repair corrects a single mismatch base pair e.g. C to A, instead of T to A.

The template strand of the DNA exists in a methylated form, while the newly synthesized strand is not methylated. This difference allows the recognition of the new strands. The enzyme GATC endonuclease cuts the strand at an adjacent methylated GATC sequence (***Fig.24.18***). This is followed by an exonuclease digestion of the defective strand, and thus its removal. A new DNA strand is now synthesized to replace the damaged one.

Hereditary nonpolyposis colon cancer (HNPCC) is one of the most common inherited cancers. This cancer is now linked with ***faulty mismatch repair*** of defective DNA.

Double-strand break repair

Double-strand breaks (DSBs) in DNA are dangerous. They result in genetic recombination which may lead to chromosomal translocation, broken chromosomes, and finally cell death. DSBs can be repaired by homologous recombination or non-homologous end joining. Homologous recombination occurs in yeasts while in mammals, non-homologous and joining dominates.

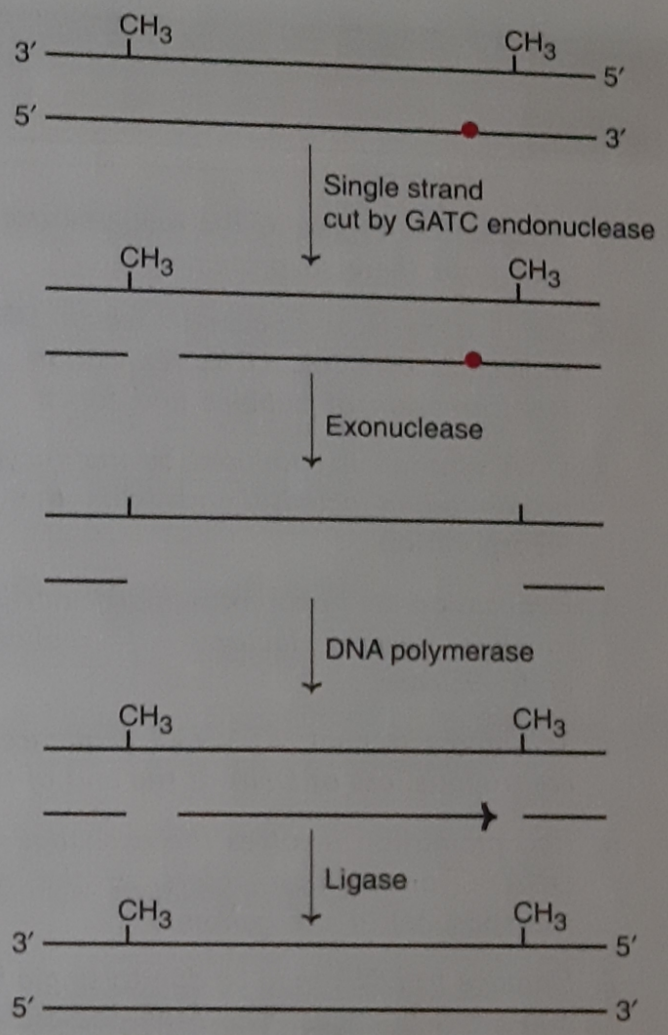

Fig. 24.18 : A diagrammatic representation of mismatch repair of DNA.

DEFECTS IN DNA REPAIR AND CANCER

Cancer develops when certain genes that regulate normal cell division fail or are altered. Defects in the genes encoding proteins involved in nucleotide-excision repair, mismatch repair and recombinational repair are linked to human cancers. For instance, as already referred above, HNPCC is due to a defect in mismatch repair.

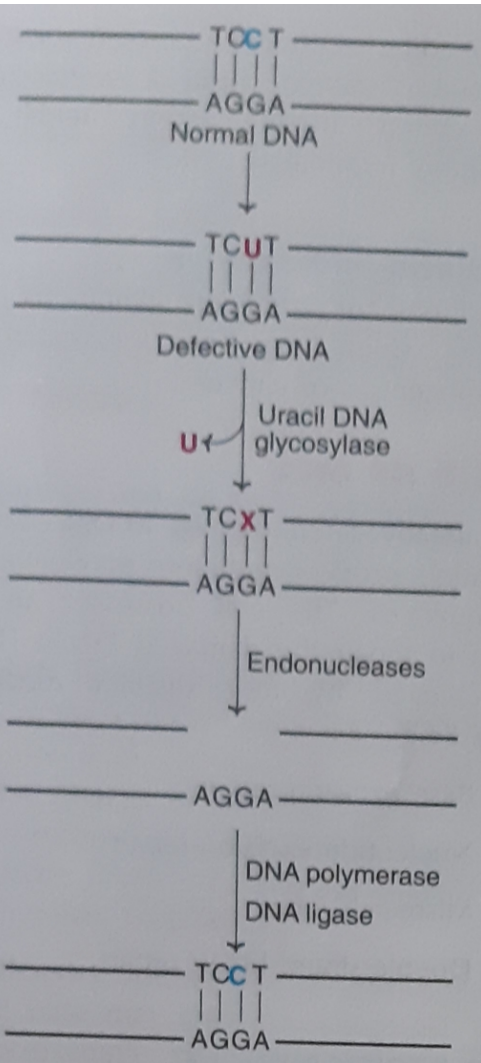

Fig. 24.16 : A diagrammatic representation of base excision-repair of DNA.

A defective DNA in which cytosine is deaminated to uracil is acted upon by the enzyme uracil DNA glycosylase. This results in the removal of the defective base uracil. An endonuclease cuts the backbone of DNA strand near the defect and removes a few bases. The gap so created is filled up by the action of repair DNA polymerase and DNA ligase.

Nucleotide excision-repair

The DNA damage due to ultraviolet light, ionizing radiation and other environmental factors often results in the modification of certain bases, strand breaks, cross-linkages etc. Nucleotide excision-repair is ideally suited for such large-scale defects in DNA. After the identification of the defective piece of the DNA, the DNA double helix is unwound to expose the damaged part. An ***excision nuclease*** (exinuclease) cuts the DNA on either side (upstream and downstream) of the damaged DNA. This defective piece is degraded. The gap created by the nucleotide excision is filled up by DNA polymerase which gets ligated by DNA ligase (***Fig.24.17***).

Xeroderma pigmentosum (XP) is a rare autosomal recessive disease. The affected patients are photosensitive and susceptible to skin cancers. It is now recognized that XP is due to a defect in the nucleotide excision repair of the damaged DNA.

Mismatch repair

Despite high accuracy in replication, defects do occur when the DNA is copied. For instance, cytosine (instead of thymine) could be incorporated opposite to adenine. Mismatch

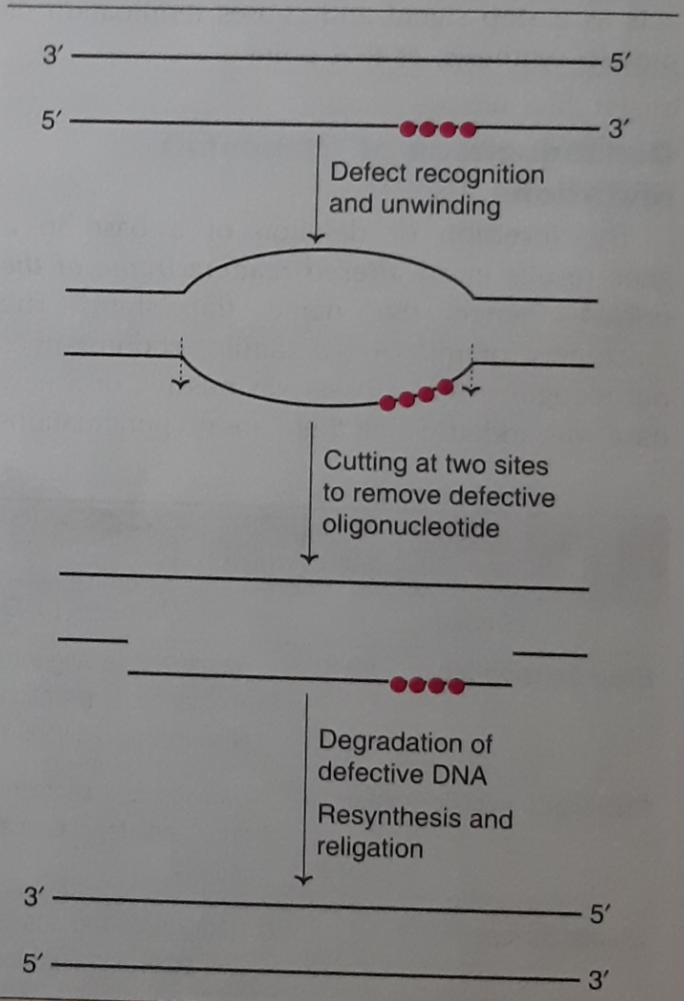

Fig. 24.17 : A diagrammatic representation of nucleotide excision-repair of DNA.

REPAIR OF DNA

As already stated, damage to DNA caused by replication errors or mutations may have serious consequences. The cell possesses an inbuilt system to repair the damaged DNA. This may be achieved by four distinct mechanisms (***Table 24.2***).

1. Base excision-repair
2. Nucleotide excision-repair
3. Mismatch repair
4. Double-strand break repair.

Base excision-repair

The bases cytosine, adenine and guanine can undergo spontaneous depurination to respectively form uracil, hypoxanthine and xanthine. These altered bases do not exist in the normal DNA, and therefore need to be removed. This is carried out by base excision repair (***Fig.24.16***).

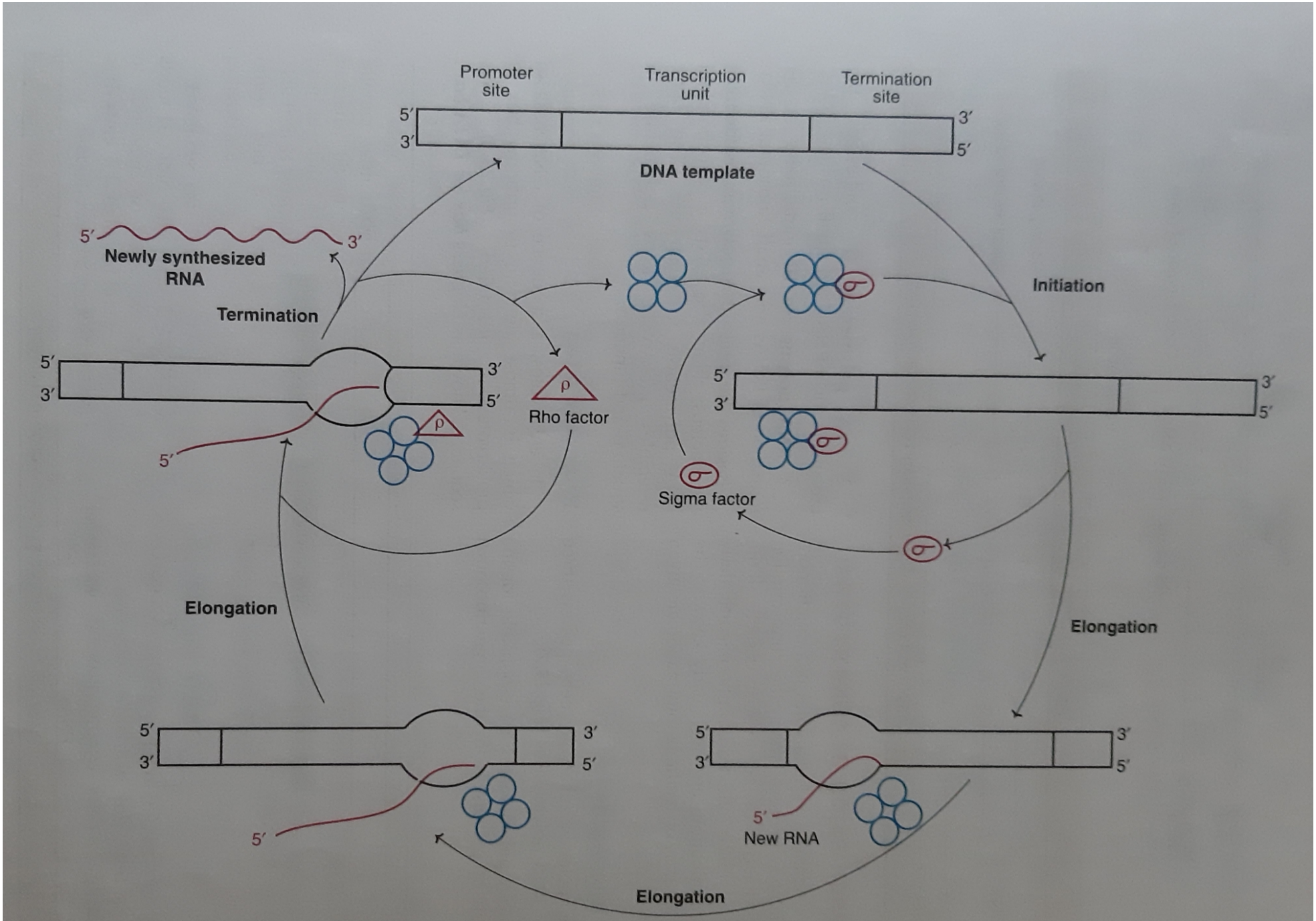

Fig. 25.4 : *Synthesis of RNA from DNA template (transcription).*

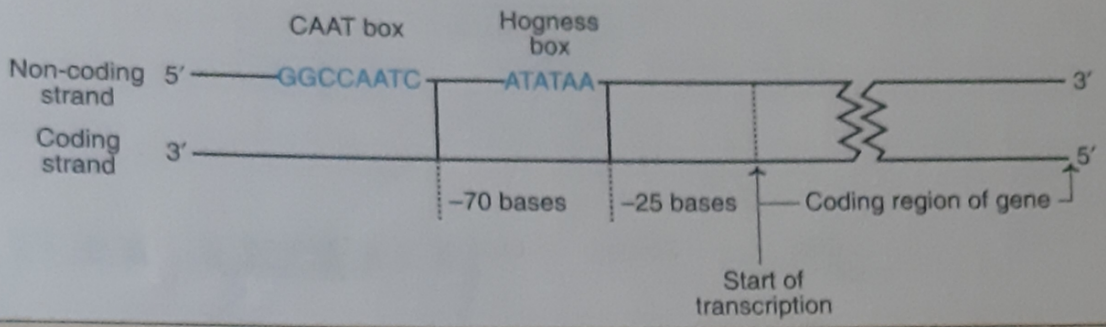

Fig. 25.8 : Promoter regions of DNA in eukaryotes.

is located on the left about 25 nucleotides away (upstream) from the starting site of mRNA synthesis. There also exists another site of recognition between 70 and 80 nucleotides upstream from the start of transcription. This second site is referred to as ***CAAT box***. One of these two sites (or sometimes both) helps RNA polymerase II to recognize the requisite sequence on DNA for transcription.

Initiation of transcription

The molecular events required for the initiation of transcription in eukaryotes are complex, and broadly involve three stages.

1. Chromatin containing the promoter sequence made accessible to the transcription machinery.

2. Binding of transcription factors (TFs) to DNA sequences in the promoter region.

3. Stimulation of transcription by enhancers.

A large number of ***transcription factors*** interact with eukaryotic promoter regions. In humans, about six transcription factors have been identified (TFIID, TFIIA, TFIIB, TFIIF, TFIIE, TFIIH). It is postulated that the TFs bind to each other, and in turn to the enzyme RNA polymerase.

Enhancer can increase gene expression by about 100 fold. This is made possible by binding of enhancers to transcription factors to form ***activators***. It is believed that the chromatin forms a loop that allows the promoter and enhancer to be close together in space to facilitate transcription.

Heterogeneous nuclear RNA (hnRNA)

The ***primary mRNA transcript*** produced by RNA polymerase II in eukaryotes is often referred to as heterogeneous nuclear RNA (hnRNA). This is then processed to produce mRNA needed for protein synthesis.

POST-TRANSCRIPTIONAL MODIFICATIONS

The RNAs produced during transcription are called primary transcripts. They undergo many alterations—***terminal base additions, base modifications, splicing*** etc., which are collectively referred to as post-transcriptional modifications. This process is required to convert the RNAs into the active forms. A group of enzymes, namely ribonucleases, are responsible for the processing of tRNAs and rRNAs of both prokaryotes and eukaryotes.

The prokaryotic mRNA synthesized in transcription is almost similar to the functional mRNA. In contrast, eukaryotic mRNA (i.e. hnRNA) undergoes extensive post-transcriptional changes.

An outline of the post-transcriptional modifications is given in ***Fig.25.9***, and some highlights are described.

Messenger RNA

The primary transcript of mRNA is the hnRNA in eukaryotes, which is subjected to many changes before functional mRNA is produced.

1. **The 5′ capping :** The 5′ end of mRNA is capped with 7-methylguanosine by an unusual

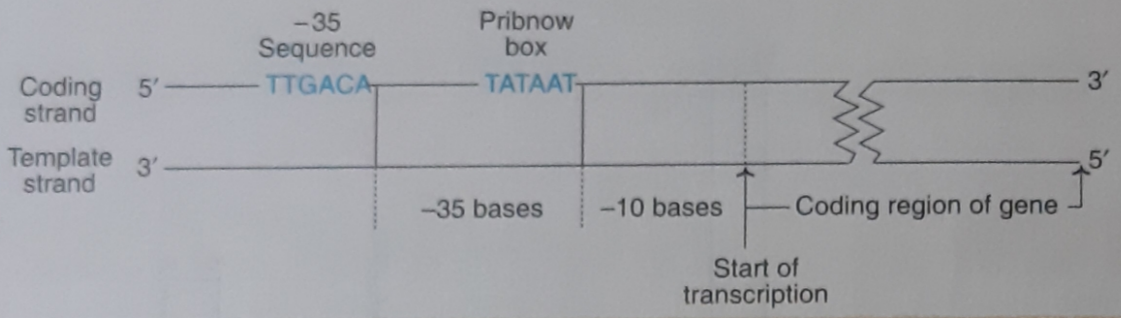

Fig. 25.5 : Promoter regions of DNA in prokaryotes.

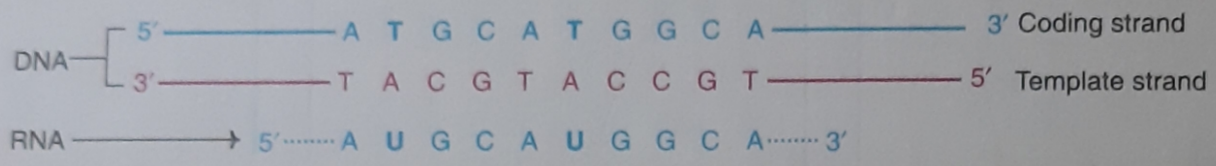

Fig. 25.6 : Transcription—Complementary base pair relationship.

TRANSCRIPTION IN EUKARYOTES

RNA synthesis in eukaryotes is a much more complicated process than the transcription described above for prokaryotes. As such, all the details of eukaryotic transcription (particularly about termination) are not clearly known. The salient features of available information are given hereunder.

RNA polymerases

The nuclei of eukaryotic cells possess three distinct RNA polymerases (***Fig.25.7***).

1. **RNA polymerase I** is responsible for the synthesis of precursors for the large ribosomal RNAs.

2. **RNA polymerase II** synthesizes the precursors for mRNAs and small nuclear RNAs.

3. **RNA polymerase III** participates in the formation of tRNAs and small ribosomal RNAs.

Besides the three RNA polymerases found in the nucleus, there also exists a mitochondrial RNA polymerase in eukaryotes. The latter resembles prokaryotic RNA polymerase in structure and function.

Promoter sites

In eukaryotes, a sequence of DNA bases—which is almost identical to pribnow box of prokaryotes—is identified (***Fig.25.8***). This sequence, known as ***Hogness box*** (or ***TATA box***),

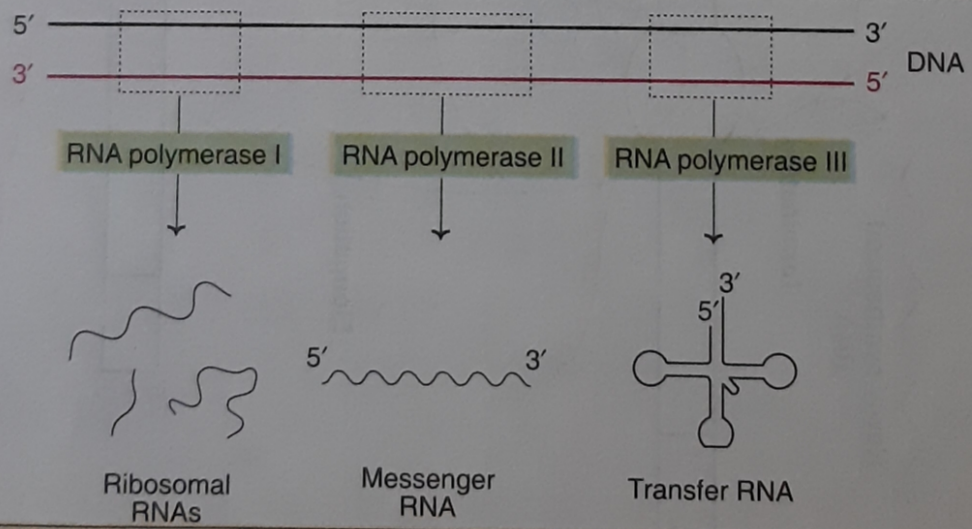

Fig. 25.7 : An overview of transcription in eukaryotes.

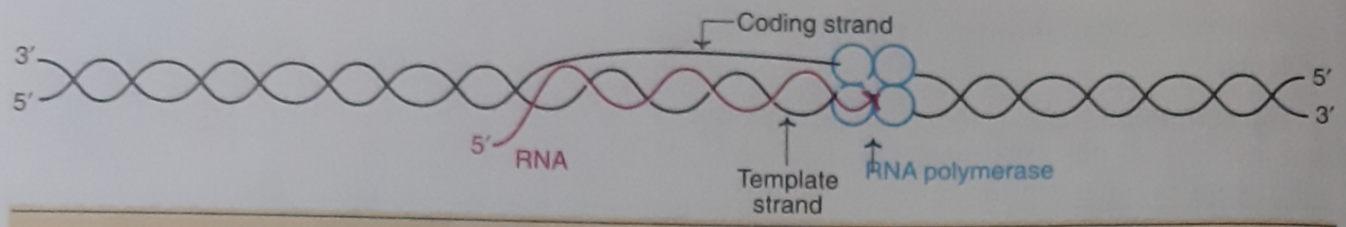

Fig. 25.3 : An overview of transcription.

Initiation

The binding of the enzyme RNA polymerase to DNA is the prerequisite for the transcription to start. The specific region on the DNA where the enzyme binds is known as ***promoter region.*** There are two base sequences on the ***coding DNA strand*** which the sigma factor of RNA polymerase can recognize for initiation of transcription (***Fig.25.5***).

1. **Pribnow box (TATA box) :** This consists of 6 nucleotide bases (TATAAT), located on the left side about 10 bases away (upstream) from the starting point of transcription.

2. **The '–35' sequence :** This is the second recognition site in the promoter region of DNA. It contains a base sequence TTGACA, which is located about 35 bases (upstream, hence –35) away on the left side from the site of transcription start.

Elongation

As the holoenzyme, RNA polymerase recognizes the promoter region, the sigma factor is released and transcription proceeds. RNA is synthesized from 5′ end to 3′ end (5′→3′) antiparallel to the DNA template. RNA polymerase utilizes ribonucleotide triphosphates (ATP, GTP, CTP and UTP) for the formation of RNA. For the addition of each nucleotide to the growing chain, a pyrophosphate moiety is released.

The sequence of nucleotide bases in the mRNA is complementary to the template DNA strand. It is however, identical to that of coding strand except that RNA contains U in place of T in DNA (***Fig.25.6***).

RNA polymerase differs from DNA polymerase in two aspects. No primer is required for RNA polymerase and, further, this enzyme does not possess endo- or exonuclease activity. Due to lack of the latter function (proof-reading activity), RNA polymerase has no ability to repair the mistakes in the RNA synthesized. This is in contrast to DNA replication which is carried out with high fidelity. It is, however, fortunate that mistakes in RNA synthesis are less dangerous, since they are not transmitted to the daughter cells.

The double helical structure of DNA unwinds as the transcription goes on, resulting in supercoils. The problem of supercoils is overcome by topoisomerases (more details in ***Chapter 24***).

Termination

The process of transcription stops by termination signals. Two types of termination are identified.

1. **Rho (ρ) dependent termination :** A specific protein, named ρ factor, binds to the growing RNA (and not to RNA polymerase) or weakly to DNA, and in the bound state it acts as ATPase and terminates transcription and releases RNA. The ρ factor is also responsible for the dissociation of RNA polymerase from DNA.

2. **Rho (ρ) independent termination :** The termination in this case is brought about by the formation of ***hairpins*** of newly synthesized RNA. This occurs due to the presence of ***palindromes***. A palindrome is a word that reads alike forward and backward e.g. madam, rotor. The presence of palindromes in the base sequence of DNA template (same when read in opposite direction) in the termination region is known. As a result of this, the newly synthesized RNA folds to form hairpins (due to complementary base pairing) that cause termination of transcription.

expression which directs the synthesis of proteins.

Transcriptomics : The study of transcriptome that involves all the RNA molecules made by a cell, tissue or an organism is transcriptomics.

PROTEOME

The ***cell's repertoire*** (repository/storehouse) of ***proteins*** with their nature ***and biological functions*** is regarded as proteome. Thus, proteome represents the entire range of proteins and their biological functions in a cell.

Proteomics : The study of the proteome.

Metabolomics : The use of genome sequence analysis for determining the capability of a cell, tissue or an organism to synthesize small molecules (metabolites) is metabolomics.

Whether the central dogma of life is represented in the conventional or more recent form, replication, transcription and translation are the key or core processes that ultimately control life. Replication of DNA has been described in ***Chapter 24***, while transcription and translation are discussed in this chapter.

TRANSCRIPTION

Transcription is a process in which ribonucleic acid ***(RNA) is synthesized from DNA***. The word ***gene*** refers to the ***functional unit of the DNA*** that can be transcribed. Thus, the genetic information stored in DNA is expressed through RNA. For this purpose, one of the two strands of DNA serves as a ***template*** (non-coding strand or ***sense strand***) and produces ***working copies of RNA molecules***. The other DNA strand which does not participate in transcription is referred to as coding strand or antisense strand (frequently referred to as coding strand since with the exception of T for U, primary mRNA contains codons with the same base sequence).

Transcription is selective

The entire molecule of DNA is not expressed in transcription. RNAs are synthesized only for some selected regions of DNA. For certain other regions of DNA, there may not be any transcription at all. The exact reason for the selective transcription is not known. This may be due to some inbuilt signals in the DNA molecule.

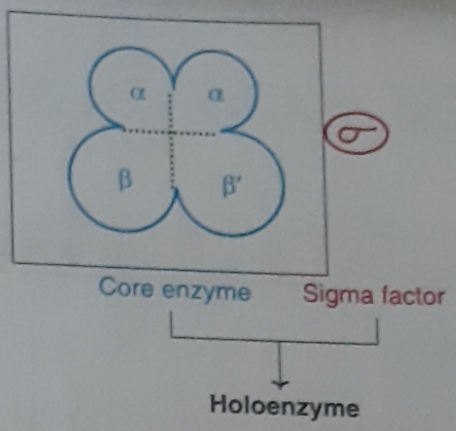

Fig. 25.2 : RNA polymerase of E. coli.

The product formed in transcription is referred to as ***primary transcript.*** Most often, the primary RNA transcripts are inactive. They undergo certain alterations (splicing, terminal additions, base modifications etc.) commonly known as ***post-transcriptional modifications***, to produce functionally active RNA molecules.

There exist certain differences in the transcription between prokaryotes and eukaryotes. The RNA synthesis in prokaryotes is given in some detail. This is followed by a brief discussion on eukaryotic transcription.

TRANSCRIPTION IN PROKARYOTES

A single enzyme—DNA dependent RNA polymerase or simply ***RNA polymerase***—synthesizes all the RNAs in prokaryotes. RNA polymerase of *E. coli* is a complex holoenzyme (mol wt. 465 kDa) with five polypeptide subunits—2α, 1β and 1β′ and one sigma(s) factor (***Fig.25.2***). The enzyme without sigma factor is referred to as core enzyme ($\alpha_2\beta\beta'$).

An overview of RNA synthesis is depicted in ***Fig.25.3***. Transcription involves three different stages—initiation, elongation and termination (***Fig.25.4***).

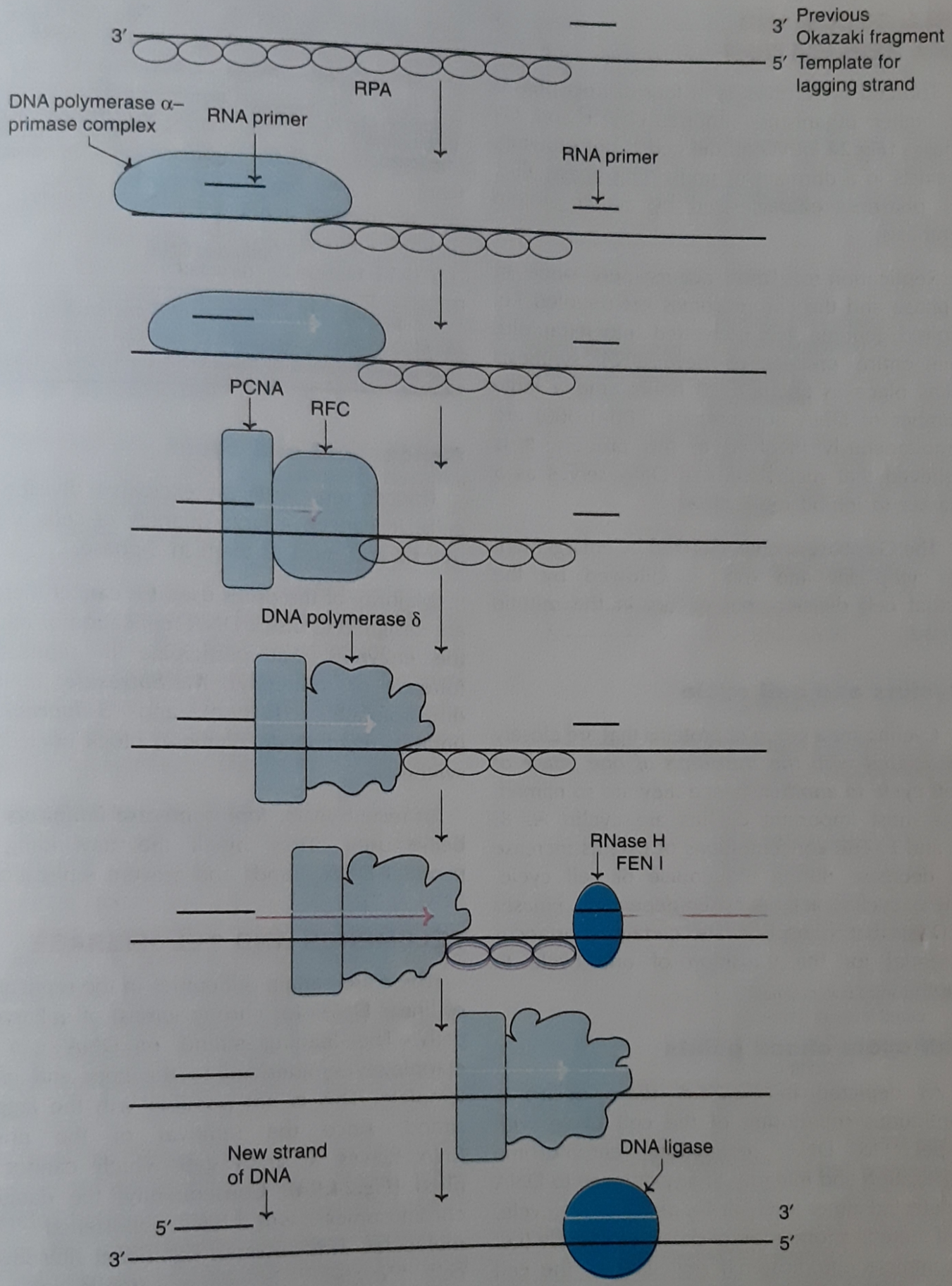

Fig. 24.7 : *An outline of DNA replication on the lagging strand in eukaryotes (RPA–Replication protein A; PCNA–Proliferating cell nuclear antigen; RFC–Replication factor C; RNase H–Ribonuclease H; FENI–Flap endonuclease I;* ***Note :*** *Leading strand not shown).*

5. **DNA polymerase** ε is involved in DNA synthesis on the lagging strand and proof-reading function.

The differences in the DNA replication between bacteria and human cells, attributed to the enzymes, are successfully used in antibacterial therapy to target pathogen (bacterial) replication and spare the host (human) cells.

The replication on the leading (continuous) strand of DNA is rather simple, involving ***DNA polymerase*** δ and a sliding clamp called ***proliferating cell nuclear antigen (PCNA)***. PCNA is so named as it was first detected as an antigen in the nuclei of replicating cells. PCNA forms a ring around DNA to which DNA polymerase δ binds. Formation of this ring also requires another factor namely ***replication factor C (RFC)***.

The ***replication on the lagging*** (discontinuous) ***strand in eukaryotes is more complex*** when compared to prokaryotes or even the leading strand of eukaryotes. This is depicted in ***Fig.24.7***, and briefly described hereunder.

The parental strands of DNA are separated by the enzyme helicase. A single-stranded DNA binding protein called ***replication protein A (RPA)*** binds to the exposed single-stranded template. This strand has been opened up by the replication fork (a previously formed Okazaki fragment with an RNA primer is also shown in ***Fig.24.4***).

The enzyme primase forms a complex with DNA polymerase α which initiates the synthesis of Okazaki fragments. The primase activity of pol α-primase complex is capable of producing 10-bp RNA primer. The enzyme activity is then switched from primase to DNA polymerase α which elongates the primer by the addition of 20–30 deoxyribonucleotides. Thus, by the action of pol α-primase complex, a short stretch of DNA attached to RNA is formed. And now the complex dissociates from the DNA.

The next step is the binding of replication factor C (RFC) to the elongated primer (short RNA-DNA). RFC serves as a clamp loader, and catalyses the assembly of proliferating cell nuclear antigen (PCNA) molecules. The DNA polymerase δ binds to the sliding clamp and elongates the Okazaki fragment to a final length of about 150–200 bp. By this elongation, the replication complex approaches the RNA primer of the previous Okazaki fragment.

The RNA primer removal is carried out by a pair of enzymes namely RNase H and flap endonuclease I (FENI). This gap created by RNA removal is filled by continued elongation of the new Okazaki fragment (carried out by polymerase δ, described above). The small nick that remains is finally sealed by DNA ligase.

Eukaryotic DNA is tightly bound to histones (basic proteins) to form nucleosomes which, in turn, organize into chromosomes. During the course of replication, the chromosomes are relaxed and the nucleosomes get loosened. The DNA strands separate for replication, and the parental histones associate with one of the parental strands. As the synthesis of new DNA strand proceeds, histones are also produced simultaneously, on the parent strand. At the end of replication, of the two daughter chromosomal DNAs formed, one contains the parental histones while the other has the newly synthesized histones.

Bacteria contain a specific type II topoisomerase namely ***gyrase***. This enzyme cuts and reseals the circular DNA (of bacteria), and thus overcomes the problem of supercoils. Bacterial gyrase is inhibited by the antibiotics ciprofloxacin, novobiocin and nalidixic acid. These are widely used as antibacterial agents since they can effectively block the replication of DNA and multiplication of cells. These antibacterial agents have almost no effect on human enzymes.

Certain compounds that ***inhibit human topoisomerases*** are used as anticancer agents e.g. adriamycin, etoposide, doxorubicin. The nucleotide analogs that inhibit DNA replication are also used as anticancer drugs e.g. 6-mercaptopurine, 5-fluorouracil.

Another enzyme—***DNA polymerase II***—has been isolated. It participates in the ***DNA repair process***.

Supercoils and DNA topoisomerases

As the double helix of DNA separates from one side and replication proceeds, supercoils are formed at the other side. The formation of supercoils can be better understood by comparing DNA helix with two twisted ropes tied at one end. Hold the ropes at the tied end in a fixed position. And let your friend pull the ropes apart from the other side. The formation of supercoils is clearly observed.

The problem of supercoils that comes in the way of DNA replication is solved by a group of enzymes called DNA topoisomerases. Type I DNA topoisomerase cuts the single DNA strand (nuclease activity) to overcome the problem of supercoils and then reseals the strand (ligase activity). Type II DNA topoisomerase (also known as DNA gyrase) cuts both strands and reseals them to overcome the problem of supercoils.

REPLICATION IN EUKARYOTES

Replication of DNA in eukaryotes closely resembles that of prokaryotes. Certain differences, however, exist. ***Multiple origins of replication*** is a characteristic feature of eukaryotic cell. Further, at least ***five distinct DNA polymerases*** are known in eukaryotes. Greek letters are used to number these enzymes.

1. **DNA polymerase** α is responsible for the synthesis of RNA primer for both the leading and lagging strands of DNA.

2. **DNA polymerase** β is involved in the repair of DNA. Its function is comparable with DNA polymerase I found in prokaryotes.

3. **DNA polymerase** γ participates in the replication of mitochondrial DNA.

4. **DNA polymerase** δ is responsible for the replication on the leading strand of DNA. It also possesses proof-reading activity.

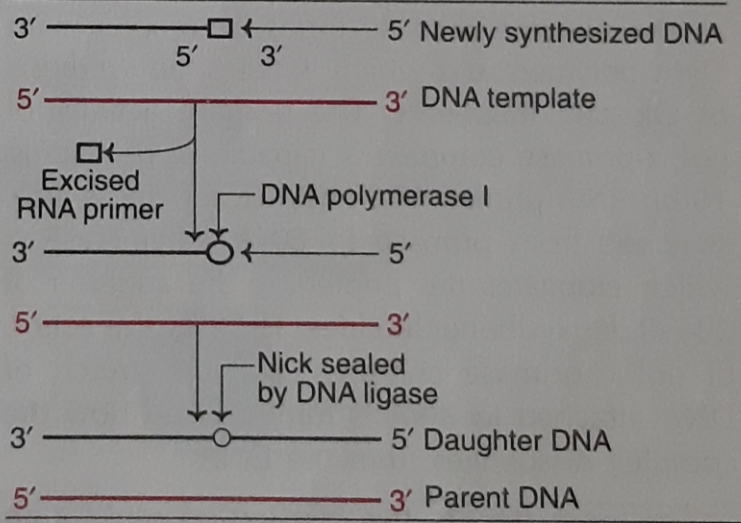

Fig. 24.6 : *Overview of the action of DNA polymerase I and DNA ligase.*