucleotides) complementary to the DNA, enabling DNA polymerase III to begin synthesis.

Fig: DNA Replication. [Authors: OpenStax Biology 2e, CC-BY-4.0]

Due to DNA's antiparallel nature, replication differs for the two strands:

The **leading strand** is synthesized continuously in the **5' to 3' direction**, toward the replication fork, allowing smooth, uninterrupted synthesis.

The **lagging strand** is synthesized discontinuously in short segments called **Okazaki fragments**, each initiated by an RNA primer. These fragments are made in the **5' to 3' direction** but apparaently away from the replication fork. The fragments are later joined by **DNA ligase** after the RNA primers are removed.

The leading strand is synthesized discontinuously in short segments which we call as Okazaki fragments, each initiated by an RNA primer. These fragments are made in the 5' to 3' direction but apparently away from the replication fork. The fragments are later joined by DNA ligase which is therefore another important component of this replication machinery after the RNA primers are removed. So, we can see that the DNA replication is taking place in two opposite directions.

In the leading strand, it is happening towards the fork, and in the lagging strand, where small, small fragments of DNA are replicated, and later on, joint by DNA ligase, the replication is in the direction which is away from this replication fork. At the replication fork, DNA's antiparallel structure presents a challenge as the two template strands run in opposite directions. The leading strand is synthesized continuously toward the fork complementing the 3' to 5' template. The lagging strand synthesized in short okazekic fragments requires a separate primer for each fragment

Replication Fork Challenge

At the replication fork, DNA's antiparallel structure presents a challenge as the two template strands run in opposite directions. The leading strand is synthesized continuously toward the fork, complementing the 3' to 5' template. The lagging strand, synthesized in short Okazaki fragments, requires separate primers for each fragment and is made away from the fork in the 3' to 5' direction, while the leading strand runs in the 5' to 3' direction.

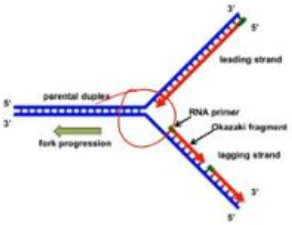
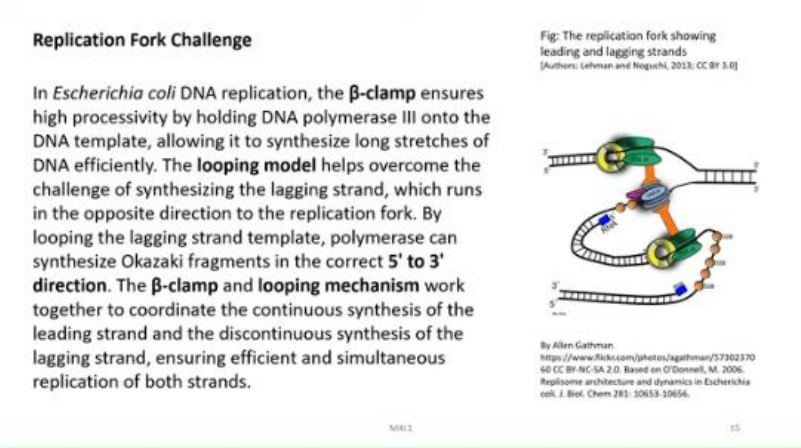


Fig: The replication fork showing leading and lagging strands
[Authors: Lehman and Noguchi, 2013; CC BY 3.0]

and is made away from the fork in the 3' to 5' direction, while the leading strand runs in the 5' to 3' direction. However, the revised model of DNA replication is not as simple as this. There is some kind of adjustment for the steric hindrance problem over here. So, in the lagging strand, there will be a loop formation, and there is a reason for this, which we will try to explain over here. What you need to remember is that the leading strand replicates



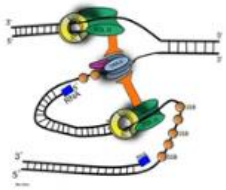
towards the fork center, and in the leading strand, there is some kind of looping and not as straight as this. However, the Okazaki fragments and other things will all be present in this model as well. So, in E. coli DNA replication, there is—I mean, overall, not only in E. coli—a beta clamp which will ensure high processivity by holding the DNA polymerase onto the DNA template, allowing it to synthesize long stretches of DNA efficiently. The looping model helps overcome the challenge of synthesizing the lagging strand, which runs in the opposite direction to the replication fork.

By looping the lagging strand template, polymerase can synthesize Okazaki fragments in the correct 5' to 3' direction. The beta clamp and looping mechanism work together to coordinate the continuous synthesis of the leading strand and the discontinuous synthesis of the lagging strand, ensuring efficient and simultaneous replication of both strands. So, if you want to look into DNA replication very briefly, there is DNA unwinding at the origin of replication.

Replication Fork Challenge

In *Escherichia coli* DNA replication, the **β-clamp** ensures high processivity by holding DNA polymerase III onto the DNA template, allowing it to synthesize long stretches of DNA efficiently. The **looping model** helps overcome the challenge of synthesizing the lagging strand, which runs in the opposite direction to the replication fork. By looping the lagging strand template, polymerase can synthesize Okazaki fragments in the correct **5' to 3' direction**. The **β-clamp** and **looping mechanism** work together to coordinate the continuous synthesis of the leading strand and the discontinuous synthesis of the lagging strand, ensuring efficient and simultaneous replication of both strands.

Fig: The replication fork showing leading and lagging strands
[Authors: Lehman and Noguchi, 2013; CC BY 3.0]



By Allen Gathman
https://www.flickr.com/photos/agathman/57302370
60 CC BY-NC-SA 2.0. Based on O'Donnell, M. 2006.
Replisome architecture and dynamics in Escherichia
coli. J. Biol. Chem 281: 10653-10656.

Then there is helicase action which separates the strands, creating replication forks in both directions. Then the strands are stabilized by single-strand binding proteins to prevent them from reforming a double helix. Then the tension is released by topoisomerases ahead of the fork. Then there is primer synthesis by primase, and this primer is an RNA primer. So, we have to remember that the aim of replication is copying DNA to DNA, but that starts with first making a small RNA primer, not a DNA primer.

We use DNA primers in polymerase chain reactions in silico methods. So, then there is DNA synthesis where the DNA polymerase adds nucleotides to the end of the primers, elongating the leading and lagging strands. Then the primers have to be finally removed from the DNA molecule. Otherwise, you know, there will be errors. So, there are exonucleases which remove these primers.

DNA REPLICATION SUMMARY

DNA Unwinding: DNA unwinds at the origin of replication.

Helicase Action: Helicase enzymes separate the strands, creating replication forks in both directions.

Strand Stabilization: Single-strand binding proteins prevent the strands from re-forming a double helix.

Tension Relief: Topoisomerases relieve supercoiling ahead of the fork.

Primer Synthesis: Primase synthesizes RNA primers complementary to the DNA template.

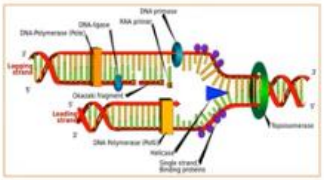


Figure: DNA replication [Source: Public domain (Wikipedia)]

DNA Synthesis: DNA polymerase III adds nucleotides to the 3'-OH end of primers, elongating the leading and lagging strands.

Primer Removal: Exonucleases remove RNA primers.

Gap Filling: DNA polymerase I fills the gaps with deoxynucleotides (dNTPs).

Strand Sealing: DNA ligase seals gaps between DNA fragments by forming phosphodiester bonds.

Then there is gap filling. DNA polymerase 1 fills the gaps with deoxy nucleotides, and then finally there is strand sealing. The DNA ligase seals the gap between DNA fragments by

forming phosphodiester bonds. Now, let us look into one important aspect of the DNA molecule, which is the transcription process where DNA is copied into RNA molecules.

So, we know about the primase copying the DNA into DNA primers or RNA primers. But in this case, when we talk about transcription, it involves an altogether different kind of process. Before understanding that, let us try to understand the genetic code because transcription has a purpose. The transcription of DNA does not stop at that stage; it will further go on to translation, where the RNA molecule, which has copied the information from the DNA, will be further used for processing and making proteins.

So, further understanding the concept of the genetic code is very, very important. And in this aspect, let us summarize the central dogma, which is very, very important for the genetic code. So, as I have briefly mentioned, DNA gets copied to RNA, and the RNA finally converts to proteins. The central dogma describes this flow of genetic information from DNA to RNA to proteins. Genes encode mRNA sequences, which are translated into proteins, converting the four-nucleotide genetic code into triplet codons that specify around 20 amino acids. This describes the different genetic codes corresponding to amino acids.

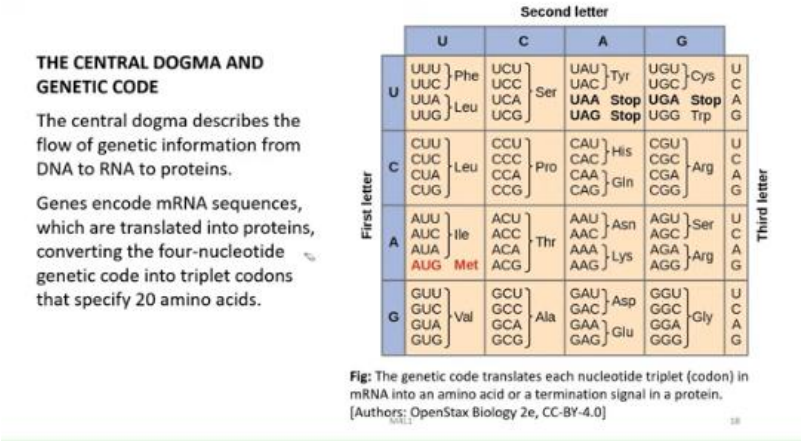


Fig: The genetic code translates each nucleotide triplet (codon) in mRNA into an amino acid or a termination signal in a protein. [Authors: OpenStax Biology 2e, CC-BY-4.0]

For example, if we have three U's—first place U, second place U, third place U—this will code for phenylalanine. But if, for example, the third place changes to A or G in this case, it will represent leucine. But then, in certain other cases, all the types of combinations available in the third place—U, A, or G—may represent a single type: leucine. So, leucine again has two sets of codes, with the first and second having both U and U.

And in the second case, the first place has all C. So, this table has been worked out for each and every amino acid, which are 20 amino acids. Now, there are certain codons; we call these triplets codons. There are certain codons which do not code for amino acids; rather,

they give certain signals like stop signals. That is, once these signals are there, the transcription will stop, or the process will stop there. So, now mRNA translation will not proceed further once these stop codons are encountered.

So, we will try to understand this process in more detail. So, some of the things that are important to know at this point are the degeneracy and termination, about which we have spoken briefly. Of these 64, 3 represent the stop codons. So, 64 minus 3, 61 specify amino acids. So, these 61 correspond to 21 amino

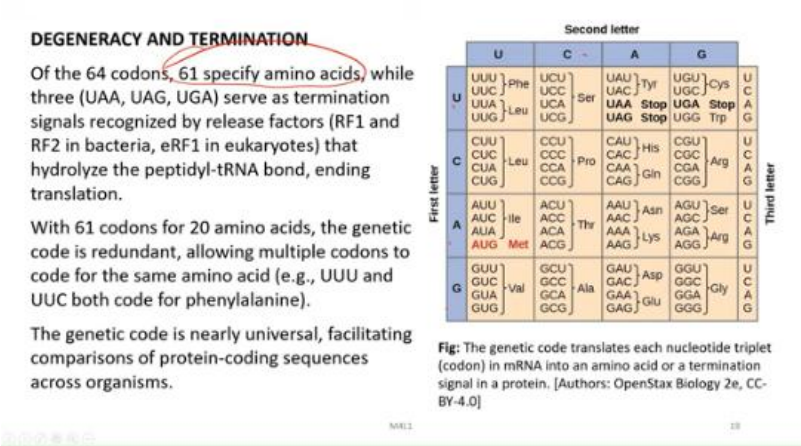
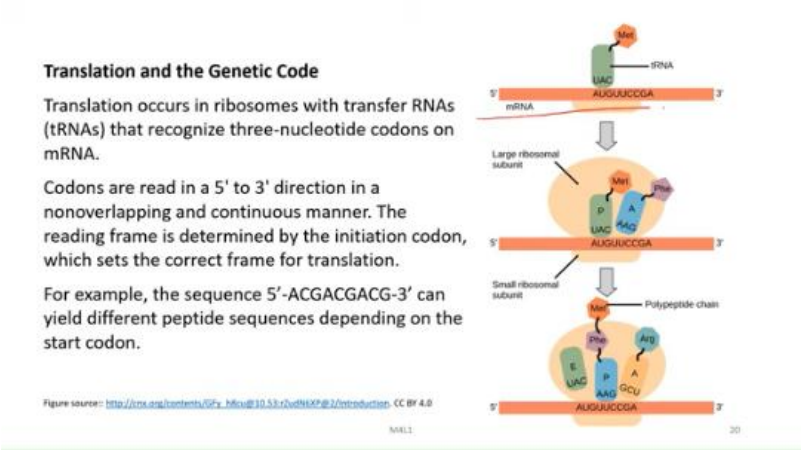


Fig: The genetic code translates each nucleotide triplet (codon) in mRNA into an amino acid or a termination signal in a protein. [Authors: OpenStax Biology 2e, CC-BY-4.0]

amino acids. So, many amino acids have more than 1 or 2 codons. So, this is what we call genetic redundancy. So, this allows multiple codons to code for the same amino acid, which we have described earlier. And then, these three stop codons serve as termination signals, recognized by release factors RF1 and RF2 in bacteria and ERF1 in eukaryotes, that hydrolyze the peptidyl-tRNA bond, ending the process of translation.

The genetic code is nearly universal, facilitating comparison of protein-coding sequences across organisms, but there are some variations between prokaryotes and eukaryotes in terms of amino acids. The translation in the genetic code. So, let us look into this diagram first. This is an mRNA molecule with directionality from 5' to 3'. This mRNA is produced from



DNA by the process of transcription once we have this mRNA in hand. So, you have U here in place of T. Now, these triplets represent one codon, and depending on the reading frame, we may have many codons from a small stretch of mRNA molecule. So, translation will occur in ribosomes with transfer RNAs. This is a tRNA that recognizes the three-nucleotide codons on mRNA. Codons are read in the 5' to 3' direction in a non-overlapping and continuous manner.

The reading frame is determined by the initiation codon, which sets the correct frame for So, this is a large ribosomal subunit, and this is the small ribosomal subunit. Then we have the tRNAs bringing in the amino acids, and there are these anticodons that will identify the triplet codons on the mRNA by complementarity. You can see U pairs with A, and C pairs with G over here. So, in this sequence 5' ACGA, CGA, CG3', this can yield different peptide sequences depending on the start codon.

Let us now look into the process of transcription. This is the process of RNA synthesis from DNA molecules. The RNA sequence is complementary to the DNA template, with adenine pairing with uracil instead of thymine, as already mentioned. There are three main types of RNA.

Transcription is the process of synthesizing RNA based on DNA instructions.

The RNA sequence is complementary to the DNA template, with adenine pairing with uracil instead of thymine.

There are three main types of RNA:

- **Transfer RNA (tRNA):** Carries amino acids for protein synthesis.
- **Ribosomal RNA (rRNA):** Forms part of ribosomes, essential for assembling proteins.

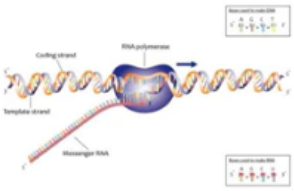


Figure: DNA is copied onto mRNA by RNA polymerase
[Source: Wikipedia; CC BY 2.0]

M4L1 22

Transfer RNA carries amino acids for protein synthesis. Then there is ribosomal RNA, which forms part of ribosomes, essential for assembling proteins. So, these are ribosomal RNA, and this is tRNA, which is a type of RNA. These ribosome structures, whether large or small subunits, also contain rRNA.

So, ribosomal RNA. We studied these in one of the earlier lectures. Then there is messenger RNA. The first two do not translate and produce proteins, but mRNA produces proteins. However, these two are essential for protein production using messenger RNA.

So mRNA carries the genetic code for protein synthesis. In bacteria, genes for related functions are often grouped into operons and transcribed together from a single promoter, producing a polycystionic mRNA with multiple coding regions separated by spaces. Each coding region is translated independently into different polypeptides. RKL mRNAs can also be polycystriotic. So you can see here a coding strand and this is the RNA polymerase which is producing the messenger RNA thereby DNA is copied into mRNA by RNA polymerase.

Messenger RNA (mRNA):

mRNA carries the genetic code for protein synthesis.

In **bacteria**, genes for related functions are often grouped into operons and transcribed together from a single promoter, producing a **polycistronic mRNA** with multiple coding regions separated by spacers.

Each coding region is translated independently into different polypeptides.

Archaeal mRNAs can also be polycistronic.

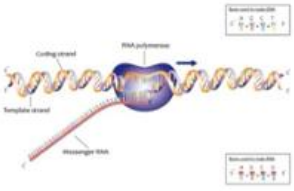


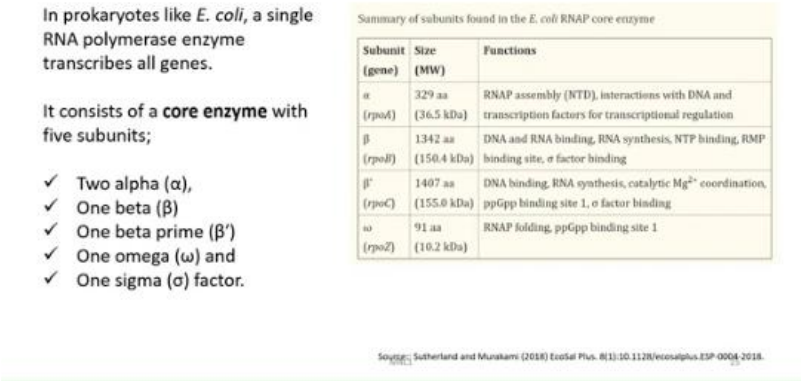
Figure: DNA is copied onto mRNA by RNA polymerase
[Source: Wikipedia; CC BY 2.0]

M4L1 23

So, some of the important features of transcription is it requires a template strand, the DNA strand which is used to create complementary mRNA. Then there is a non-template strand or the coding strand. The DNA strand similar to the mRNA not used as a template because the mRNA is complementary to the template strand and the non-template strand is complementary to the template strand. So the non-template strand and the mRNA will be having a similar sequence with similar basis.

So another important thing is the plus one site which is called as the initiation site, the location where the transcription begins. Then there is upstream or downstream position. The nucleotides before the plus one site are upstream. Negative numbers after that are downstream, which are the positive numbers transcription requires. Unwinding of the DNA to form a transcription bubble where mRNA synthesis will be occurring.

In prokaryotes like E. coli, a single RNA polymerase enzyme transcribes all the genes. It consists of a core enzyme with five subunits: alpha, which has two units, then beta, and beta prime and omega and sigma factor, all in single copies. You can see the molecular size of these different subunits over here. Alpha is around 329 amino acids or 36.5 kilodaltons.

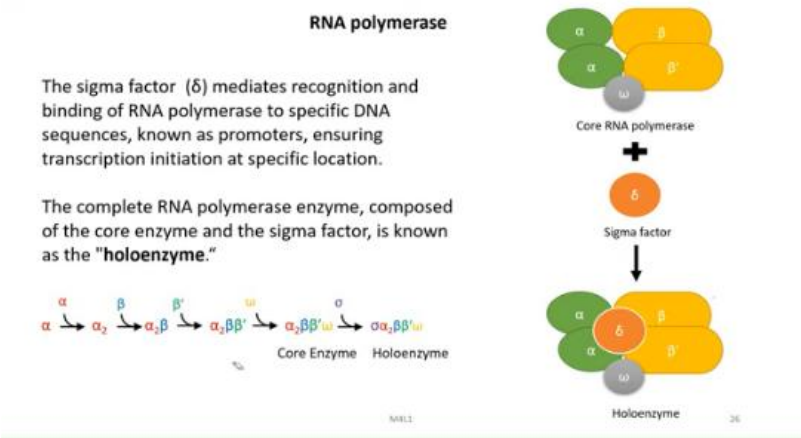


Beta is 150.4 kilodaltons. Beta prime is 155 kilodaltons, and omega is around 10.2 kilodaltons. These are the genes that produce these particular subunits of the enzyme. RPO A produces alpha, RPO B beta, RPO C beta prime, and RPO Z omega. The functions of these subunits are: alpha helps in RNAP assembly, interactions with DNA, and transcription factors for transcriptional regulation.

Beta helps in DNA and RNA binding, RNA synthesis, NTP binding, RMP binding site, and sigma factor binding. Beta prime helps in DNA binding, RNA synthesis, catalytic

magnesium coordination, PPGPP binding sites 1, and sigma factor binding, while omega helps in RNA folding and PPGPP binding site 1. This structure has 2 alpha, 1 beta, 1 beta prime, and omega. This is known as the core RNA polymerase, which, when coupled to the sigma factor, forms what is called the holoenzyme. The sigma factor mediates recognition and binding of RNA polymerase to specific DNA sequences known as promoters.

So this will bind to the promoters and ensure transcription initiation at the correct and specific location. The complete RNA polymerase enzyme, composed of the core enzyme and the sigma factor, is known as the holoenzyme, and it is a kind of sequential assembly. Two alpha subunits join to form alpha 2, to which a beta subunit adds to form alpha 2 beta. Then joined by beta prime and finally omega (alpha 2 beta beta prime omega), and then the sigma factor, which recognizes promoters, will bind to it, forming the holoenzyme. This is the functional enzyme.



Initiation of transcription in prokaryotes: how prokaryotic RNA polymerase starts the process. The sigma factor provides specificity by ensuring RNA polymerase binds to the correct promoter sequence and initiates transcription at the proper site. Without the sigma factor, the core enzyme would not recognize promoter sequences accurately. So, holoenzyme minus the sigma factor (the core enzyme) is unable to bind to promoter sequences, leading to non-specific or inefficient transcription. Once transcription begins, the sigma factor can be released.

The sigma factor is like a person who introduces you to the house you want to visit. After introducing you to the house you want to visit, it will go away. Once transcription begins, the sigma factor can be released, allowing the core enzyme to continue. The core enzyme has a crab-claw or clamp-jaw shape with an internal channel that runs the length of the

enzyme. Eukaryotic and archaeal RNA polymerases have a similar core structure but include additional subunits.

INITIATION OF TRANSCRIPTION IN PROKARYOTES

Prokaryotic RNA Polymerase

The sigma factor provides specificity by ensuring that RNA polymerase binds to the correct promoter sequence, initiating transcription at the proper site.

Without the sigma factor, the core enzyme would not recognize promoter regions accurately, leading to non-specific or inefficient transcription.

Once transcription begins, the sigma factor can detach, allowing the core enzyme to continue. The core enzyme has a "crab claw" or "clamp-jaw" shape with an internal channel that runs the length of the enzyme.

Eukaryotic and archaeal RNA polymerases have a similar core structure but include additional subunits.

M4L1 27

Let us now look into the prokaryotic promoters. What is a promoter? A promoter is a specific DNA sequence where the transcription machinery, including the RNA polymerase, attaches to initiate transcription. Promoters are typically located upstream of the genes they regulate. The sequence per promoter is crucial because it influences the frequency of transcription, determining whether it occurs constantly, intermittently, or rarely.

So here you can see a promoter sequence. It has two regions: the minus 35 region and the minus 10 region. And then this is the plus one transcription start site. And you can see here this RNA polymerase, or the core enzyme, and this blue one is the sigma factor. This sigma factor recognizes this promoter and after that, it may detach after starting the process, and this RNA polymerase will do the transcription as discussed.

Prokaryotic Promoters

A promoter is a specific DNA sequence where the transcription machinery, including RNA polymerase, attaches to initiate transcription.

Promoters are typically located upstream of the genes they regulate.

The sequence of a promoter is crucial because it influences the frequency of transcription, determining whether it occurs constantly, intermittently, or rarely.

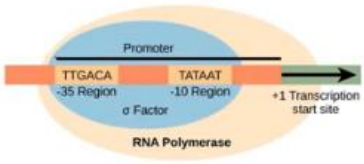


Fig: The σ subunit of prokaryotic RNA polymerase recognizes consensus sequences found in the promoter region upstream of the transcription start sight.
[Authors: OpenStax Biology for AP®, CC-BY-4.0]

M4L1 28

Prokaryotic promoters share conserved elements across species. Notably, two consensus sequences are located at minus 10 and minus 35 positions upstream of the transcription

start site. The minus 10 region is known as the Pribnow box, and the consensus sequence is TATAAT. Then there is the minus 35 region, where the consensus sequence is TTGACA. The sigma factor of RNA polymerase initiates transcription by binding to the promoter. The minus 10 region, which is AT-rich, aids in DNA unwinding.

During initiation, sort aborted transcripts of about 10 nucleotides are produced and released early in the transcription process. So, this is not a single strike process so it may keep on trying several times. It is like kick starting on vehicle in old times where you start the process again and again and ultimately when the knock is optimum the whole process begins. Then once the initiation is over, the next phase is the elongation phase.

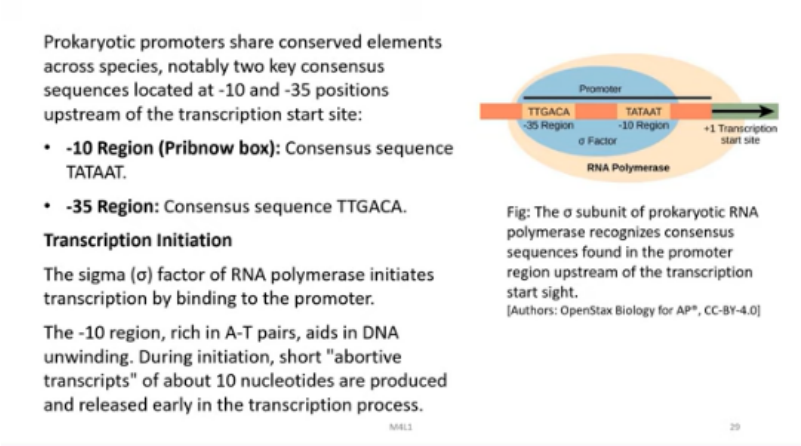


Fig: The σ subunit of prokaryotic RNA polymerase recognizes consensus sequences found in the promoter region upstream of the transcription start sight.
[Authors: OpenStax Biology for AP®, CC-BY-4.0]

So the mRNA synthesis begins after the sigma subunit is released from the RNA polymerase, allowing the core enzyme to synthesize mRNA in the 5' to 3' direction of about 40 nucleotides per second. And then the key steps in these elangosis process are number one, DNA unwinding. The DNA double helix unwinds ahead of the polymerase to provide a template for RNA synthesis. DNA rewinding also happens. The DNA reforms its double helix behind the polymerase.

So then there is stabilization. Ongoing stabilization is required for the base pairing between the DNA and the RNA. And after the elongation process comes the termination in prokaryotic transcription. Termination occurs via two mechanisms. One is known as the Rho-dependent mechanism, another is the Rho-independent mechanism.

Elongation

mRNA synthesis begins after the sigma (σ) subunit is released from RNA polymerase, allowing the core enzyme to synthesize mRNA in the 5' to 3' direction at about 40 nucleotides per second.

Key steps during elongation are:

- **DNA Unwinding**: The DNA double helix unwinds ahead of the polymerase to provide a template for RNA synthesis.
- **DNA Rewinding**: The DNA re-forms its double helix behind the polymerase.
- **Stabilization**: Ongoing stabilization is required for the base pairing between DNA and RNA.

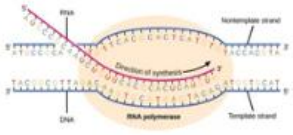


Figure: During elongation, the prokaryotic RNA polymerase tracks along the DNA template, synthesizes mRNA in the 5' to 3' direction, and unwinds and rewinds the DNA as it is read.

[Authors: OpenStax Biology for AP®, CC-BY-4.0]

The Rho protein catches up to RNA polymerase at a Z-rich region, triggering mRNA release. The Rho-independent method, a CZ-rich Sequence causes the mRNA to form a hairpin loop, stabilizing the polymerase. Weak UV interactions lead to the release of the polymerase and mRNA. Once termination occurs, transcription is complete.

Now, let us study about another important process, the translation. Translation is basically the process of protein synthesis from mRNA. So basically here the genetic information copied from DNA into mRNA is finally translated into polypeptides or proteins and these are vital for cell function as proteins are involved in nearly every cellular process including replication, transcription and translation itself, not to speak about other metabolic pathways. So overall, decatalyze reactions and regulate gene expression.

These complex and precise processes are highly energy intensive. A lot of energy is required in the protein synthesis. Translation occurs in ribosomes where tRNA delivers amino acids that are assembled into polypeptide chains based on the mRNA sequence. Ribosomes are crucial for efficiently assembling proteins and so in protein which are essential for cellular functions are synthesized accurately. So, once again let us look into the structure of ribosome which has two subunits in prokaryotes.

Protein Synthesis

Protein synthesis translates genetic information from mRNA into polypeptides, vital for cell function as proteins are involved in nearly every cellular process, such as catalyzing reactions and regulating gene expression.

This complex and precise process is highly energy-intensive.

Translation occurs in ribosomes, where tRNA delivers amino acids that are assembled into polypeptide chains based on the mRNA sequence.

Ribosomes are crucial for efficient protein assembly, ensuring proteins essential for cellular functions are synthesized accurately.

M4L1 33

The 30S small subunit is made up of around 16S ribosomal RNA molecules and 21 proteins. The 50S large subunit contains 23S rRNA, 5S rRNA, and 34 proteins, and its function is to help in protein synthesis. When the 30S and 50S subunits join together, they form the 70S ribosome. This ribosome moves along an mRNA strand to connect amino acids and build proteins in prokaryotic cells.

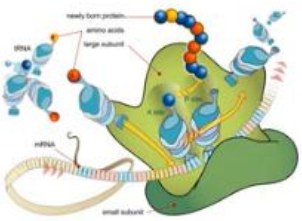
The Ribosome

Structure:

The ribosome consists of two subunits:

1. **30S Small Subunit:** made up of a 16S ribosomal RNA (rRNA) molecule and 21 proteins.
2. **50S Large Subunit:** contains a 23S rRNA, a 5S rRNA, and 34 proteins.

Function: When the 30S and 50S subunits join together, they form a 70S ribosome. This ribosome moves along an mRNA strand to connect amino acids and build proteins in prokaryotic cells.



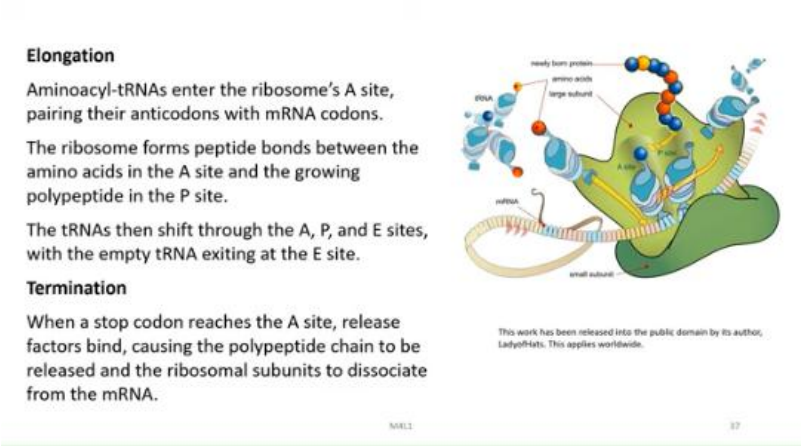
M4L1 34

The ribosome undergoes a cycle of association and dissociation during translation. It binds to mRNA, synthesizes the protein, and then dissociates to be used again. This is what we call the ribosome cycle. Then there are tRNA binding sites. The ribosome has three sites for tRNA binding.

The first one is the A site, then the P site, and the E site. The A site is the aminoacyl tRNA site where the tRNA with the next amino acid binds. The P site is where the tRNA holds the growing polypeptide chain, and then there is the E site, where the tRNA exits the ribosome after delivering its amino acid. Now, let us look into the mechanism of protein synthesis in the ribosome. Protein synthesis in ribosomes involves three main stages.

The initiation stage, the elongation stage, and the termination stage. In initiation, the 30S subunit binds to the mRNA and the Shine-Dalgarno sequence, positioning the start codon in the P-site. So, you can see here the P-site. The initiator tRNA is carrying formyl methionine (fMet) binds to this start codon, and the 50S subunit then joins the complex, forming the 70S ribosome.

Now, next is elongation, where aminoacyl tRNA enters the ribosome's A site. You can see here the A site, pairing its anticodon with mRNA codons. The ribosome forms peptide bonds between the amino acids in the A site and the growing polypeptide in the P-site.

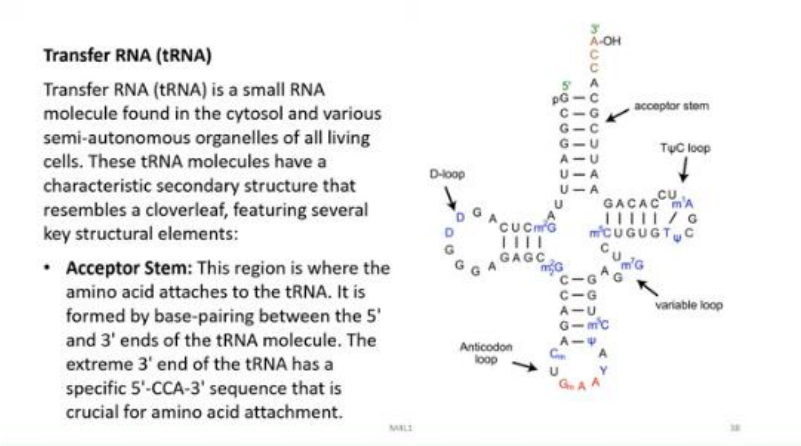


The tRNAs then shift through the A, P, and E sites, with empty tRNA exiting at the E site. Then comes the termination step, when a stop codon reaches the A site here. Release factors will bind, causing the polypeptide chain to be released and the ribosomal subunits to dissociate from the mRNA. Now, in this entire process, tRNA plays a crucial role apart from the ribosomal subunits. So, let us look into the structure of the tRNA.

tRNA is transfer RNA, which is a small RNA molecule found in the cytosol and various semi-autonomous organelles of all living cells. These tRNA molecules have a characteristic secondary leaf structure that resembles a clover leaf, featuring several key structural elements like an acceptor stem, a TsiC loop, a variable loop, and an anticodon loop. D-loop. The acceptor stem is the region where the amino acid attaches to the tRNA. It is formed by base pairing between the 5' and 3' ends of the tRNA molecule.

This is actually a single molecule folded on its own, with the 5' and 3' ends forming the acceptor stem where the amino acid will be loaded. The extreme end of the 3' terminus of tRNA has a specific 5' CCA 3' sequence. This sequence is crucial for amino acid

attachment. Then there is a pseudo-uridine loop, named for its inclusion of pseudo-uridine, an unusual base. This loop is a distinct feature of tRNA.



Then there is the D-loop. It contains dihydrouridine, which is responsible for its name. This loop contributes to the overall tRNA structure. Then there is the anticodon loop. This loop contains the anticodon.

A 3-nucleotide sequence that pairs with the complementary mRNA codon during translation. The anticodon is flanked by a purine base on its 3-prime end and a uracil on its 5-prime end. Then there is this variable loop, which is positioned between the anticodon loop and the CU loop. The variable loop. The loop can range in size from 3 to 21 bases and varies among different tRNA molecules.

After folding via intermolecular base pairing, a tRNA molecule has one end that contains the anticodon, which interacts with the mRNA codon, and the CCA amino acid-binding end. Attachment of amino acids to tRNA. How these amino acids get loaded into a tRNA molecule. What are the states of tRNA? Charged tRNA and uncharged tRNA.

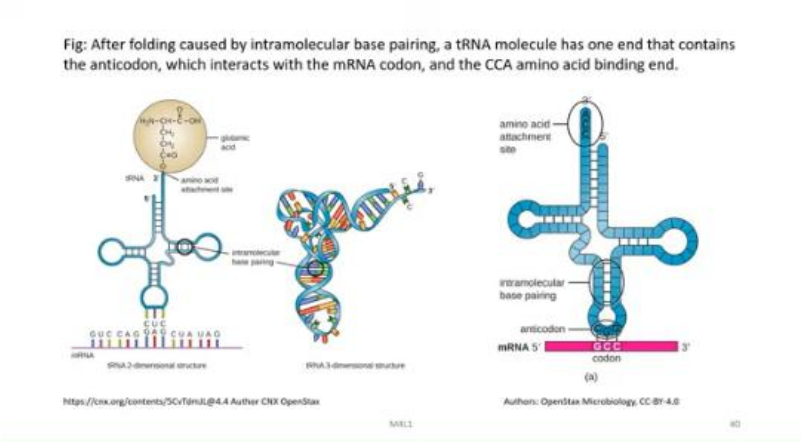
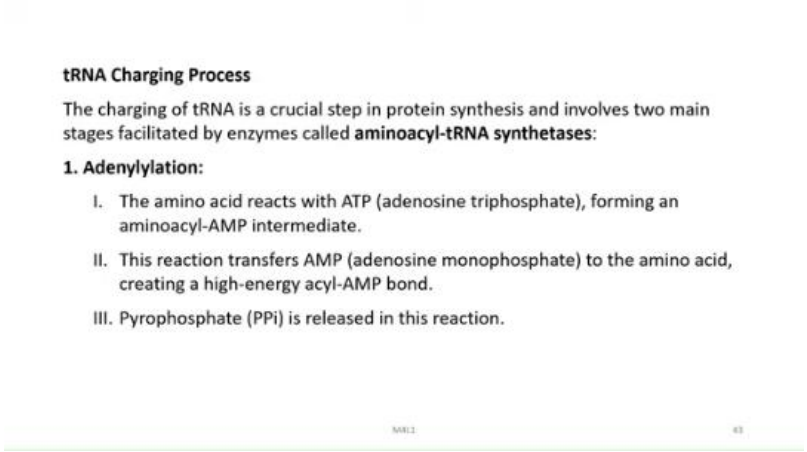


Fig: After folding caused by intramolecular base pairing, a tRNA molecule has one end that contains the anticodon, which interacts with the mRNA codon, and the CCA amino acid binding end.

The charged tRNA has a covalently attached amino acid, which is essential for carrying the amino acid to the ribosome during protein synthesis. The process of attaching an amino acid to the tRNA is known as charging. Then there is uncharged tRNA. This tRNA lacks an amino acid. It is considered empty until it undergoes the charging process.

Aminoacyl-tRNA synthetases are enzymes. They come in two classes based on the attachment sites. Class 1 is the enzyme that attaches amino acids to the 2' OH group of tRNA. They are typically single monomeric proteins. Then there is the class 2 enzyme, which attaches amino acids to the 3' OH group of tRNA.

They are usually composed of two or four subunits, dimeric or tetrameric. Both classes are crucial for ensuring the accurate pairing of amino acids with their corresponding tRNA molecules, which is vital for correct protein synthesis during translation. tRNA charging process. The charging of tRNA is a crucial step in protein synthesis and involves two main stages facilitated by enzymes called aminoacyl-tRNA synthetases. Number one is adenylation.



tRNA Charging Process

The charging of tRNA is a crucial step in protein synthesis and involves two main stages facilitated by enzymes called **aminoacyl-tRNA synthetases**:

1. Adenylylation:

I. The amino acid reacts with ATP (adenosine triphosphate), forming an aminoacyl-AMP intermediate.

II. This reaction transfers AMP (adenosine monophosphate) to the amino acid, creating a high-energy acyl-AMP bond.

III. Pyrophosphate (PPi) is released in this reaction.

The amino acid reacts with ATP, forming an aminoacyl-AMP intermediate. This reaction transfers adenosine monophosphate to the amino acid, creating a high-energy acyl-AMP bond. Pyrophosphate is released in this reaction. Then tRNA charging. The aminoacyl-AMP intermediate binds to the aminoacyl-tRNA synthetase enzyme.

The amino acid is then transferred from the enzyme to the 3' end of the tRNA molecule, specifically to the 2' or 3' hydroxyl group of the tRNA's acceptor stem. AMP is released during this transfer. The outcome of these steps is a charged tRNA with a specific amino acid attached, ready for participation in protein synthesis. So, this is, in brief, the

diagrammatic representation of the loading of a tRNA molecule by the process that we have just described. So, how does translation initiation happen?

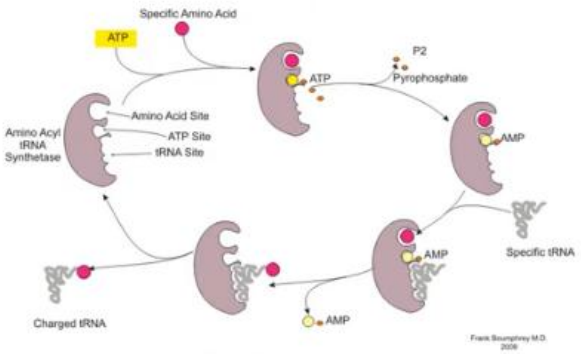


Figure: tRNA charging with an amino acid
[Source: Wikipedia; CC BY-SA 3.0]

Protein synthesis begins with the formation of an initiation complex. In E. coli, this complex involves several key components: the 30S ribosomal subunit, the mRNA template (the messenger RNA that carries the genetic code), initiation factors called IF-1, IF-2, and IF-3, which are involved in this process, and the initiator tRNA, a specific initiator tRNA called tRNA-fMet, which carries formyl methionine. It is usually removed after translation in subsequent elongation phases.

A regular methionine is inserted by tRNA. This is tRNA-Met, and this is tRNA-fMet. So this needs to be remembered, as it is a very important point. So you can see here the tRNA with the first amino acid, and then there is the anticodon. There is the start codon, then you have the large subunit and the small subunit, and this is the overall mRNA, with a stop codon where the process will stop.

Translation Initiation

Protein synthesis begins with the formation of an initiation complex. In *E. coli*, this complex involves several key components:

- **30S Ribosomal Subunit:** The small ribosomal subunit.
- **mRNA Template:** The messenger RNA that carries the genetic code.
- **Initiation Factors:** Three factors—IF-1, IF-2, and IF-3—are involved in the process.
- **Initiator tRNA:** A specific initiator tRNA, called $tRNA^{fMet}$, carries formylmethionine (fMet). It is usually removed after translation. In subsequent elongation phases, a regular methionine is inserted by $tRNA^{Met}$.

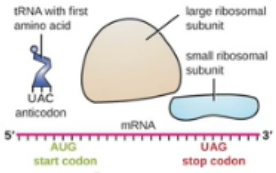
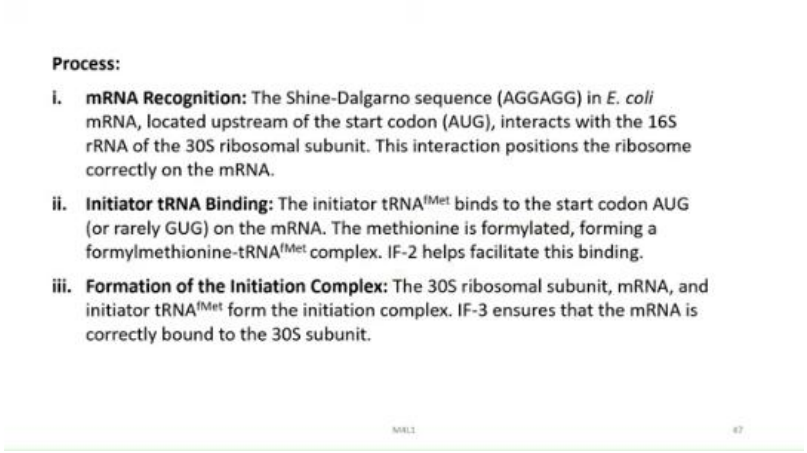


Figure: Formation of TIC (translation initiation complex)

Let us look into the process with a little bit more detail. It starts with mRNA recognition of the Shine-Dalgarno sequence in E. coli mRNA. Located upstream of the start codon, it interacts with the 16S rRNA of the 30S ribosomal subunit. This interaction positions the ribosome correctly on the mRNA. Then there is the initiator tRNA binding.

The initiator tRNA fMet binds to the start codon AUG on the mRNA and sometimes to GUG, though rarely. The methionine is formylated, forming a formyl methionine complex. IF2 helps facilitate this binding. Then there is the formation of the initiation complex. The 30S ribosomal subunit, mRNA, and initiator tRNA fMet form the initiation complex, and IF3 ensures that the mRNA is correctly bound to the 30S subunit.



Process:

i. **mRNA Recognition:** The Shine-Dalgarno sequence (AGGAGG) in *E. coli* mRNA, located upstream of the start codon (AUG), interacts with the 16S rRNA of the 30S ribosomal subunit. This interaction positions the ribosome correctly on the mRNA.

ii. **Initiator tRNA Binding:** The initiator $tRNA^{fMet}$ binds to the start codon AUG (or rarely GUG) on the mRNA. The methionine is formylated, forming a formylmethionine-$tRNA^{fMet}$ complex. IF-2 helps facilitate this binding.

iii. **Formation of the Initiation Complex:** The 30S ribosomal subunit, mRNA, and initiator $tRNA^{fMet}$ form the initiation complex. IF-3 ensures that the mRNA is correctly bound to the 30S subunit.

Then there is the joining of the subunits: the 50S ribosomal subunit associates with the 30S subunit, forming the complete 70S ribosome ready for the elongation phase of translation. In this process, the energy source comes from guanosine triphosphate for the various steps in the initiation process, including the binding of the initiator tRNA and the joining of the ribosomal subunits. Then comes translation elongation. Once the translation complex is established, the ribosome has three tRNA binding sites: A, P, and E (aminoacyl, peptidyl, and exit, respectively). In initiation, the initiating methionine tRNA occupies the P site, while the A and E sites are available for incoming tRNAs. Then there is codon recognition, and the ribosome moves along the mRNA.

Each mRNA codon aligns with the appropriate charged tRNA anticodon. This precise pairing is crucial for accurate protein synthesis, as the ribosome would otherwise bind tRNAs randomly without mRNA. Then, there is amino acid addition. Charged tRNAs enter the A site, where their amino acids are added to the growing polypeptide chain. The amino acid at the A-site tRNA forms a peptide bond with the amino acid at the P-site tRNA.

Amino Acid Addition: Charged tRNAs enter the A site, where their amino acids are added to the growing polypeptide chain.

The amino acid at the A-site tRNA forms a peptide bond with the amino acid at the P-site tRNA.

This bond formation is catalyzed by peptidyl transferase, an RNA-based enzyme in the 50S ribosomal subunit.

The energy for peptide bond formation comes from the high-energy bond between each amino acid and its tRNA.

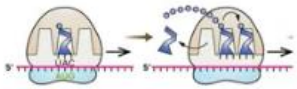


Figure: tRNAs bring amino acids which are added to the growing chain

This bond formation is catalyzed by peptidyl transferase, an RNA-based enzyme in the 50S ribosomal subunit. The energy for peptide bond formation comes from the high-energy bond between each amino acid and its tRNA. Then, the next step is translocation. After peptide bond formation, the ribosome undergoes a conformational change, moving three bases along the mRNA. This shifts the tRNA from the A site to the P site, and the tRNA in the P site moves to the E site, where it exits the ribosome.

This movement is facilitated by elongation factors that hydrolyze GTP, providing the necessary energy for each step. In E. coli, the translation machinery can add an amino acid every 0.05 seconds, allowing for the synthesis of a 200-amino-acid protein in about 10 seconds. Finally, termination occurs when a stop codon is encountered in the A site of the ribosome. Here is how this process occurs. Recognition is the first step.

When a stop codon is positioned in the A site, it is recognized by protein release factors, which are analogous to tRNAs. But instead, they function to terminate translation, and then there is release factor action. These release factors promote the addition of a water molecule to the carbonyl end of the polypeptide chain, attached to the tRNA in the P site. This reaction, catalyzed by peptidyl transferase, hydrolyzes the bond between the polypeptide and the tRNA, releasing the newly synthesized protein. Then there is disassembly.

Translation Termination

Translation terminates when a stop codon (UAA, UAG, or UGA) is encountered in the A site of the ribosome. Here's how this process occurs:

i. **Recognition of Stop Codon:** When a stop codon is positioned in the A site, it is recognized by protein release factors (RFs), which are analogous to tRNAs but instead function to terminate translation.

ii. **Release Factor Action:** These release factors promote the addition of a water molecule to the carboxyl end of the polypeptide chain attached to the tRNA in the P site.

This reaction, catalyzed by peptidyl transferase, hydrolyzes the bond between the polypeptide and the tRNA, releasing the newly synthesized protein.

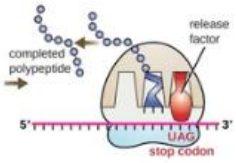


Figure: Stop codons recognized by release factors leading to the dissociation of the translational complex and release of polypeptide chain

After the polypeptide is released, the ribosomal subunits—both small and large—detach from the mRNA and each other. And then there is recycling. The ribosomal subunits are then available to be assembled into new translation initiation complexes. Meanwhile, the mRNA is degraded, and its nucleotides are recycled for future transcription and translation processes. This process ensures the efficient release of the newly synthesized protein and the reusability of the ribosomal machinery and mRNA components.

So overall, we can see here the schematics of this process, where there is initiation, You have this tRNA with the first amino acid, which we already described, along with the large and small subunits and the mRNA. There is a start codon and a stop codon, and then initiation happens, followed by elongation and finally termination. So with this, we come to the end of this lecture. Thank you for your patient hearing.

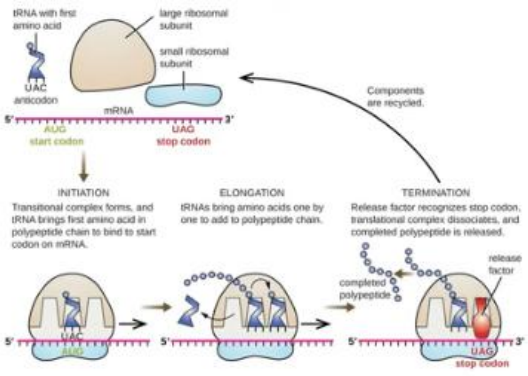


Figure: Diagrammatic representation of Initiation, Elongation and Termination of Translation
[Authors: OpenStax Microbiology, CC BY 4.0]