PROCESSING AND MODIFICATION OF RNA

RNA Splicing

RNA Editing

Dr. R. Prasad
Department of Zoology,
Eastern Karbi Anglong College,
Sarihajan

Although transcription is the first and most highly regulated step in gene expression, it is usually only the beginning of the series of events required to produce functional RNA. Most of the newly synthesized RNAs must be modified in various ways to be converted to their functional forms. You have studied that the transcription and translation in bacterial cells take place simultaneously and hence they are coupled. Therefore, bacterial mRNA has little opportunity to be modified before protein synthesis. In contrast, transcription and translation are separated in both time and space in eukaryotic cells. Transcription takes place in nucleus and translation in cytoplasm. Substantial processing is required in the nucleus to convert primary transcripts into mature mRNA molecules that are ready to be transported to the cytoplasm and translated. Primary transcripts are often very long, ranging from 2000 to 20,000 nucleotides. This heterogeneity in size is reflected in the term heterogeneous nuclear RNA (hnRNA). hnRNA consists of a mixture of mRNA molecules and their precursors i.e. pre-mRNA. The processing of mRNA includes modification of both ends of the initial transcripts as well as alteration of protein-coding section of RNA molecule. The C-terminal domain (CTD) of *RNA polymerase* II plays a key role in coordinating these processes by serving as binding site for the enzyme complexes involved in mRNA processing. The association of these processing enzymes with CTD of *polymerase* II accounts for their specificity in processing mRNAs. *RNA polymerase* I and III lack CTD, so their transcripts are not processed by the same enzyme complexes.

The mRNA processing in eukaryotes involves (i) addition of addition of 5´ –cap at its 5´ – end (ii) addition of poly (A) tail at its 3´ – end, (iii) RNA splicing, and (iv) RNA editing.

Addition of the 5’ Cap

The first step in mRNA processing is the modification of the 5´ end of the transcript by the addition of a structure called a 7-methylguanosine cap.

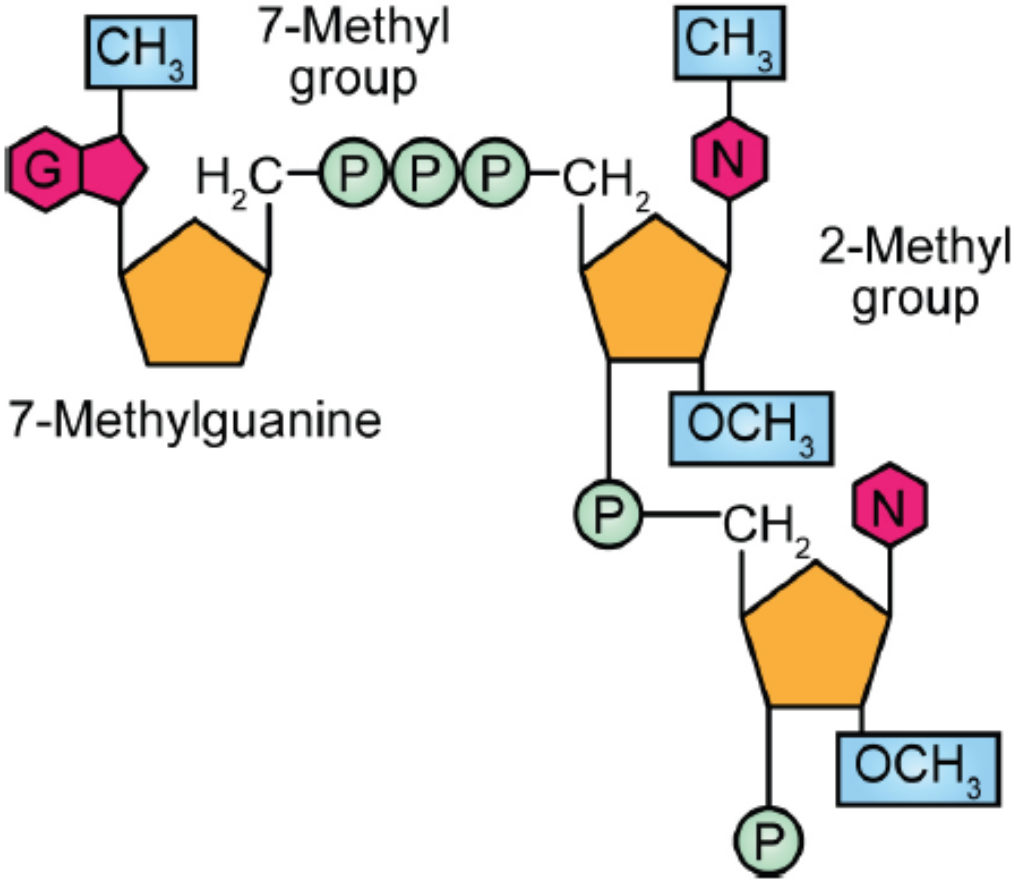

Fig.12.18: Eukaryotic mRNA 5´ cap consisting of nucleotide with 7 – methylguanine attached to the pre-mRNA by a unique 5´ – 5´ bond.

The enzyme responsible for capping are recruited to the phosphorylated CTD following initiation of transcription, and the cap is added after transcription of the 20 to 30 nucleotides of the RNA. Capping is initiated by the addition of a guanosine triphosphate (GTP) in reverse orientation to the 5′ nucleotides of the RNA. Then methyl groups are added to the G residues and to the ribose moieties of one of the 5′ nucleotides of the RNA chain (see Fig. 12.18). The 5′ cap contributes to mRNA stability by protecting the molecule from degradation by nucleases that attack RNA at the 5′ end. The 5′ cap also plays an important role in positioning and stabilizing mRNA on the ribosome for initiation of translation.

Addition of Poly (A) Tail at 3’ End

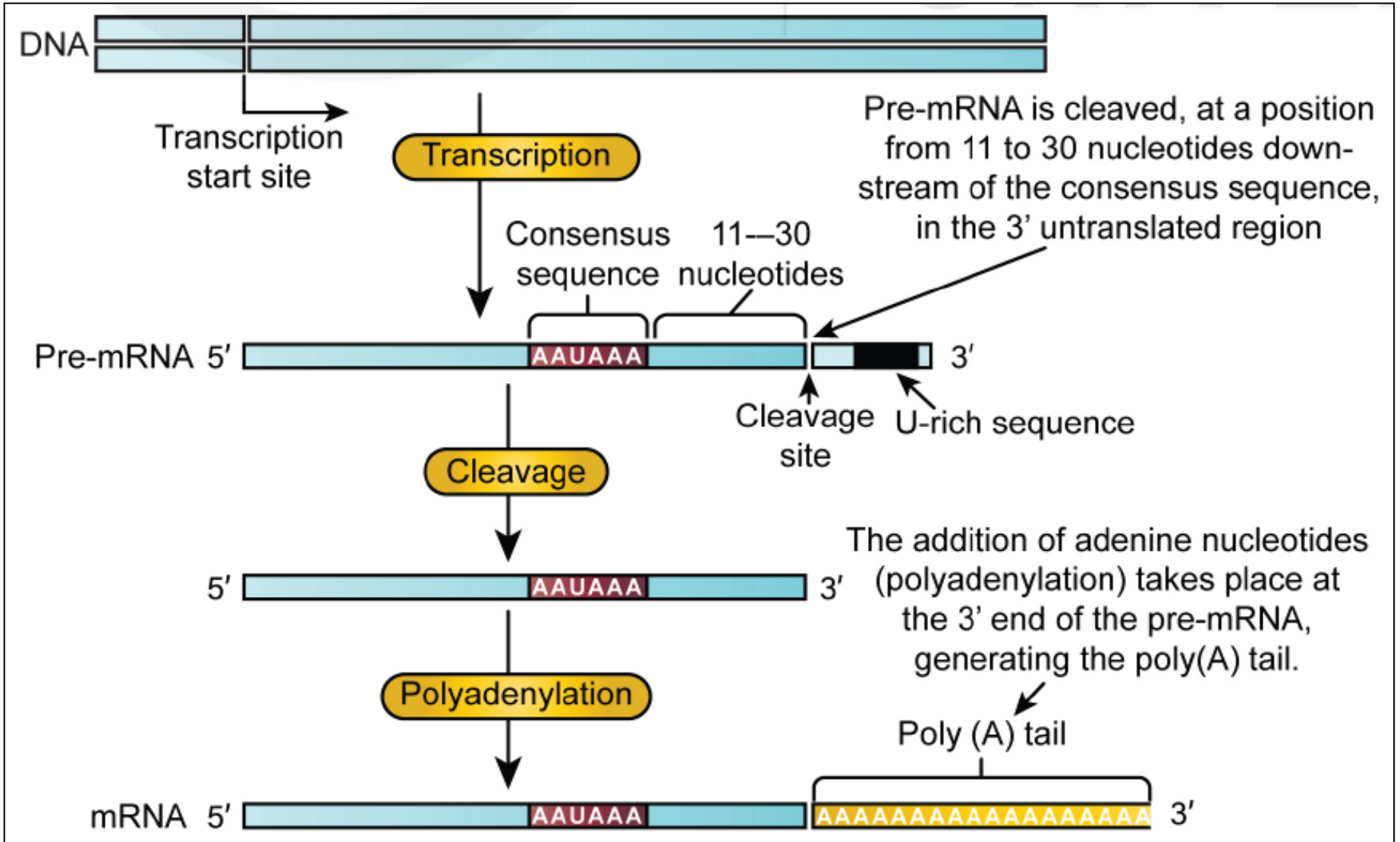

Fig.12.19: Eukaryotic mRNA with 3´ poly (A) tail. In pre-mRNA processing poly (A) tail is added through cleavage and polyadenylation.

In addition to the 5′ cap, a poly (A) tail ranging from 50 – 250 nucleotides in length is present at the 3′ end of the most of the eukaryotic mRNA molecules. The mRNAs coding for the major histones in animal cells are among the few mRNAs known to lack such a poly (A) tail. The poly (A) tail is always added after transcription as genes do not contain long stretches of thymine (T) nucleotides that could serve as a template for the addition of poly (A) tail. This process of addition of poly (A) tail is known as adenylation (see Fig. 12.19).The isolation of the enzyme *poly (A) polymerase*, which catalyzes the addition of poly (A) sequences to RNA without requiring a DNA template has given direct support to the above conclusion. The addition of poly (A) is a part of the process that creates the 3′ end of the most eukaryotic mRNA molecules. Processing of the 3′ end of pre-mRNA requires sequences both upstream and downstream of the cleavage site. Unlike bacteria, where specific termination sequences halt transcription at 3′ end of newly forming mRNAs, the transcription of eukaryotic pre-mRNAs often proceeds hundreds or even thousands of nucleotides beyond the site destined to become 3′ end of the final mRNA molecule. A special signaling (consensus) sequence AAUAAA is usually located slightly upstream of the cleavage site and a sequence rich in uracil (U) nucleotides or GU-rich nucleotides downstream of the cleavage site determines where the poly (A) tail should be added. This signaling element triggers cleavage of the primary transcript 10 -35 nucleotides downstream from the AAUAAA sequence and *poly (A) polymerase* catalyses the formation of poly (A) tail.

Like the 5´ cap, poly (A) tail protects mRNA from *nuclease* attack and, as a result, the length of the poly (A) influences mRNA stability (the longer the tail, the longer the life span of the mRNA in the cytoplasm). In addition, poly (A) is recognized by specific proteins involved in exporting mRNA from the nucleus to the cytoplasm, and it may also help ribosomes recognize mRNA as molecule to be translated. Figure 12.20 shows a mRNA with both 5´ cap and poly(A) tail.

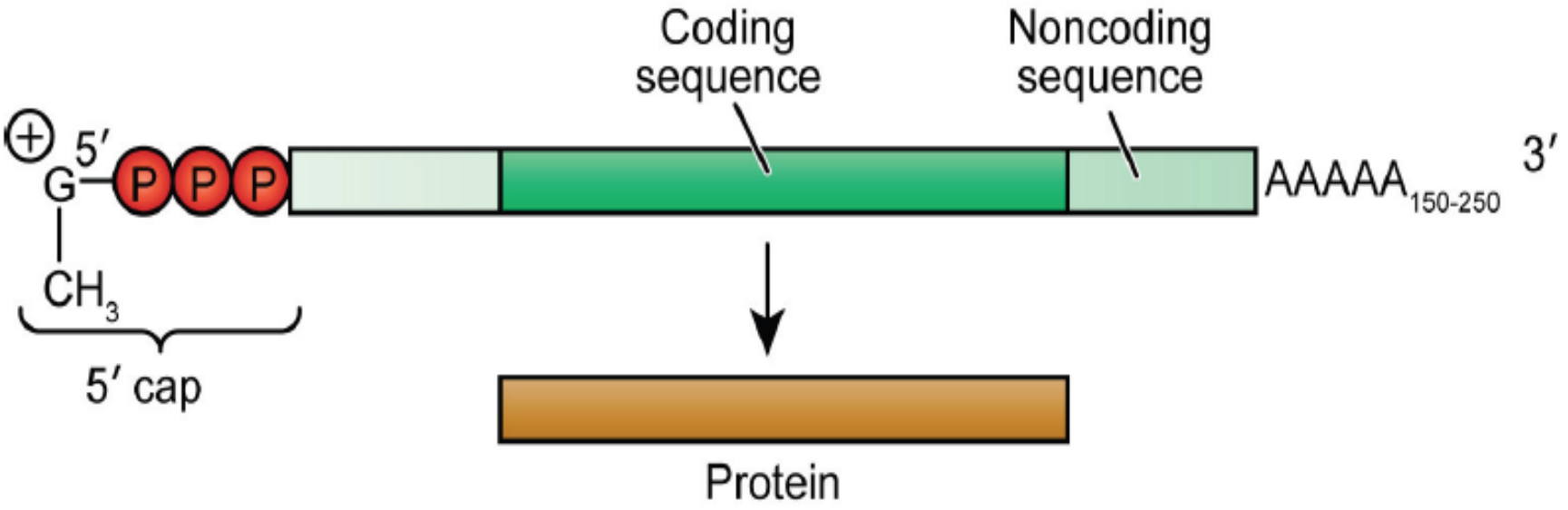

Fig.12.20: Formation of eukaryotic mRNAs by addition of a 5´ cap, cleavage of pre-mRNA transcript, and the addition of poly (A) tail.

RNA Splicing

One of the most exciting breakthroughs in the history of molecular genetics occurred in 1977, when internal DNA sequences in eukaryotes were discovered that are not expressed in the amino acid sequence of the protein they code. These internal DNA sequences are represented in initial RNA transcripts, but they are removed before the mature mRNA is translated. Such nucleotides segments are called **intervening sequence** and the genes that contain them are called **split genes**. DNA sequences that are not represented in the final mRNA product are also called **introns** and those retained and expressed are called **exons** (see Fig.12.21). Although the vast majority of eukaryotic genes examined thus far contain introns, there are several exceptions. Interestingly, the genes coding for histones and interferon lack introns. It is not clear why or how the genes encoding these molecules have been maintained throughout evolution without acquiring introns which is the characteristics of all other genes.

Fig.12.21: Split genes in eukaryotes. The intervening sequences contain various exons and introns. The numbers indicate the number of nucleotides present in various introns and exons regions.

The most striking modification of pre-mRNAs therefore, is the removal of introns through splicing and rejoining together the remaining RNA segments (exons).

Splicing requires the presence of three **consensus sequences** in the intron. One end of the intron is referred to as the **5´ splice site**, and the other end is the **3´ splice site.** These splice sites possess short consensus sequences (see Fig. 12.22). Most introns in pre-mRNA begin with GU and end with AG. This is indicative of the fact that these sequences play crucial role in splicing. The third important requisite for splicing is at the branch point, which is an adenine nucleotide that lies from 18 to 40 nucleotides upstream of the 3´ splicing site. The sequences surrounding the branch point does not have strong consensus. The deletion or mutation of the adenine nucleotide at the branch point will prevent splicing. Splicing takes place within a large structure called **spliceosome**, which is one of the largest and most complex of all molecular complexes. The spliceosome consists of five RNA molecules and almost 300 proteins. The RNA components are small nuclear RNAs (snRNAs) ranging in length from 107 to 210 nucleotides. These snRNAs associate with proteins to form **snRNPs** (small ribonucleoproteins). Each snRNPs (pronounced "snurp") contains a single RNA molecule and multiple proteins. The spliceosome is composed of five snRNPs (U1, U2, U4, U5 and U6), and some proteins not associated with an snRNA.

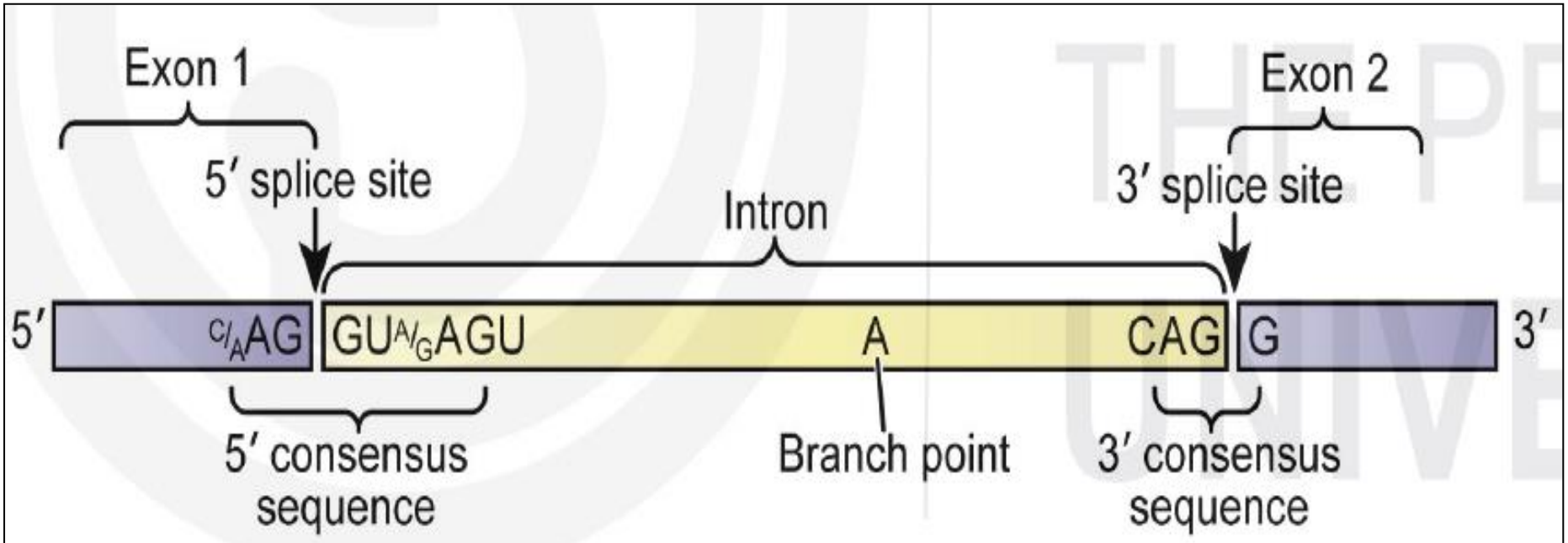

Fig. 12.22: Pre-mRNA showing consensus sequences required for splicing.

RNA Editing

Now you will be introduced to another very interesting aspect. It has been believed since long that all information about the amino acid sequence of a protein resides in DNA. This may not always be true. The process of RNA editing weakens this assumption. In RNA editing, the coding sequence of an mRNA molecule is altered after transcription, and so the protein has an amino acid sequence that differs from that encoded by the gene. RNA editing was first detected in 1986 when the coding sequences of mRNAs were compared with the coding sequences of the DNAs from which they had been transcribed.

In some nuclear genes in mammalian cells and some mitochondrial genes in plant cells, there had been substitutions in some of the nucleotides of mRNAs. The mitochondrial gene in trypanosome parasites (which causes African sleeping sickness) undergoes extensive RNA editing. Different types of RNA editing have now been observed in mRNAs, tRNAs, and rRNAs from the wide range of organisms that include insertion and deletion of the nucleotides and conversion of one base to another. A variety of mechanisms can bring about changes in RNA sequences. In some cases, molecules called **guide RNAs** (gRNAs) play a crucial role. A gRNA contains sequences that are partly complementary to segments of the pre-edited RNA, and the two molecules undergo base pairing in these sequences (Fig. 12.24). After the mRNA is anchored to the gRNA, the mRNA undergoes cleavage and nucleotides are added, deleted, or altered to the template provided by gRNA. A recently discovered RNA editing technique that has revolutionized the field of genome engineering is CRISPER/cas9.

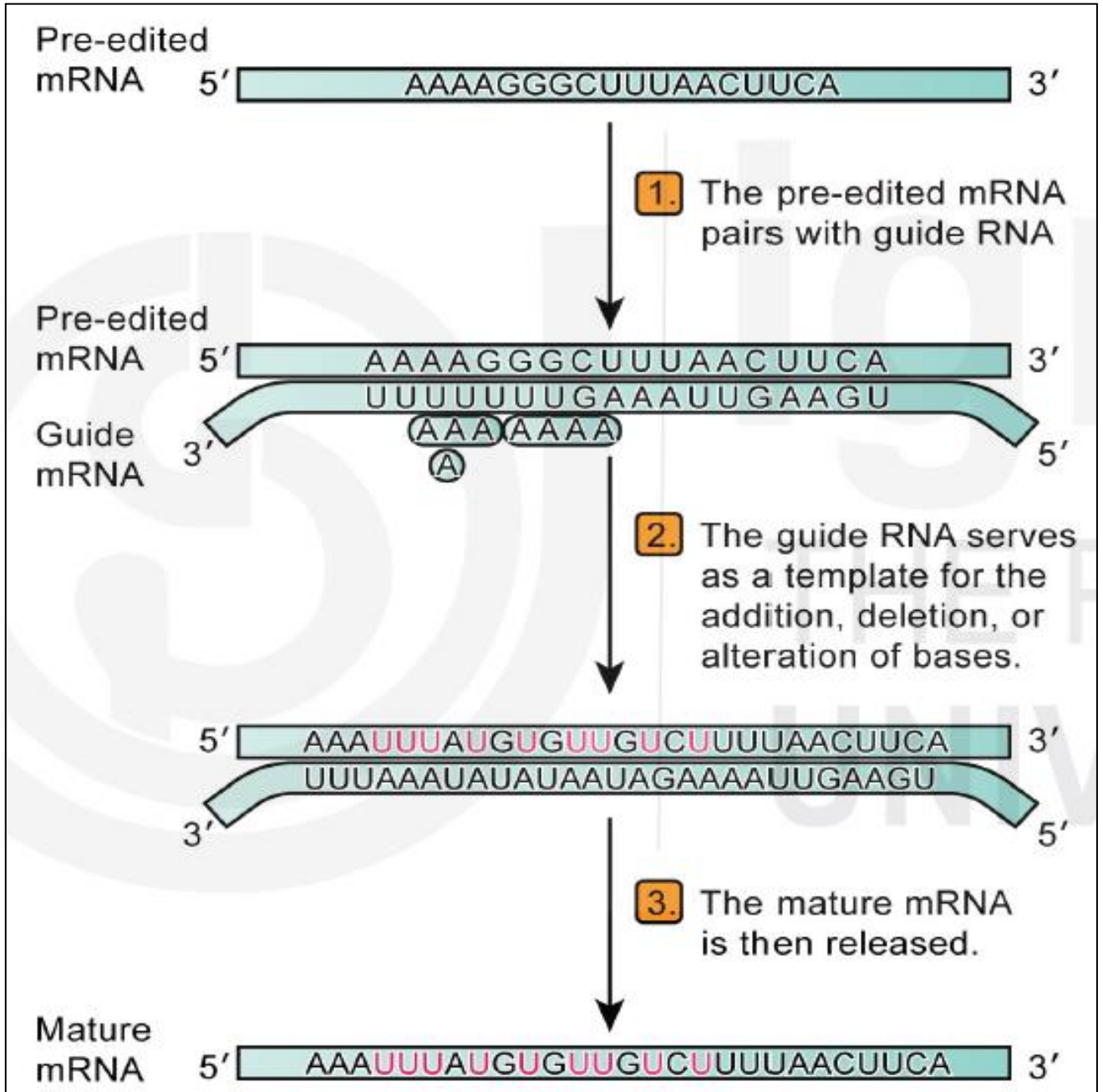

Fig. 12.24: Process of RNA editing by guide RNAs.