From DNA to Protein

Carsten Müssig

eLearning in Studienbereichen – MNF – Universität Potsdam

Content

Transcription

Quiz 2
Overview 2
mRNA 2
Bacterial mRNA 3
Structure of bacterial mRNA 4
Operons 4
Eukaryotic mRNA 5
Splicing of eukaryotic mRNA 6
Splicing apparatus 6
Alternative splicing: increasing protein diversity 7
tRNA 7
rRNA 8
Additional classes of ncRNA 8
Regulatory DNA sequences 9
Prokaryotic transcription 9
δ subunit 9
Initiation 10
Elongation 10
Termination 10
Eukaryotic transcription 11
Eukaryotic promoters 12
Enhancers 13
RNA transport 13
Further Readings 13
References 13
OER and OCW 14

Translation

Quiz 15
Overview 15
Initiation 15
Bacteria 15
Eukaryotes 15
Ribosomes 16
A, P, and E sites of ribosomes and translocation 17
Termination 18
Genetic code 18
Mutation 19
Wobble 19
Further Readings 20
References 20
OER and OCW 20

Control of gene expression

Quiz 21
Overview 21
Control of transcription 22
Initiation: Promoters and enhancers 22
Chromatin remodelling 23
The histone code and DNA methylation 23
Control of transcription elongation and termination 25
Control of translation 25
Control of mRNA degradation 26
Further Readings 27
References 27
OER and OCW 27

Dieses Material steht unter der **Creative-Commons-Lizenz** Namensnennung - Weitergabe unter gleichen Bedingungen 4.0 International. Um eine Kopie dieser Lizenz zu sehen, besuchen Sie http://creativecommons.org/licenses/by-sa/4.0/. Trotz sorgfältiger inhaltlicher Kontrolle wird keine Haftung für Inhalte externer Links übernommen. Für den Inhalt dieser Seiten sind ausschließlich deren Betreiber verantwortlich. Für die hier dargebotenen Informationen wird kein Anspruch auf Vollständigkeit, Aktualität, Qualität und Richtigkeit erhoben. Es kann keine Verantwortung für Schäden übernommen werden, die durch das Vertrauen auf die Inhalte oder deren Gebrauch entstehen.

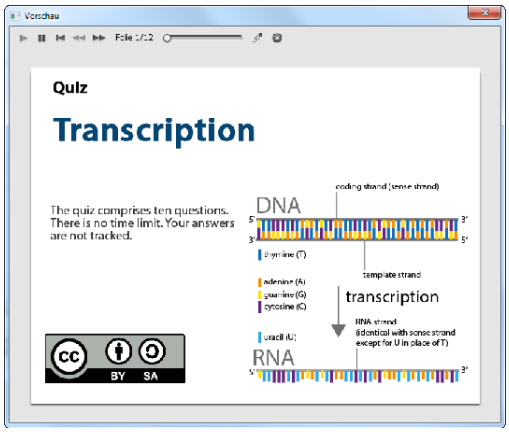
Letzte Überarbeitung: **Oktober 2014**.

Transcription

Transcription is the process by which the information contained in DNA is transferred to a newly assembled RNA.

Quiz

Start this chapter with the quiz to test for your mastery. You may reattempt the quiz later to check your progress.



Quiz 1. Transcription. Click to play - HTML5 or Flash (opens new window in your standard web browser).

Overview

Chromosomal DNA is the template for RNA molecules. All cellular RNA molecules (messenger RNA and non-coding RNAs) are made on DNA templates.

Transcription is the process whereby one strand of a DNA molecule is used as a template for synthesis of a complementary RNA by **RNA polymerase**.

Splicing (see page 6) is a modification of mRNAs after transcription, in which introns of precursor mRNA (pre-mRNA) are removed and exons of it are joined. Splicing only occurs in eukaryotes.

In prokaryotic genomes introns do not exist.

All proteins are determined by mRNA templates. The mature mRNA is translated by a ribosome.

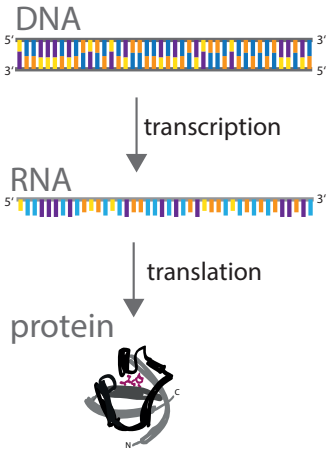


Figure 1. Genetic information directs the synthesis of proteins. This is the process that converts a mRNA sequence into a protein. Translation is necessarily preceded by transcription.

In **Translation** (see page 15), mRNA is decoded to produce a polypeptide according to the rules specified by the genetic code (Figure 1).

In **prokaryotic cells**, which have no nuclear compartment, the processes of transcription and translation occur in the cytoplasm and may be linked together.

In **eukaryotic cells**, the site of transcription (the cell nucleus) is separated from the site of translation (the cytoplasm). The RNA molecules move to the cytoplasm, where they can be bound by ribosomes.

Fore more details on the difference between prokaryotes and eukaryotes see:

S. Clancy (2008) RNA transcription by RNA polymerase: prokaryotes vs eukaryotes. Nature Education 1: 125

mRNA

A gene is not directly translated into protein, but is expressed via the production of a mRNA, a nucleic acid intermediate actually used to synthesize a protein.

All of the RNA in a cell is made by transcription. mRNA corresponds to one strand of the DNA double helix (Figure 2).

- The sequence of mRNA is complementary with the sequence of one strand

of DNA and is identical (apart from the replacement of T with U) with the other strand of DNA.

- The strand of DNA that directs synthesis of the mRNA via complementary base pairing is called the **template strand** or **antisense strand**. Antisense is used as a general term to describe a sequence of DNA or RNA that is complementary to mRNA.

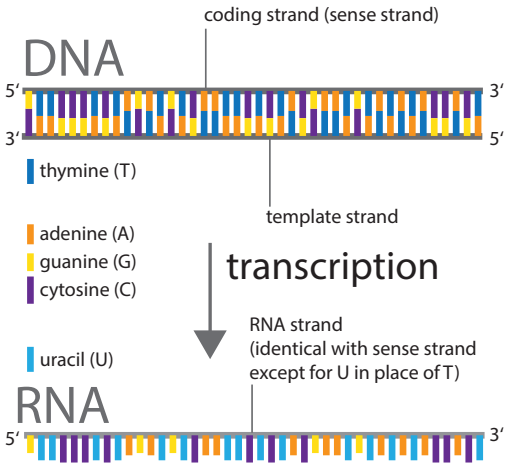


Figure 2. The coding strand of DNA has the same sequence as that of the RNA transcript except for thymine (T) in place of uracil (U). The RNA sequence is identical with one strand of the DNA (the coding or sense strand). It is complementary to the other strand, which provides the template strand for its synthesis.

- The other DNA strand bears the same sequence as the mRNA (except for possessing T instead of U), and is called the **coding strand** or **sense strand**.
- The convention for writing DNA sequences is that the top strand is the coding strand and runs **5′ to 3′**, with the sequence that is the same as RNA.
- The entire length of a mRNA is not translated, but each mRNA contains at least one coding region that is related to a protein sequence by the genetic code: each nucleotide triplet (codon) of the coding region represents one amino acid.

Messenger RNA has the same function in all cells, but there are important differences in the details of the synthesis and structure of prokaryotic and eukaryotic mRNA.

Bacterial mRNA

In bacteria, mRNA is **transcribed and translated** in the single cellular compartment. Transcription and translation occur **simultaneously**. As soon as transcription begins, ribosomes attach to the 5′ end of the mRNA and start translation, even before the rest of the message has been synthesized.

Since ribosomes attach to bacterial mRNA even before its transcription has been completed, the ribosomes are likely still to be attached to DNA.

A bunch of ribosomes moves along the mRNA while it is being synthesized. **Polyribosomes** (or polysomes) are a cluster of ribosomes, bound to a mRNA molecule. Polyribosomes read one strand of mRNA simultaneously, helping to synthesize the same protein at different spots on the mRNA. Bacterial transcription and translation take place at similar rates (~45 nucleotides or ~15 amino acids per second).

The 3′ end of the mRNA is generated when transcription terminates.

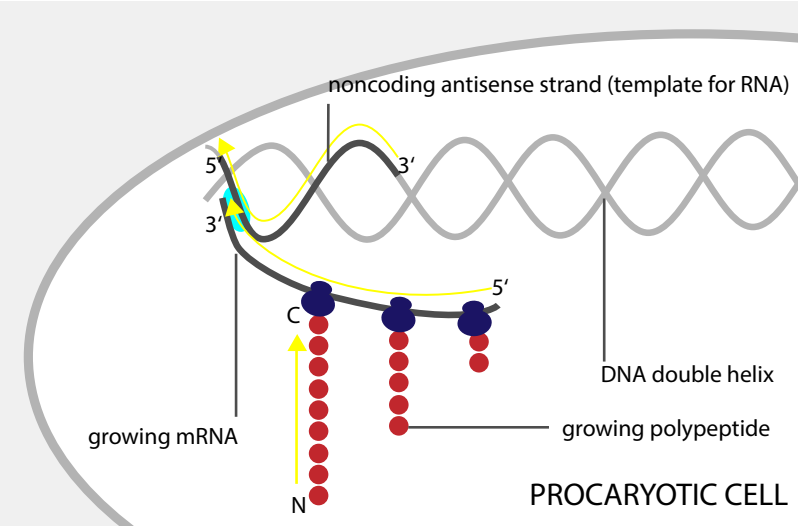


Figure 3. Transcription, translation, and degradation all occur in the same direction.

CC BY SA

Bacterial mRNA usually is **unstable**, and is therefore translated into proteins only a few minutes. The mRNA is synthesized, translated by the ribosomes, and degraded, all in rapid succession. The 5′ end of the mRNA starts to decay before the 3′ end has been synthesized or translated.

Structure of bacterial mRNA

Some mRNAs represent only a single protein: they are **monocistronic**. Others (the majority) carry sequences coding for several proteins: they are **polycistronic** (Figure 4).

All mRNAs contain two types of region.

The **coding region** consists of a series of codons representing the amino acid sequence of the protein, starting with AUG and ending with a termination codon.

But the mRNA is always longer than the coding region, extra regions are present at both ends. An additional sequence at the 5′ end, preceding the start of the coding region, is described as the leader or **5′ UTR** (untranslated region, leader sequence). An additional sequence following the termination signal, forming the 3′ end, is called the trailer or **3′ UTR**. These sequences are not used to code for protein.

A polycistronic mRNA also contains **intercistronic regions**. They vary greatly in size.

Operons

Bacterial genes are organized into operons. Operons are clusters of coregulated genes. An **operon** includes an operator, a common promoter, and one or more structural genes that are controlled as a unit to produce mRNA.

One of the best-studied examples is the ***lac*** operon from the bacterium *Escherichia coli*.

The *lac* operon is rapidly induced when *E. coli* cells get lactose as the only carbon source. It comprises three proteins required for lactose metabolism. In the absence of lactose (and presence of glucose), these enzymes are not necessary and the expression of the three proteins is repressed by the *lac* repressor protein.

For more details on the *lac* operon see:

J.M. Berg, J.L. Tymoczko, L. Stryer (2002) Prokaryotic DNA-binding proteins bind specifically to regulatory sites in operons. In: Biochemistry, 5th edition. W.H. Freeman: http://www.ncbi.nlm.nih.gov/books/NBK22512/

H. Lodish, A. Berg, S.L. Zipursky et al. (2000) Molecular Cell Biology, 4th edition. W.H. Freeman: http://www.ncbi.nlm.nih.gov/books/NBK21683/

A. Ralston (2008) Operons and prokaryotic gene regulation. Nature Education 1: 216

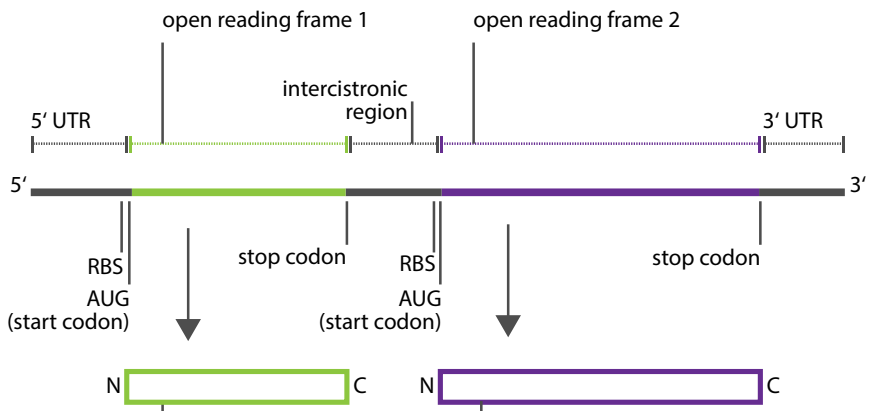


Figure 4. Structure of a polycistronic bacterial mRNA.

CC BY SA

Eukaryotic mRNA

The production of eukaryotic mRNA involves additional stages after transcription (Figure 5). Both ends of the transcript are modified by additions of further nucleotides (involving additional enzymes).

The 5′ end of the RNA is modified by addition of a "**cap**" virtually as soon as it appears. This replaces the triphosphate of the initial transcript with a nucleotide in reverse (3′ to 5′) orientation, thus "sealing" the end.

- A 5′ cap is formed by adding a G to the terminal base of the transcript via a 5′ to 5′ link (Figure 6). The new G residue added to the end of the RNA is in the reverse orientation from all the other nucleotides.
- One to three methyl groups are added. From yeast to human, this modification is necessary for efficient gene expression and cell viability.

The 5′ cap is required for the translation of the majority of mRNAs.

It stabilizes mRNA against attack by exonucleases and promotes transcription, splicing, polyadenylation and nuclear export of mRNA

The 3′ end is modified by addition of a series of adenylic acid nucleotides, polyadenylic acid or **poly(A)**.

- Polyadenylation of mRNA precursors is a two-step process. An endonucleolytic cleavage first generates a free hydroxyl group at the 3′ end of the pre-mRNA. The second step is the addition of a poly(A) tail to the 3′ end of the nascent transcript.
- The average poly(A) tail is species-specific (e.g. yeast 70 and humans 250 adenosyl residues). The poly(A) sequence is not coded in the DNA. The poly(A) tail of both nuclear RNA and mRNA is associated with a protein called the poly(A)-binding protein (PABP).
- Only after the completion of all modifications and processing events can the mRNA be exported from the nucleus to the cytoplasm. Once the mRNA has entered the cytoplasm, it is recognized by ribosomes and translated.

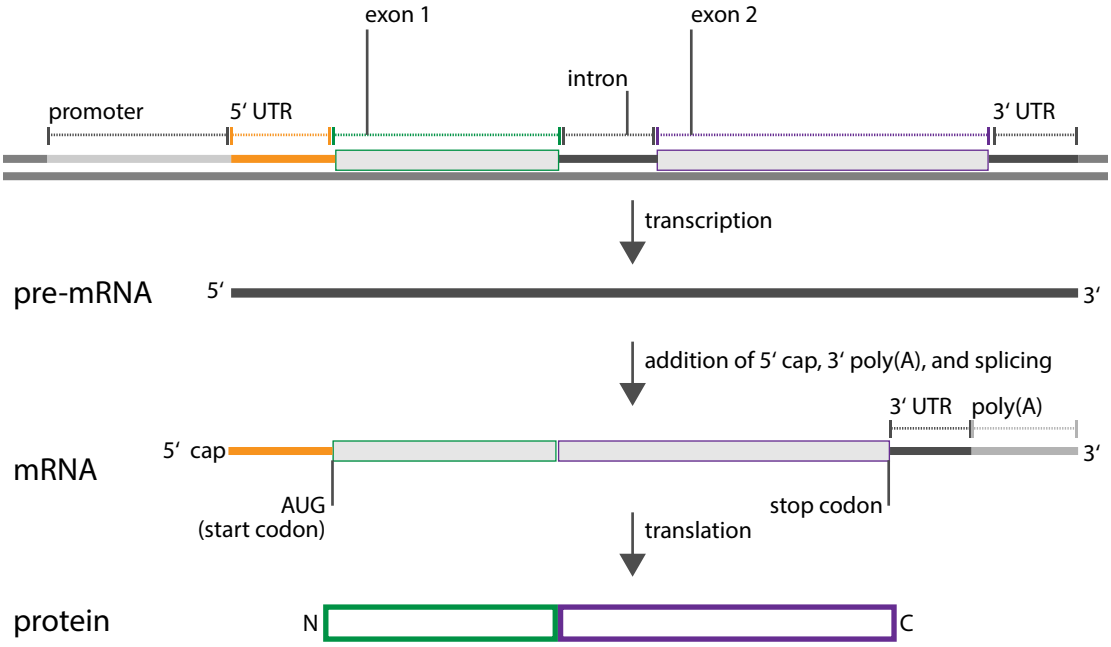


Figure 5. Structure of eukaryotic gene and production of mRNA. The pre-mRNA comprises introns and exons. Exons comprise untranlated region (UTRs) and coding sequences. The 5′ cap and 3′ poly(A) tail are added to the mRNA. The mature mRNA is transported into the cytoplasma and translated into protein. In contrast to bacterial mRNA, eukaryotic mRNA is relatively stable.

CC BY SA

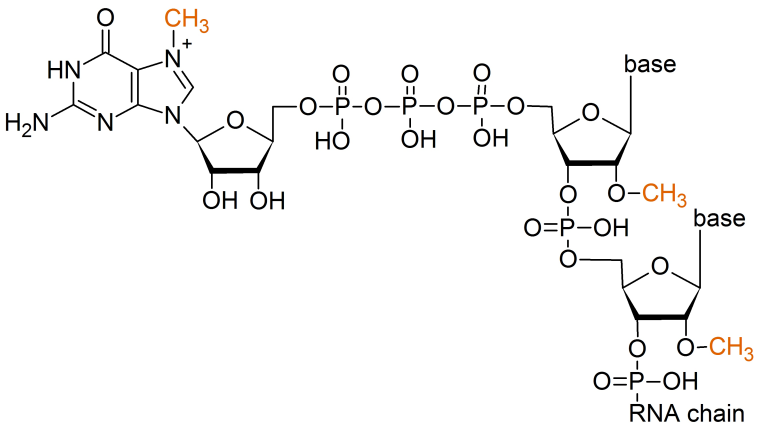


Figure 6. 5′ cap of eukaryotic mRNA.

Splicing of eukaryotic mRNA

Most eukaryotic genes have the coding sequences (**exons**) interrupted by non-coding sequences (**introns**) (Figure 7). The primary transcript has the same organization as the gene, and is called the **pre-mRNA**.

Removal of the introns from pre-mRNA is called RNA **splicing**. Splicing occurs in the nucleus, together with other modifications (i.e. capping and addition of poly(A) tail) that are made to newly synthesized RNAs.

Each intron contains a few nucleotides that act as cues for its removal. The **splice sites** are the sequences immediately surrounding the exon-intron boundaries.

The splice sites have well conserved, though rather short, consensus sequences. The 5′ splice site at the 5′ (left) end of the intron includes the consensus sequence **GU**. The 3′ splice site at the 3′ (right) end of the intron includes the consensus sequence **AG**. Splicing depends on recognition of these pairs of splice junctions.

The intron starts with the dinucleotide GU and ends with the dinucleotide AG (therefore, the junctions are often described as conforming to the **GT-AG rule**). The two sites have different sequences and so they define the ends of the intron directionally. The 5′ splice site is called the left or **donor site**, the 3′ splice site is called the right or **acceptor site**.

In addition, a short sequence (including an invariant A) at the **branch site** is essential for the formation of a **lariat** structure during the splicing process (Figure 7).

Splicing apparatus

The 5′ and 3′ splice sites and the branch sequence are recognized by components of the splicing apparatus that assemble to form a large complex. The splicing apparatus contains both proteins and non-coding RNAs (in addition to the pre-mRNA).

The RNAs (termed small nuclear RNAs or **snRNAs**) exist as **ribonucleoprotein particles** (small nuclear ribonucleoproteins or **snurps**).

The snRNPs involved in splicing, together with many additional proteins, form a large complex called the **spliceosome**.

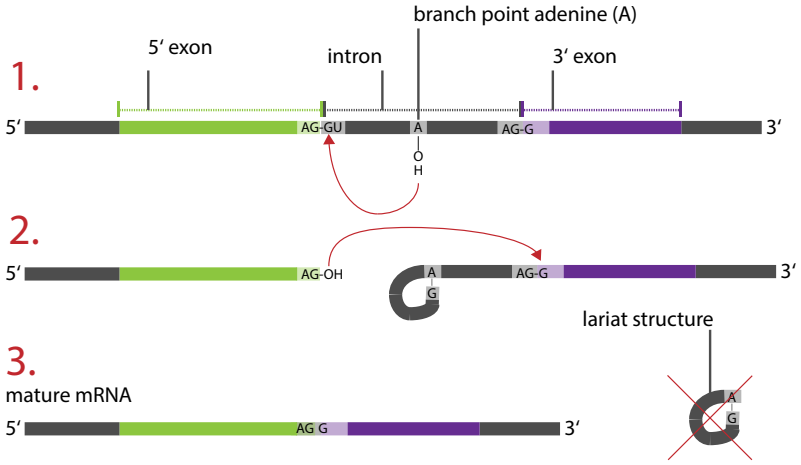


Figure 7. Splicing of eukaryotic pre-mRNA. Step 1: The branch point adenine (A) attacks the 3′ end of the 5′ exon. Step 2: The bond that forms the lariat structure goes from the 5′ position of the G that was at the 5′ end (donor site) of the intron to the 2′ position of the invariant A in the branch site. The left (upstream or 5′) exon takes the form of a linear molecule. The right intron-exon molecule forms a lariat, in which the 5′ terminus generated at the end of the intron becomes linked by a 5′-2′ bond to the conserved A of the branch site. Step 3: The free 3′ OH group of the 5′ exon attacks the 5′ end of the downstream (3′) exon. The exons are joined and the intron is released as lariat. The lariat is rapidly degraded.

CC BY SA

Alternative splicing: increasing protein diversity

The number of proteins in the proteome is not equivalent to the number of genes, but can exceed it by orders of magnitude.

Alternative splicing can generate more transcripts from a single gene than the number of genes in an entire genome.

In addition, the intron-exon structure may have speeded up the evolution of new proteins, because the introns make genetic recombination between different genes more likely.

Different **alternative splicing mechanisms** exist.

- **Alternative donor or acceptor splice sites**: Exons can have multiple 5′ or 3′ splice sites that are alternatively used.
- **Exon skipping**: Single exons can reside between two constitutive exons such that the alternative exon is either included or skipped.
- **Multiple cassette exons (mutually exclusive exons)**: One of two exons is retained in mRNAs after splicing, but not both.
- **Intron retention**: Introns can be retained in the mRNA and become translated. This is distinguished from exon skipping because the retained sequence is not flanked by introns.

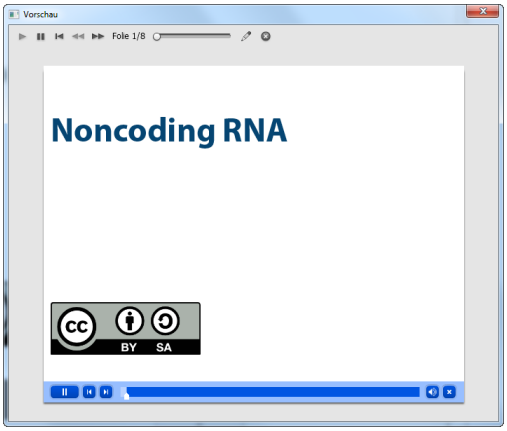
For more information on splicing see:

B. Alberts, A. Johnson, J. Lewis et al. (2002) Post-transcription controls. In: Molecular Biology of the Cell. 4. edition. Garland Science. http://www.ncbi.nlm.nih.gov/books/NBK26890/#A1366

S. Clancy (2008) RNA Splicing: Introns, Exons and Spliceosome. Nature Education 1: 31

Noncoding RNA

Non-coding RNA is RNA that does not encode protein (Video 1).



Video 1. Noncoding RNA. Click to play - HTML5 or Flash (opens new window in your standard web browser).

tRNA

Transfer RNAs (tRNA) are small RNAs that are used to provide amino acids corresponding to each particular codon in mRNA. A tRNA has a sequence of 74-95 bases that folds into a **cloverleaf secondary structure** (Figure 8).

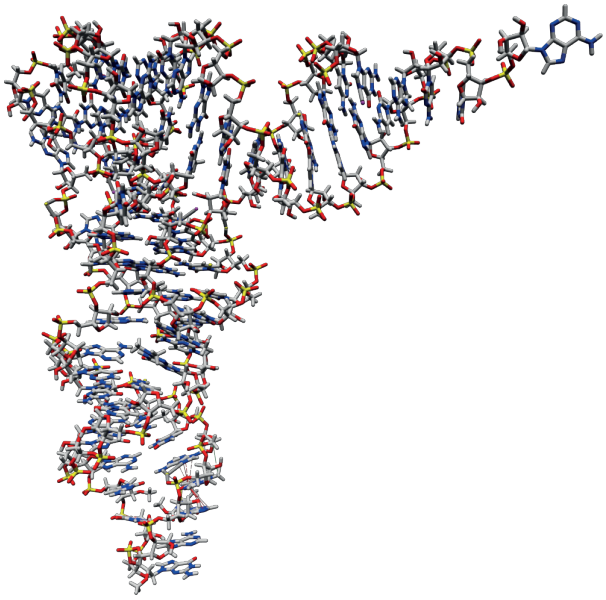


Figure 8. Yeast phenylalanine tRNA. The anticodon loop is at the bottom of the structure (drawn from 1EHZ.pdb with YASARA) (Shi and Moore, 2000).

There is at least one tRNA (usually more) for each amino acid. A tRNA represents a single amino acid, to which it is *covalently linked*. When a tRNA is charged with the amino acid corresponding to its anticodon, it is called **aminoacyl-tRNA**.

The amino acid is linked by an ester bond from its carboxyl group to the 2′ or 3′ hydroxyl group of the ribose of the 3′ terminal base of the tRNA (which is always adenine). The process of charging a tRNA is catalyzed by a specific enzyme, **aminoacyl-tRNA-synthetase**.

tRNA contains a trinucleotide sequence, the **anticodon**, which is complementary to the codon representing its amino acid. The anticodon enables the tRNA to recognize the codon via complementary base pairing.

rRNA

Ribosomal RNAs (rRNA) are components of the **ribosome**, a large ribonucleoprotein complex that contains many proteins as well as its RNA components, and which provides the apparatus for polymerizing amino acids into a polypeptide chain.

Ribosomal RNA is coded by a large number of identical genes that are tandemly repeated to form one or more clusters. Ribosomal RNA is the predominant product of transcription, constituting some 80-90% of the total mass of cellular RNA in both eukaryotes and prokaryotes.

Additional classes of ncRNA

Several additional classes of non-coding RNAs with essential functions in gene expression exist (Table 1).

snRNAs (spliceosomal RNAs): Interactions between pre-mRNA and snRNA are responsible for splice-site (U1) and branch-site (U2) recognition.

snoRNAs (small nucleolar RNAs): snoRNAs direct the site-specific modification of ribosomal RNAs and other ncRNAs in Eukaryota and Archaea. The RNAs act as guides, base pairing with complementary regions of the target RNA, while the catalytic function resides with proteins in the snoRNP complex.

miRNAs (microRNAs): miRNAs are short (~22 nt) sequences that inhibit the expression of target genes by binding to complementary regions of transcripts, triggering translational repression or transcript degradation.

The mature miRNA is excised from a stem loop called the precursor miRNA (pre-miRNA) by the ribonuclease enzyme DICER (see page 26).

Small regulatory RNAs (siRNAs): Includes miRNAs, Piwi-interacting RNAs, and small interfering or silencing RNAs. siRNAs are derived from double-stranded RNAs that use the RNA interference (RNAi) pathway to regulate a specific gene containing a complementary sequence.

Table 1. Non-coding RNAs.

Type of RNA	Function
rRNA	Part of a ribosome
snRNA	Involved in the processing (splicing) of mRNA
snoRNA	Involved in the modification of non-coding RNA
miRNA	Typically 21 or 22 bp in length, regulates gene expression
siRNA	21 to 25 bp in length, Includes miRNA and other small regulatory RNAs
tRNA	Transfers amino acids to the ribosome

For more details on the structure of RNAs see:

S. Clancy (2008) Chemical structure of RNA. Nature Education 7: 60

CC BY SA

Regulatory DNA sequences

RNA synthesis is catalyzed by a large enzyme called RNA polymerase. Transcription starts when RNA polymerase binds to a special region, the **promoter**, at the start of the gene.

The promoter surrounds the first base pair that is transcribed into RNA, the transcription **startpoint**. From this point, RNA polymerase moves along the template, synthesizing RNA, until it reaches a **terminator** sequence.

Sequences prior to the startpoint are described as **upstream** of it; those after the startpoint (within the transcribed sequence) are **downstream** of it (Figure 9).

Sequences are conventionally written so that transcription proceeds from left (upstream) to right (downstream). This corresponds to writing the mRNA in the usual 5'→3' direction.

Base positions are numbered in both directions away from the startpoint, which is assigned the value **+1**. Numbers are increased going downstream, and the negative numbers increase going upstream. There is no base assigned the number 0.

RNA synthesis, like nearly all biological polymerisation reactions, takes place in three stages: initiation, elongation, and termination.

The **initiation** of transcription is a critical process because here the cell selects which RNAs and proteins are synthesized.

To begin transcription, RNA polymerase must be able to recognize the startpoint. A DNA sequence termed promoter contains specific nucleotide sequences which indicate the starting point for RNA synthesis. Promoters are sequences of DNA that direct the RNA polymerase to the proper initiation site for transcription. This is a classic example of a *cis*-acting site.

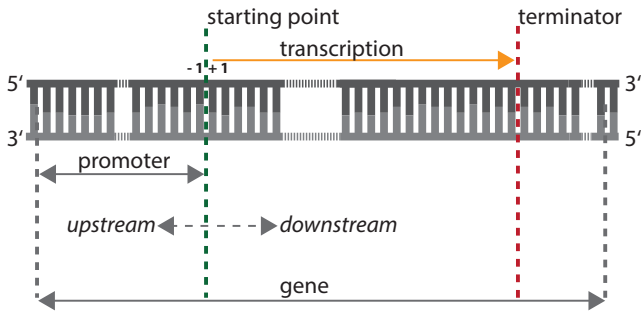


Figure 9. Transcription - terms and definitions.

One DNA strand is the template. **Elongation** than occurs until the RNA polymerase encounters the **terminator** (stop) site. At the terminator site, the DNA template and RNA are released.

Prokaryotic transcription

Transcription starts at promoters on the DNA template. In prokaryotes, the promoter consists of two short sequences at -10 and -35 positions upstream from the transcription start site. These two common motifs are known as the -10 sequence and the -35 sequence. These sequences are each 6 bp long.

The sequence at -10 is also called the **Pribnow box**, or -10 element, and usually consists of the six nucleotides TATAAT. The Pribnow box is absolutely essential to start transcription in prokaryotes. The other sequence at -35 (the **-35 element**) usually consists of the six nucleotides TTGACA. Its presence allows a very high transcription rate.

δ subunit

Transcription starts when RNA polymerase binds to a special region, the promoter. The complete **RNA polymerase holoenzyme** includes a δ subunit which is essential for initiation at the correct start site.

The δ subunit contributes to specific initiation in two ways. First, it *decreases* the affinity of RNA polymerase for general

CC BY SA

regions of DNA by a factor of 10^4. In its absence, the core enzyme binds DNA indiscriminately and tightly.

The holoenzyme binds to duplex DNA and moves along the double helix in search of a promoter, forming transient hydrogen bonds (Figure 10).

The search is rapid because RNA polymerase slides along DNA instead of repeatedly binding and dissociating from it. In other words, the promoter site is encountered by a random walk in one dimension rather than in three dimensions.

Second, the δ subunit enables RNA polymerase to recognize promotes sites. The holoenzyme binds to promoters and confers ability to recognize specific **binding sites**.

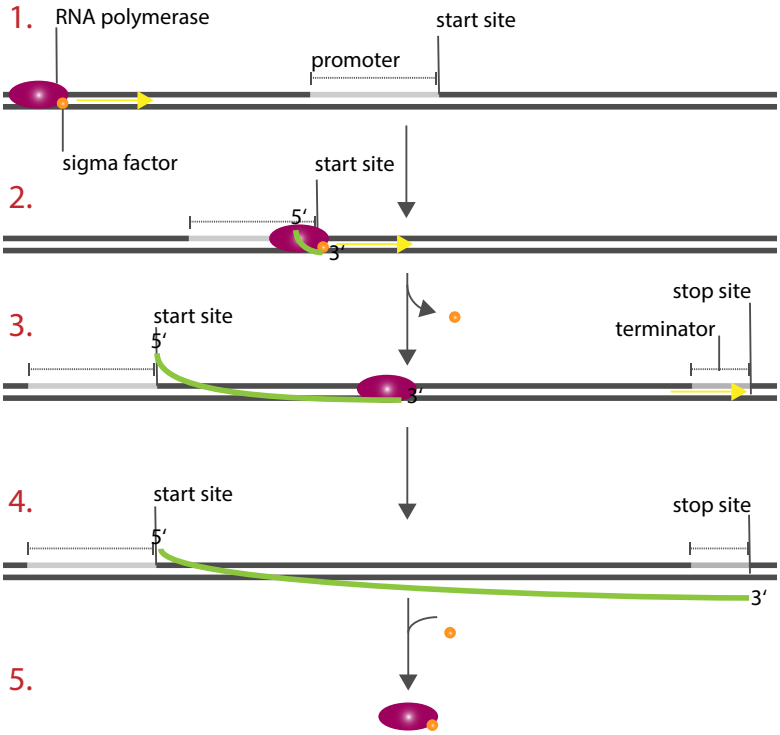


Figure 10. Bacterial transcription. 1. Bacterial RNA polymerase contains a sigma factor and slides along the double helix. 2. RNA polymerase reaches the promoter, binds tightly to the DNA, and initiates transcription (initiation). 3. The sigma factor is released and transcription continues (elongation). 4. RNA polymerase stops transcription at a termination signal (termination). 5. The new RNA strand and RNA polymerase are released. 5. Sigma factor rebinds RNA polymerase. The process begins again at another promoter.

The δ subunit is released when the nascent RNA chain reaches nine or ten nucleotides in length. After its release, it can assist initiation by another core enzyme. *E.coli* contains multiple δ factors to recognize several types of promoter sequences.

Initiation

Initiation begins with the binding of RNA polymerase to the double-stranded DNA at a promoter. In contrast with DNA synthesis, RNA synthesis can start *de novo*, without the requirement for a primer.

Then the strands of DNA are separated to form the **open complex** that makes the template strand available for base pairing with ribonucleotides. The δ factor is involved in the melting reaction. Initiation is accomplished if and when the enzyme manages to move along the template.

Elongation

During elongation the RNA polymerase moves along the DNA and extends the growing RNA chain. The elongation phase of RNA synthesis begins after the formation of the first phosphodiester bond.

Nucleotides are covalently added to the 3′ end of the growing RNA chain. An important change is the loss of δ (recall the core enzyme without δ binds more strongly to the DNA template). RNA polymerase stays bound to its template until a termination signal is reached.

Termination

The termination of transcription is as precisely controlled as its initiation. Termination involves the recognition of the point at which no further bases should

CC BY SA

be added to the chain.

When the last base is added to the RNA chain, the transcription bubble collapses as the RNA-DNA hybrid is disrupted, the DNA reforms in duplex state, and the RNA polymerase and RNA are both released.

The sequence of DNA required for these reactions defines the **terminator**. Terminators are distinguished in *E. coli* according to whether RNA polymerase requires any additional factors to terminate *in vitro*.

- RNA polymerase can terminate *in vitro* at certain sites in the absence of any other factor. These sites are called **intrinsic terminators**.
- Rho-dependent terminators are defined by the need for addition of **rho (ρ) factor** *in vitro*.

For more information on bacterial transcription see:

B. Alberts, A. Johnson, J. Lewis et al. (2002) From DNA to RNA. In: Molecular Biology of the Cell. 4. edition. Garland Science. http://www.ncbi.nlm.nih.gov/books/NBK26887/

S. Clancy (2008) DNA transcription. Nature Education 1: 41

H. Lodish, A. Berk, S.L. Zipursky et al. (2000) Transcription termination. In: Molecular Cell Biology. 4. edition. W.H. Freeman. http://www.ncbi.nlm.nih.gov/books/NBK21601/

Eukaryotic transcription

In eukaryotes, initiation of transcription requires the enzyme RNA polymerase and transcription factors. Any protein that is needed for the initiation of transcription but which is not itself part of RNA polymerase is defined as a **transcription factor**.

A significant difference between transcription of eukaryotic and prokaryotic mRNAs is that initiation at eukaryotic promoters involves a large number of factors that bind to a variety of *cis*-acting elements. The promoter contains all these binding sites.

Transcription factors are needed for initiation, but are not required subsequently. For the eukaryotic RNA polymerases, the transcription factors, rather than the RNA polymerase itself, are principally responsible for recognizing the promoter. For all eukaryotic RNA polymerases, the transcription factors create a structure at the promoter to provide the target that is recognized by the enzyme.

This is different from bacterial RNA polymerase, where it is the polymerase and sigma factor that recognize the promoter sequences.

Transcription in eukaryotic cells is divided into three classes. Each class is transcribed by a different RNA polymerase. **RNA polymerase I** transcribes rRNA. RNA polymerase **II** transcribes mRNA (Kornberg, 2007) (Figure 11). RNA polymerase **III** transcribes tRNA and other small RNAs.

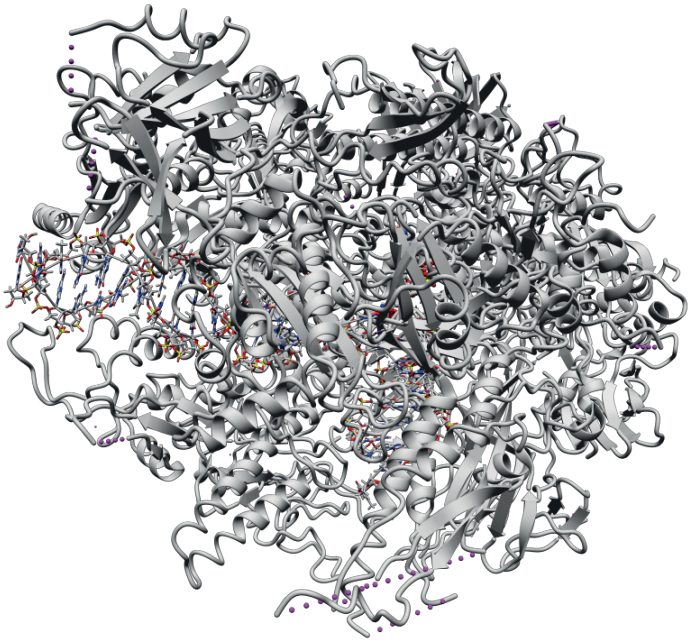


Figure 11. Yeast RNA polymerase II transcribing complex. RNA polymerase II (ribbon model) forms specific structures that bind DNA (coloured double helix) and mediate recognition of correct nucleoside triphosphates (NTPs) and formation of phosphodiester bonds (drawn from 2E2I.pdb with YASARA; Wang et al., 2006).

CC BY SA

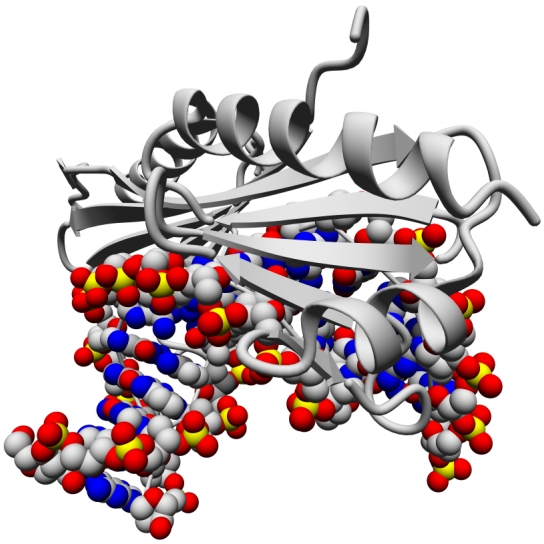


Figure 12. Human TBP (grey) binds TATA-element. During transcription initiation, TBP recognizes the TATA box in a directional manner, unwinds DNA and attracts the transcription machinery (drawn from 1CDW.pdb with YASARA) (Nikolov et al., 1996).

For more details on eukaryotic transcription see:

J.M. Berg, J.L. Tymoczko, L. Stryer (2002) Eukaryotic transcription and translation are separated in space and time. In: Biochemistry, 5th edition, W.H. Freeman. http://www.ncbi.nlm.nih.gov/books/NBK22433/

G.M. Cooper (2000) Eukaryotic RNA polymerases and general transcription factors. In: The Cell: A Molecular Approach. 2nd edition. Sinauer Accociates. http://www.ncbi.nlm.nih.gov/books/NBK9935/

Scitable by Nature Education (2014) The information in DNA is decoded by transcription.

Eukaryotic promoters

Promoters utilized by RNA polymerase II have a **modular organization**. Short sequence elements that are recognized by transcription factors lie upstream of the startpoint.

Some examples are the following elements: TATA, GC boxes, CAAT boxes, and the octamer (an 8 bp element). Some of these elements and the factors that recognize them are common. They are found in a variety of promoters and are used constitutively.

Others are specific: they identify particular classes of genes and their use is regulated. The elements occur in different combinations in individual promoters.

Initiation requires the transcription factors to act in a defined order to build a complex that is joined by RNA polymerase. The eukaryotic RNA polymerase II can bind to the promoter only after separate transcription factors have bound. The basal elements of the promoter (the **TATA box** and **Inr**) primarily determine the location of the startpoint.

The **basal apparatus** (several transcription factors) assembles at the promoter. In TATA-box promoters, the initial event is binding of the TATA-binding protein (TBP) to the TATA-box (Figure 12). Other transcription factors bind to the complex in a defined order. When RNA polymerase II binds to the complex, it initiates transcription (Figure 13).

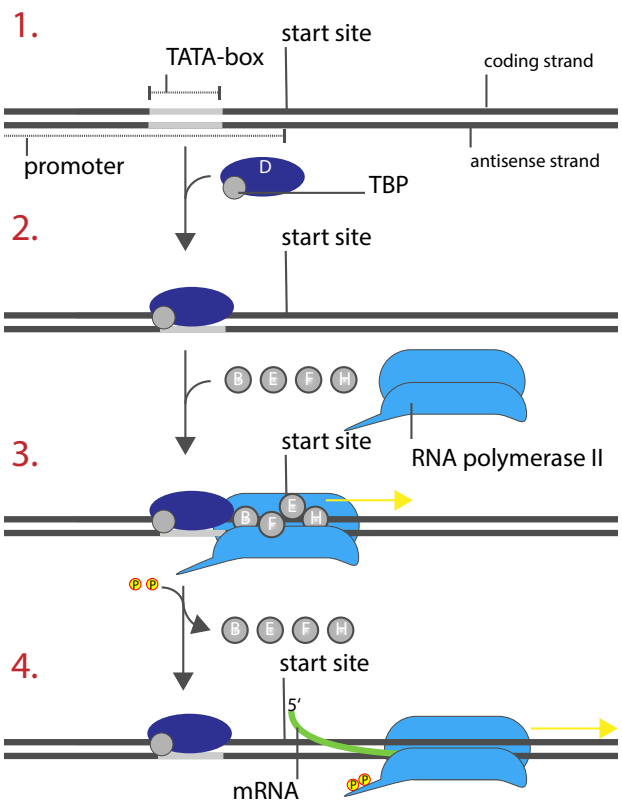


Figure 13. Initiation of eukaryotic transcription. 1. Promoters are located on the 5' side of the transcription start site. 2. The TATA box is recognized by the TATA-box binding protein, which is bound to the transcription factor TFIID. 3. Additional general transcription factors (TFIIB, E, F, and H) and RNA polymerase II are attracted to the promoter. 4. TFIIH opens the DNA double helix and phosphorylates the carboxyl-terminal domain of RNA polymerase II. After the release of general transcription factors, transcription begins.

CC BY SA

For more details on promoters and tools to analyze them see:

Weizmann Institute of Science (2014) Toolbox>-Sequence Analysis>Promoters

Enhancers

We have considered the promoter as an isolated region responsible for binding RNA polymerase. But eukaryotic promoters do not necessarily function alone. In some cases, the activity of a promoter is enormously increased by the presence of an enhancer, which consists of another group of elements, but is located at a variable distance from the promoter.

Enhancers work much like promoters, except for their ability to function in either orientation and at variable distances from the startpoint. The position of the enhancer relative to the promoter can vary substantially.

It can be either upstream or downstream. An enhancer can stimulate any promoter placed in its vicinity.

Enhancers can be located within genes. However, enhancers work only in *cis* configuration (on the same DNA molecule) with a target promoter. The organization of DNA must be flexible enough to allow the enhancer and promoter to be closely located. This requires the intervening DNA to be extruded as a large loop.

Elements analogous to enhancers, called **upstream activator sequences** (UAS), are found in yeast.

For more details on enhancers see:

B. Alberts, A. Johnson, J. Lewis (2002) How genetic switches work. Molecular Biology of the cell, 4th edition, Garland Science: http://www.ncbi.nlm.nih.gov/books/NBK26872/#A1283

J.M. Berg, J.L. Tymoczko, L. Stryer (2002) Enhancers. Biochemistry, 5th edition, W.H. Freeman: http://www.ncbi.nlm.nih.gov/books/NBK22433/#A3994

RNA transport

Since all RNA is synthesized in the nucleus, the entire cytoplasmic complement of RNA (mRNA, rRNA, tRNA, and other small RNAs) must be derived by export from the nucleus.

mRNA is transported as a ribonucleoprotein (mRNP), which forms on the RNA transcript in the nucleus.

The nuclear pores are used for both import and export of material. All pores are identical. Since all proteins are synthesized in the cytosol, any proteins required in the nucleus must be transported there.

The **nuclear pore complex** (NPC) is a very large structure that extends through the nuclear envelope, providing a channel for bidirectional transport of molecules between the nucleus and the cytosol.

For more details on RNA export from the nucleus to the cytoplasm see:

A. Köhler and E. Hurt (2007) Exporting RNA from the nucleus to the cytoplasm. Nature Rev. Mol. Cell Biol. 8: 761-773

Further Readings

References

B. Alberts, D. Bray, K. Hopkin, A. Johnson, J. Lewis, M. Raff, K. Roberts and P. Walter (2009) Essential Cell Biology, 3rd edition. Garland Science

H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, and P.E. Bourne (2000) The Protein Data Bank. Nucleic Acids Res. 28: 235-242

T.A. Brown (2002) Genomes. 2nd edition. Garland Science. http://www.ncbi.nlm.nih.gov/books/NBK21128/

R.D. Kornberg (2007) The molecular basis of eukaryotic transcription. Proc. Natl. Acad. Sci. USA 104: 12955-12961

E. Krieger and G. Vriend (2014) YASARA View - molecular graphics for all devices - from smartphones to workstations. Bioinformatics doi:10.1093/bioinformatics/

CC BY SA

btu426

J. Krebs, E.S. Goldstein and S.T. Kilpatrick (2012) Lewin's GENES XI. 11th edition

Jones & Bartlett LearningD.B. Nikolov, H. Chen, E.D. Helay, A. Hoffmann, R.G. Roeder and S.K. Burley (1996) Crystal structure of a human TATA box-binding protein/TATA element complex. Proc. Natl. Acad. Sci. USA 93: 4862-4867

H. Shi and P.B. Moore (2000) The crystal structure of yeast phenylalanine tRNA at 1.93 A resolution: a classic structure revisited. RNA 6: 10091-1105

D. Wang, D.A. Bushnell, K.D. Westover, C.D. Kaplan, R.D. Kornberg (2006) Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis. Cell 127: 941-954

OER and OCW

S. Clancy (2008) DNA transcription. Nature Education 1: 41

L. Hoopes (editor) (2014) Gene expression and regulation. Scitable, Nature Education

E. Lander, R. Weinberg, T. Jacks, H. Sive, G. Walker, S. Chisholm and M. Mischke (2011) Fundamentals of Biology – Molecular Biology. MIT OCW

North Dakota State University (2014) Introducing: Transcription. Virtual Cell Animation Collection

L.A. Pray (2008) What is a gene? Colinearity and transcription units. Nature Education 1: 97

OpenLearn (2009) DNA, RNA and protein formation. The Open University

CC BY SA

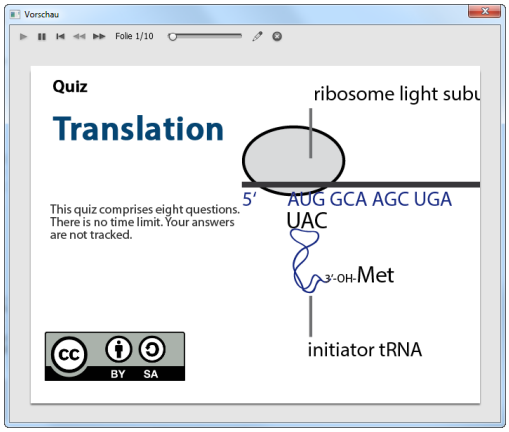
Translation

In translation, mRNA is decoded to produce a specific polypeptide according to the rules specified by the genetic code. Protein synthesis falls into the three stages: initiation, elongation, and termination.

The ribosome provides the environment for controlling the interaction between a codon of mRNA and the anticodon of aminoacyl-tRNA.

Quiz

Start this chapter with the quiz to test for your mastery. You may reattempt the quiz later to check your progress.



Quiz 1. Translation. Click to play - HTML5 or Flash (opens new window in your standard web browser).

Overview

Translation is the process that converts an mRNA sequence into a chain of amino acids that form a protein. A mRNA contains a series of codons (nucleotide triplets) that interact with the anticodons of aminoacyl-tRNAs so that a corresponding series of amino acids is incorporated into a polypeptide chain.

Protein synthesis falls into the three stages (Figure 1):

1. **Initiation** involves the reactions that precede formation of the first peptide bond between the first two amino acids of the protein.

2. **Elongation** includes all the reactions from formation of the first peptide bond to addition of the last amino acid. Amino acids are added to the chain one at a time. The ribosome moves along mRNA (in a process termed translocation) and the length of the protein chain extends.

3. **Termination** encompasses the steps that are needed to release the completed polypeptide chain; at the same time, the ribosome dissociates from the mRNA.

Initiation

Bacteria

Initiation in bacteria needs the ribosomal 30S subunit and accessory factors. A 30 S subunit carrying initiation factors binds to an initiation site on mRNA (ribosome-binding site) to form an initiation complex.

Ribosome-binding sites include the upstream **Shine-Dalgarno sequence** and the initiation codon. The rRNA of the 30 S subunit has a complementary sequence that base pairs with the Shine-Dalgarno sequence during initiation.

Then a large ribosomal subunit (50 S) joins the complex to generate a complete ribosome.

Protein synthesis starts with a methionine amino acid encoded by **AUG**.

Different methionine tRNAs are involved in initiation and elongation. The **initiator tRNA** has unique structural features that distinguish it from all other tRNAs.

Eukaryotes

Eukaryotic protein synthesis differs from prokaryotic protein synthesis primarily in translation initiation. As in prokaryotes, a special tRNA participates in initiation. The initiating codon in eukaryotes is always AUG.

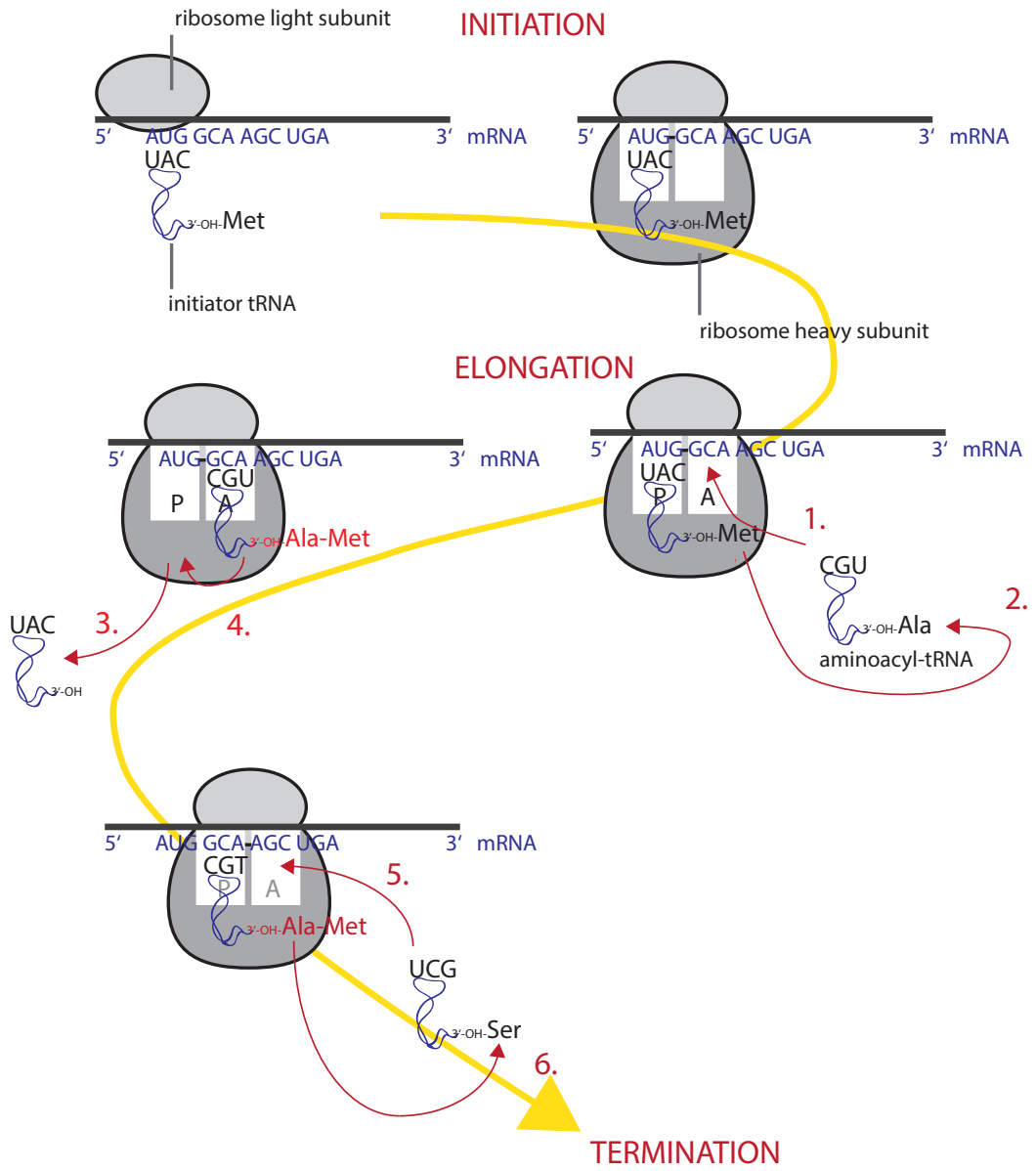


Figure 1. Translation. INITIATION: Initiator tRNA (loaded with Met) binds to startcodon (AUG). Ribosome assembles. ELONGATION: 1. Incoming aminoacyl-tRNA (exemplarily loaded with Ala) binds to A site. 2. Transfer of Met from tRNA in P site to Ala of tRNA in A site (peptide bond formation). 3. Initiator tRNA (or later another `empty` tRNA) exists ribosome. 4. tRNA at A site (loaded with Met-Ala) moves to P site. 5. Next aminoacyl-tRNA (exemplarily loaded with Lys) enters A site. 6. Met-Ala peptide is transferred from tRNA at P site to tRNA at A site, second peptide bond forms. This process continues until a termination codon (i.e., UAG, UAA, or UGA) is recognized by protein release factors. TERMINATION: The tRNA, mRNA, and ribosome dissociate.

However, the order of events is different, and the number of accessory factors is greater.

- In eukaryotes, small subunits (40 S) first recognize the 5′ end (the cap) of the mRNA, and then move to the initiation site, where they are joined by large subunits. In prokaryotes, small subunits bind directly to the initiation site.
- Binding of 40S subunits to mRNA requires several initiation factors, including proteins that recognize the structure of the cap.
- The presence of poly(A) on the 3′ tail of an mRNA stimulates the formation of an initiation complex at the 5′ end.

Ribosomes

The ribosome provides the environment for controlling the interaction between a codon of mRNA and the anticodon of aminoacyl-tRNA.

Ribosomes are traditionally described

CC BY SA

in terms of their rate of sedimentation (measured in **Svedbergs**, in which a higher S value indicates a greater rate of sedimentation and a larger mass). Bacterial (70S) ribosomes have **subunits** that sediment at 50S and 30S. The subunits of eukaryotic cytoplasmic (80S) ribosomes sediment at 60S and 40S.

Each subunit has an RNA component and many proteins. The two subunits work together as part of the complete ribosome. The small subunit mediates the correct interactions between the anticodons of the tRNAs and the codons in the mRNA. The large subunit catalyzes the formation of peptide bonds.

All ribosomes of a given cell compartment are identical. An mRNA is simultaneously translated by several ribosomes. Each ribosome is at a different stage or progression along the mRNA.

A mRNA that is simultaneously being translated by several ribosomes is termed polyribosome or **polysome**.

A, P, and E sites of ribosomes and translocation

A protein is assembled by the sequential addition of amino acids in the direction from the N-terminus to the C-terminus as a ribosome moves along the mRNA.

- A ribosome begins translation at the 5′ end of a coding region (initiation at specific **AUG start codon**).
- It translates each triplet codon into an amino acid as it proceeds towards the 3′ end (elongation).
- At each codon, the appropriate amino-acyl-tRNA associates with the ribosome, donating its amino acid to the polypeptide chain.
- At any given moment, the ribosome can accommodate the two aminoacyl-tRNAs corresponding to successive codons, making it possible for a peptide bond to form between the corresponding amino acids.

Any amino-acyl-tRNA except the initiator tRNA can enter the **A site** (Aminoacyl-tRNA site) (Figure 2). Its entry is mediated by an elongation factor. Aminoacyl-tRNA is loaded into the A site in two stages. First the anticodon end binds to the A site of the 30S subunit. Then codon-anticodon recognition triggers a change in the conformation of the ribosome.

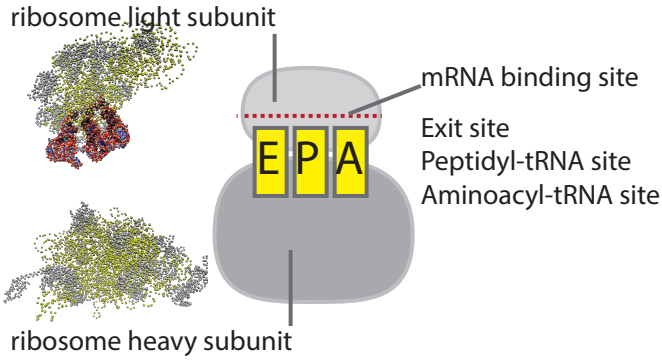


Figure 2. Ribosome structure. Ribosomes contain a small subunit and a large subunit. Each ribosome has a mRNA binding site and three tRNA binding sites. The three-dimensional structures show the small (30 S) and large (50 S) subunit of the *Thermus thermophilus* ribosome. The small subunit contains three tRNA molecules (drawn from 1GIX.pdb and 1GIY.pdb).

The **P** (Peptidyl-tRNA site) is occupied by peptidyl-tRNA, a tRNA carrying the nascent polypeptide chain. Peptide bond formation occurs when the polypeptide carried by the peptidyl-tRNA is transferred to the amino acid carried by the aminoacyl-tRNA.

Peptide bond synthesis generates deacylated tRNA in the P site and peptidyl-tRNA in the A site. The activity responsible for synthesis of the peptide bond is called **peptidyl transferase**. As mentioned above, peptidyl transferase acivity is a function of the large (50S or 60S) ribosome subunit.

The deacylated tRNA, lacking an amino acid, lies in the P site, while a new peptidyl-tRNA has been created in the A site. This peptidyl-tRNA is one amino acid longer than the peptidyl-tRNA that had been in the P site.

The ribosome moves one triplet (three bases) along the messenger. This stage is

CC BY SA

called **translocation**. Translocation moves deacylated tRNA into the **E** site (Exit site), peptidyl-tRNA into the P-site, and empties the A site.

- The movement transfers the deacylated tRNA out of the P site, the deacylated tRNA leaves the ribosome via the E site.
- The peptidyl-tRNA empties the A site and moves into the P site.
- The next codon to be translated now lies in the A site, ready for a new aminoacyl-tRNA to enter.
- A new aminoacyl-tRNA enters the A site. It receives the polypeptide chain from peptidyl-tRNA (peptide bond formation occurs), and is transferred into the P site for the next cycle of elongation.

The flow of tRNA is into the A site, through the P site, and out through the E site.

For more details on ribosomes see:

J.M Berg, J.L. Tymoczko, L. Stryer (2002) A ribosome is a ribonucleoprotein particle (70S) made of a small (30S) and a large (50S) subunit. Biochemistry, 5th edition. W.H. Freeman. http://www.ncbi.nlm.nih.gov/books/NBK22335/

T.A. Brown (2002) Synthesis and processing of the proteome. Genomes, 2nd edition. http://www.ncbi.nlm.nih.gov/books/NBK21111/

T.A. Steitz (2008) A structural understanding of the dynamic ribosome machine. Nature Reviews Molecular Cell Biology 9: 242-253

Termination

Only 61 triplets are assigned to amino acids. The other three triplets are **termination codons** (or stop codons) that end protein synthesis (see below).

The UAG, UAA, and UGA triplet sequences are necessary and sufficient to end protein synthesis.

Termination codons are recognized by protein release factors (not by aminoacyl-tRNAs). None of the termination codons is represented by a tRNA. They function in an entirely different manner from other codons, and are recognized directly by protein factors.

The termination reaction involves release of the protein chain from the last RNA. Subsequently, the tRNA and mRNA are released, and the ribosome dissociates into its subunits.

Genetic code

In translation, messenger RNA (mRNA) is decoded to produce a specific polypeptide according to the rules specified by the genetic code. The genetic code is the correspondence between triplets in DNA (or RNA) and amino acids in protein.

The sequence of a coding strand of DNA, read in the direction from 5′ to 3′, consists of nucleotide triplets (codons) corresponding to the amino acid sequence of a protein read from N-terminus to C-terminus (Figure 3).

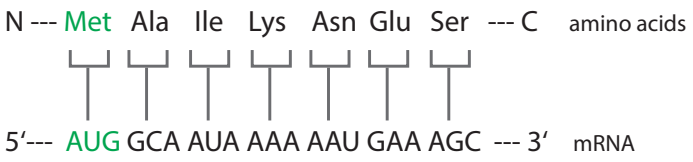


Figure 3. Three nucleotides encode an amino acid.

The universality of the genetic code indicates that it must have been established very early in evolution.

There are **64 codons** (each of 4 possible nucleotides can occupy each of the 3 positions of the codon, making $4^3 = 64$ possible trinucleotide sequences). Each of these codons has a specific meaning in protein synthesis: 61 codons represent amino acids.

Because there are more codons (61) than there are amino acids (20), almost all amino acids are represented by more than one codon. The only exceptions are methionine and tryptophan.

The **initiation codon** is a special codon (usually AUG) used to start synthesis of a protein.

CC BY SA

A **stop codon** (termination codon) is one of three triplets (UAG, UAA, UGA) that causes protein synthesis to terminate.

Codons representing the same or related amino acids tend to be similar in sequence. Often the base in the third position of a codon is not significant, because the four codons differing only in the third base represent the same amino acid.

For more details on the genetic code see:

J.M. Berg, J.L. Tymoczko, L. Stryer (2002) Amino acids are encoded by groups of three bases starting from a fixed point. In: Biochemistry, 5th edition. W.H. Freeman. http://www.ncbi.nlm.nih.gov/books/NBK22358/

A.J.F. Griffiths, J.H. Miller, D.T. Suzuki et al. (2000) Genetic code. In: An Introduction to Genetic Analysis. 7th edition. W.H. Freeman. http://www.ncbi.nlm.nih.gov/books/NBK21950/#A1864

NCBI Learning Center (2012) Single letter codes for nucleotides

Mutation

A mutation is a change in a genetic sequence. The tendency for similar amino acids to be represented by related codons minimizes the effects of mutations. It increases the probability that a single random base change will result in no amino acid substitution (silent mutation) or in one involving amino acids of similar character.

For example, a mutation of CUC to CUG has no effect, since both codons represent leucine; and a mutation of CUU to AUU results in replacement of leucine with isoleucine, a closely related amino acid.

There are three possible ways of translating any nucleotide sequence into protein, depending on the starting point. These are called **reading frames**. Usually only one reading frame is translated and the other two are blocked by frequent termination signals (Figure 4).

But a mutation that inserts or deletes a single base will change the triplet sets for the entire subsequent sequence. A change of this sort is called a **frameshift**.

For more details on mutations see:

DNA from the beginning (2011) Mutations are changes in genetic information. Cold Spring Harbor Laboratory. http://www.dnaftb.org/27/

Scitable by Nature Education (2014) DNA is constantly changing through the process of mutation.

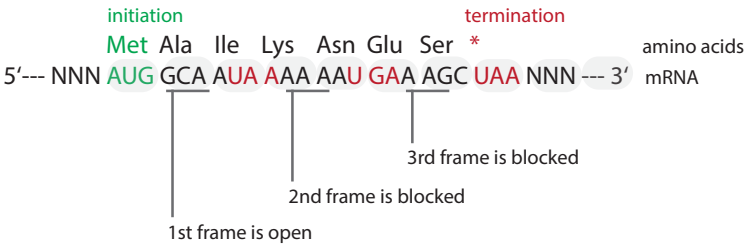


Figure 4. Only one reading frame (the open reading frame) is translated. The other two are blocked by frequent termination signals.

Wobble

Multiple codons that represent the same amino acid most often differ at the third base position (**degeneracy of the genetic code**). For example, a codon is said to be fourfold degenerate if any nucleotide at its third position specifies the same amino acid. The degeneracy of the genetic code is what accounts for the existence of silent mutations.

Often one tRNA can recognize more than one codon. This means that the base in the first position of the anticodon must be able to partner alternative bases in the corresponding third position of the codon. Base pairing at this position cannot be limited to the usual G-C and A-U partnerships.

The **wobble hypothesis** states that the pairing between codon and anticodon at the first two codon positions always follows the usual rules, but that exceptional wobbles occur at the third position.

Wobbling occurs because the conformation of the tRNA anticodon loop permits

CC BY SA

flexibility at the first base of the anti-codon.

Wobble reduces the number of tRNAs needed in a cell by enabling one tRNA to read two or three codons. For example, the human genome has 48 tRNAs, and 16 are predicted to decode two codons each. The remaining 32 tRNAs are specific for only one codon.

For more details on wobbling see:

Scitable by Nature Education (2014) Wobble may exist in the pairing of a codon on mRNA with an anticodon on tRNA

T.A. Brown (2002) Synthesis and Processing of the Proteome. In: Genomes, 2nd edition. http://www.ncbi.nlm.nih.gov/books/NBK21111/#A7603

Further Readings

References

B. Alberts, D. Bray, K. Hopkin, A. Johnson, J. Lewis, M. Raff, K. Roberts and P. Walter (2009) Essential Cell Biology, 3rd edition. Garland Science

H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, and P.E. Bourne (2000) The Protein Data Bank. Nucleic Acids Res. 28: 235-242

T.A. Brown (2002) Genomes. 2nd edition. Garland Science. http://www.ncbi.nlm.nih.gov/books/NBK21128/

R.D. Kornberg (2007) The molecular basis of eukaryotic transcription. Proc. Natl. Acad. Sci. USA 104: 12955-12961

J. Krebs, E.S. Goldstein and S.T. Kilpatrick (2012) Lewin's GENES XI. 11th edition

E. Krieger and G. Vriend (2014) YASARA View - molecular graphics for all devices - from smartphones to workstations. Bioinformatics. doi:10.1093/bioinformatics/btu426

OER and OCW

L. Hoopes (editor) (2014) Gene expression and regulation. Scitable, Nature Education

E. Lander, R. Weinberg, T. Jacks, H. Sive, G. Walker, S. Chisholm and M. Mischke (2011) Fundamentals of Biology – Molecular Biology. MIT OCW

OpenLearn (2009) DNA, RNA and protein formation. The Open University

CC BY SA

Control of gene expression

Gene expression is regulated at many steps in the pathway from DNA to protein: transcription (in particular initiation), mRNA processing (including alternative splicing), export of mature mRNA, translation (in particular initiation), and stability (e.g. microRNA-mediated mRNA degradation) (Figure 1).

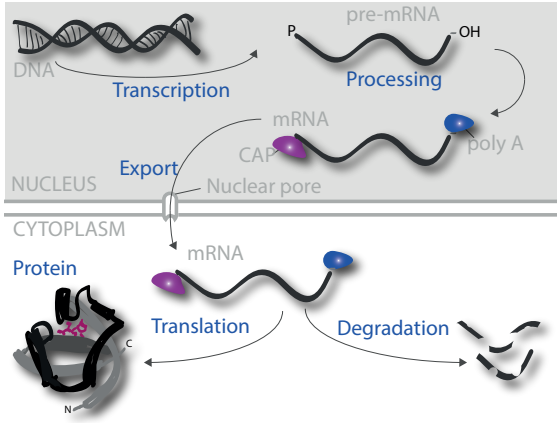
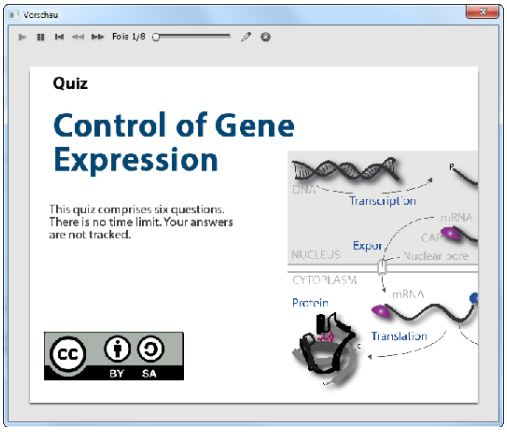


Figure 1. Gene expression is regulated at many steps in the pathway from DNA to protein.

Quiz

Start this chapter with the quiz to test for your mastery. You may reattempt the quiz later to check your progress.



Quiz 1. Control of gene expression. Click to play - HTML5 or Flash (opens new window in your standard web browser).

Overview

All steps involved in expressing a gene can be regulated. Different levels of control occur.

- **Initiation**: The most important point of control is the initiation of transcription. In eukaryotes, the local structure of chromatin determines accessibility of promoter regions.
Initiation of transcription requires that nucleosomes change their structure, becoming less compact. The **chromatin structure** is governed by the epigenetic code and proteins (transcriptional regulators) that read the code.

The **epigenetic code** comprises the histone code (i.e. covalent modifications of the core histones of the nucleosomes) and DNA methylation. Transcriptional regulators locally modify the chromatin structure of promoter regions.

In addition, transcription regulators can directly affect the assembly process of RNA polymerase and the general transcription factors at the promoter.

Since the bacterial chromosome is not packaged into chromatin, the control of initiation is simpler than in eukaryotes. Transcription regulators bind to regulatory DNA sequences (promoters) and activate or repress transcription.

- **Elongation**: Cells can also regulate gene expression by controlling events that occur after initiation of transcription.
- **Processessing** (in eukaryotic cells): The processes of capping, splicing, polyadenylation, and export are interconnected and controlled.
- **Posttranscriptional control**: The stability and translation efficiency of transcripts is controlled. For example, microRNAs (miRNAs) regulate the stability of many mRNAs and their translation.

BY SA

Control of transcription

Initiation: Promoters and enhancers

The **prokaryotic promoter** consists of short sequences upstream from the transcription start site (Figure 2). Initiation begins with the binding of the RNA polymerase to double-stranded DNA at a promoter. The strands of DNA are separated to form the "open complex" that makes the template strand available for base pairing with ribonucleotides.

Prokaryotic promoters differ in their efficacy. Strong promoters (with efficient initiation of transcription) have **-10 region** (Pribnow box) and -35 region (**-35 element**) sequences that correspond closely to consensus sequences, whereas weak promoters (with low efficiency of transcription) have base substitutions at these sites.

An example for bacterial control of gene expression is the *lac* operon (see page 4) in which *E. coli* cells turn on a suite of genes that metabolize lactose when this sugar becomes available.

All **eukaryotic cells** of an organism contain exactly the same DNA which encodes all the RNA and protein molecules that are needed. However, a typical eukaryotic cell expresses only a fraction of its genes, because cells differentiate into different cell types in different organs.

Furthermore, cells exposed to hormones, growth factors and other stimuli will change substantially in shape, size, metabolic pathways, and other characteristics. The diverse properties are the result of differences in gene expression.

In eukaryotic cells, the site of transcription (the nucleus) is separated from the site of translation (the cytoplasm). The RNA molecules move to the cytoplasm, where they can be bound by ribosomes. Translation is necessarily preceded by transcription.

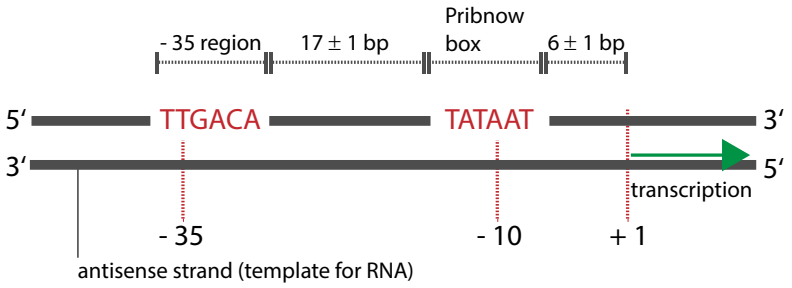


Figure 2. Promoter sites for transcription in prokaryotes. The first nucleotide to be transcribed is numbered + 1. The **Pribnow box** in bacteria has a function similar to the TATA box in eukaryotes.

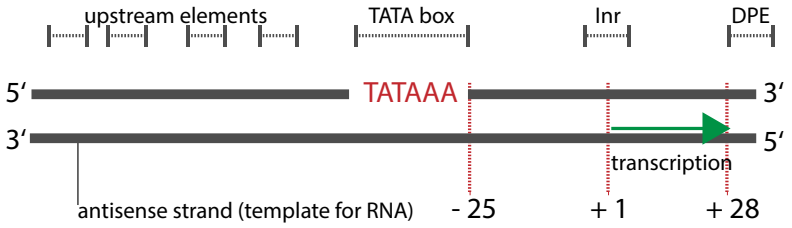


Figure 3. Promoter sites for RNA polymerase II (protein-coding genes) in eukaryotes. TATA-box: Component of many but not all core promoters. Bound by the TATA box-binding protein (TBP) subunit of the TFIID complex. **Inr**: Initiator motif encompasses the transcription start site. **DPE**: Downstream core promoter element, found in TATA-less promoters (although some promoters contain both DPE and TATA motifs) (Juven-Gershon et al., 2008). In addition, various upstream promoter elements control gene specifc expression of genes.

Eukaryotic promoters utilized by RNA polymerase II have a modular organization (Figure 3). Short sequence elements that are recognized by transcription factors lie upstream of the transcription startpoint.

Some of the promoter elements are found in many promoters and are used constitutively. Other elements are specific and identify particular genes which are active only in specific tissues or under specific metabolic and environmental conditions. Promoter elements can occur in different combinations in individual promoters.

The **core promoter** defines the shortest sequence at which RNA polymerase II can initiate transcription. It comprises the minimum sequence (of about 40 nt)

CC BY SA

that enables the general transcription factors to assemble at the transcription startpoint. Several transcription factors act in a defined order to build a complex that is joined by RNA polymerase. The transcription factors create a local chromatin structure at the promoter that is recognized by RNA polymerases (Figure 4).

However, a core promoter functions at low transcription efficiency. Other proteins called **activators** are required for higher transcription efficiency.

In some cases, the activity of a promoter is enormously increased by the presence of an **enhancer**, which consists of another group of elements, but is located at a variable distance from the promoter. Enhancers work much like promoters, except for their ability to function in either orientation and at variable distances from the transcription startpoint.

Transcription in the nucleus is concentrated in structures termed **transcription factories**. Thus, nascent RNAs and RNA polymerase II are not equally distributed in the nucleus, but are concentrated in specific nuclear areas. This favours a more economical utilization of the transcription machinery. The total number of transcription factories in nuclei of human (HeLa) cells is 2000-4000.

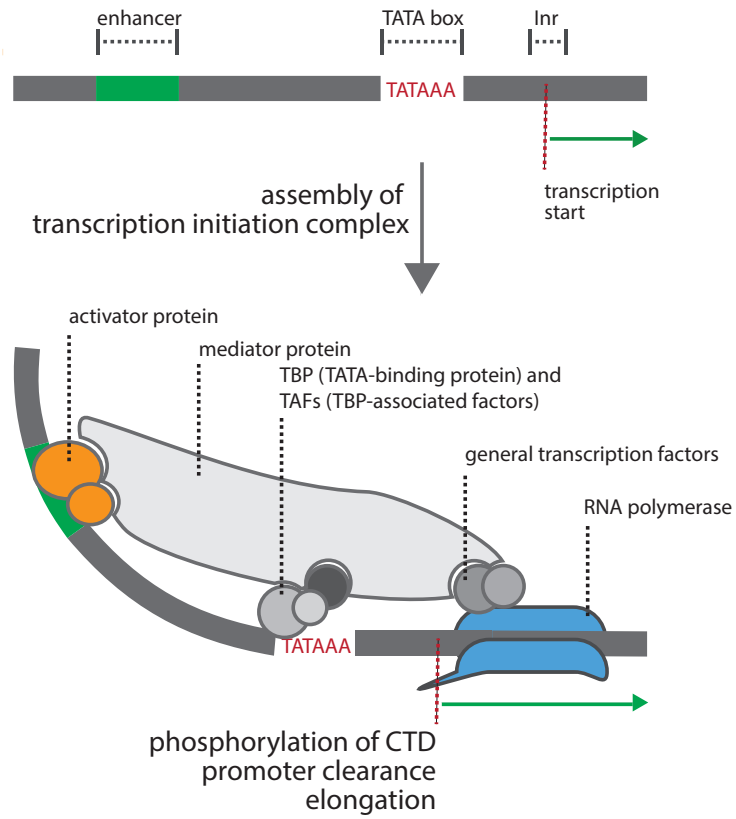


Figure 4. Initiation of transcription in eukaryotes. Transcriptional regulators such as activators and mediators aid in the assembly of the **general transcription factors** (GTFs, including TBP and TAFs) and **RNA polymerase II** at the promoter. Additional transcriptional regulators make changes in chromatin or attract other proteins that modulate chromatin structure. Many of the GTFs don't bind DNA but are part of the large initiation complex that interacts with RNA polymerase II. After the assembly of the transcription complex, the **C-terminal domain** of RNA polymerase II becomes phosphorylated. Contact with initiation factors is lost and stable association with the nascent transcript is established in a process termed **promoter clearance** and elongation takes place.

Chromatin remodelling

DNA is packed into nucleosomes which again are arranged in higher-order structures. Initiation of transcription in eukaryotic cells depends on the **chromatin structure**. The chromatin structure depends on covalent modifications of histones and DNA methylation.

For more details on chromosome structure and gene expression see:

C.L. Woodcock and R.P. Ghosh (2010) Chromatin higher-order structure and dynamics. Cold Spring Harb. Perspect. Biol. 2: a000596

Chromatin remodelling and gene expression (MP3 or OGG audio opens in your standard brower)

The histone modifications and DNA methylation represent the **epigenetic code** (see definitions in Table 1).

The histone code and DNA methylation

Histones are integral and dynamic components of the machinery responsible for regulating transcription. The position of nucleosomes and their components (in terms of covalent modifications and histone variants) are highly dynamic.

CC BY SA

Table 1. Definitions.

Term	Definition
Epigenetics	Study of heritable changes in genome function that occur without a change in DNA sequence.
Epigenome	Refers to the entirety of epigenetic features possessed by an organism.
Epigenetic code	DNA methylation and histone modifications (histone code) represent the epigenetic code. The epigenetic modifications do not change the DNA code itself, but rather, influence the availability of the code to the factors that read it and translate it into its product.
Epiallele (epigenetic allele)	An allele that shows heritable difference in expression as a consequence of epigenetic modifications (and not changes in DNA sequence).
Paramutation	Allelic interactions that cause a meiotically heritable change in the expression of one of the alleles.

The **histone modifications** comprise the attachment of modifying groups to amino acid residues of core histone molecules, especially those amino acids located at their amino (N)-terminal tails. Several types of covalent modifications are known, for example:

- Acetyl, methyl-, ubiquityl-, and sumoyl groups to lysine (Figure 3),
- methyl-group to arginine,
- phosphoryl-group to serine or threonine, and
- ADP ribosyl-group to glutamine.

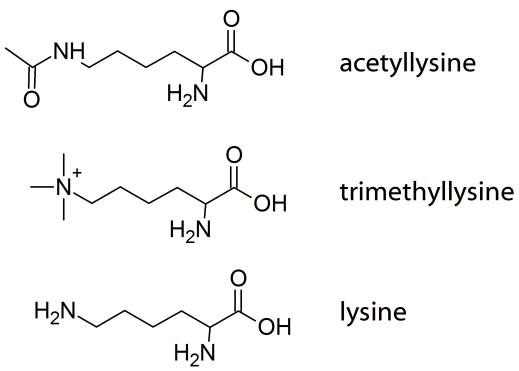


Figure 3. Molecular structure of acetyl and trimethyl modification of lysine.

Histones can be covalently modified at different positions simultaneously. Information arises from combination of marks (therefore histone code).

Histone modifications are associated with either active or inactive chromatin states, as well as with particular cellular processes such as mitosis and DNA repair (Figure 4).

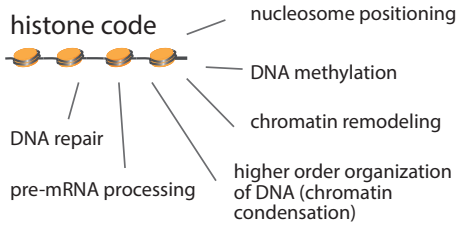


Figure 4. Factors that interact with modified histones to establish the epigenetic regulation of transcription and other processes.

The histone marks can act synergistically or antagonistically. Some modifications depend on previous modifications. Several types of covalent modifications (such as lysine acetylation and lysine methylation) are reversible.

Epigenetic modifications of chromatin recruit chromatin-associated proteins which decipher the code. Thus, modified histone residues serve as recognition marks that facilitate or prevent binding of proteins and protein complexes.

Activator and repressor proteins exploit the chromatin structure to turn genes on and off.

Histone modifications go along with **DNA methylation** patterns. In eukaryotes, only cytosine can become methylated (5-methylcytosine) (Figure 5).

In contrast, Archae and eubacteria have a variety of methylated nucleotides including N4-methylcytosine, 5-methylcytosine, and N6-methyladenine.

DNA methylation protects the genome by silencing the expression of transposable elements. In addition, DNA-methylation patterns have a major impact on

CC BY SA

gene expression. DNA methylation can be both stable (heritable) and flexible in time and space.

Chromatin structure can become modified by

- **ATP-dependent chromatin-remodelling complexes** which restructure nucleosomes and by
- enzymes that **covalently modify the core histones**. For example, many transcription activators recruit histone acetylases which add an acetyl-group to lysine residues of the histones. This modification confers greater accessibility of transcription factors to the DNA and, ultimately, helps to attract RNA polymerase.

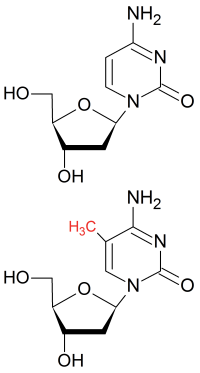


Figure 5. 2′-deoxy-cytidine (top) and 2′-deoxy-5-methylcytidine (bottom). For more details on the role of methylation in gene expression see Philips (2008).

Control of transcription elongation and termination

Riboswitches are RNA sensors that can detect changes in metabolites (e.g. amino acids, nucleotides, or metal ions) or other stimuli. Upon binding a ligand, the riboswitch undergoes a conformational change. The conformational change causes transcription termination (or, in eukaryotes, changes in splicing and mRNA degradation).

An example in *E. coli* is the riboswitch which controls purine biosynthesis (Figure 4).

For more details on riboswitches see:

A.D. Garst, A.L. Edwards and R.T. Batey (2011) Riboswitches: Structures and Mechanisms. Cold Spring Harb. Perspectiv. Biol. doi: 10.1101/csh-perspect.a003533

D. Goodsell (2010) Riboswitches. RCSB PDB

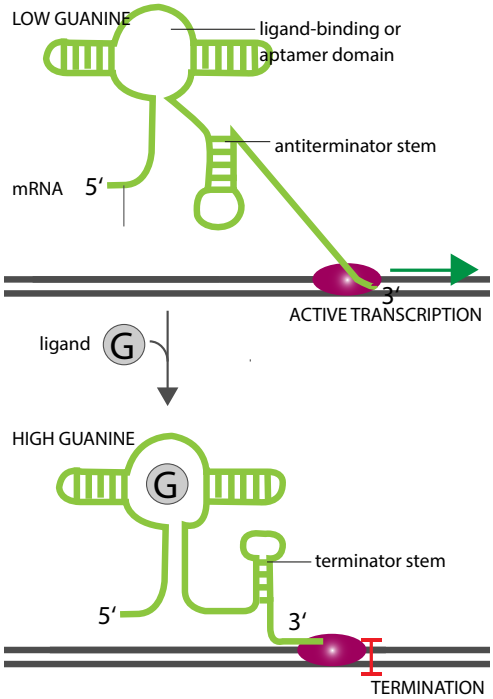


Figure 4. The expression of an enzyme needed for guanine synthesis is controlled by a riboswitch in bacteria. When guanine is scarce, the riboswitch allows active transcription. When guanine is plentiful, guanine binds to the aptamer domain and causes formation of a terminator stem. The new structure terminates transcription.

Control of translation

Translation initiation can be modulated in **bacteria** by metabolite-dependent **sequestration of a Shine-Dalgarno sequence** (e.g. TPP riboswitch).

The Shine-Dalgarno sequence is a short ribosome-binding sequence located upstream of the AUG start codon where translation begins. The sequence positions the initiating AUG codon within the ribosome and is essential for transla-

CC BY SA

tion of an mRNA.

The TPP riboswitch controls accessibility of the Shine-Dalgarno sequence. Binding of TPP (thiamine pyrophosphate) masks the Shine-Dalgarno sequence and prevents ribosome binding.

Protein synthesis in **eukaryotes** is principally regulated at the initiation stage. Translation initiation is a complex and highly regulated process that requires several protein factors.

It involves variable associations of **initiator factors**, structural **features of the mRNA** (e.g. 5′ cap and 3′ UTR and poly (A) tail, see page 5), RNA-binding proteins (e.g. **PABP** binding to the poly(A) tail), intracellular transport and localization of mRNAs, and small RNAs (e.g. microRNA-mediated repression of translation).

There are few clear cases of **regulated elongation**. For example, phosphorylation of elongation factors can affect ribosome transit along the mRNA.

by base-paring interactions. In cases of complementarity to the miRNA, target mRNAs can be degraded.

Dicer also cleaves double stranded RNA from **viruses** and **transposable elements** and generates small interfering RNAs (**siRNA**). These double-stranded RNAs are then incorporated into RISC and used to attack the potentially dangerous sequences by means of the pathway that also destroys endogenous mRNAs.

This targeted RNA degradation mechanism is termed **RNA interference** (RNAi).

For more details on RNAi see:

Nobelprize.org (2014) The Nobel Prize in Physiology or Medicine 2006

Nature Reviews Genetics (2009) RNAi technique animation tutorial from Nature Reviews Genetics. Scitable by Nature Education

G. Ruvkun (2010) What is RNA interference? Scitable by Nature Education.

Control of mRNA degradation

miRNA genes are transcribed by RNA polymerase II to generate stemloop containing primary miRNA (**pri-miRNA**). Like other noncoding RNAs (such as tRNA and rRNA), the pri-miRNA precursor transcript is processed to yield the mature miRNA.

First, a ~70 nt hairpin precursor miRNA (termed **pre-miRNA**) is formed (Figure 5). Next, the pre-miRNA is exported into the cytoplasm and cleaved to produce the mature ~22-nt **miRNA:miRNA*** duplex by an RNase called **Dicer**.

Next, the miRNA is assembled with specific proteins to form an RNA-induced silencing complex (**RISC**). The miRNA strand is incorporated into RISC, whereas the miRNA* strand is typically degraded.

Once incorporated into RISC, the miRNA guides the complex to its RNA targets

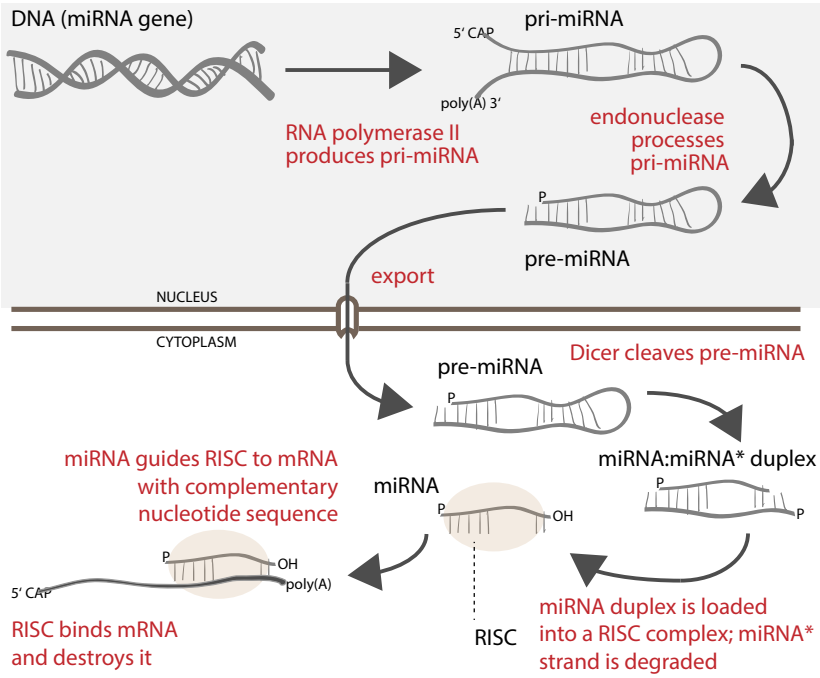


Figure 5. Biogenesis of a microRNA (miRNA).

CC BY SA

Further Readings

References

B. Alberts, D. Bray, K. Hopkin, A. Johnson, J. Lewis, M. Raff, K. Roberts and P. Walter (2009) Essential Cell Biology, 3rd edition. Garland Science

T.A. Brown (2002) Genomes. 2nd edition. Garland Science. http://www.ncbi.nlm.nih.gov/books/NBK21128/

T. Juven-Gershon, J.-Y. Hsu, J.W.M. Theisen, and J.T. Kadonaga (2008) The RNA Polymerase II Core Promoter – the Gateway to Transcription. Curr. Opin. Cell Biol. 20: 253-259

R.D. Kornberg (2007) The molecular basis of eukaryotic transcription. Proc. Natl. Acad. Sci. USA 104: 12955-12961

J. Krebs, E.S. Goldstein and S.T. Kilpatrick (2012) Lewin's GENES XI. 11th edition

T. Phillips (2008) The role of methylation in gene expression. Nature Education 1: 116

OER and OCW

L. Hoopes (editor) (2014) Gene expression and regulation. Scitable by Nature Education

E. Lander, R. Weinberg, T. Jacks, H. Sive, G. Walker, S. Chisholm and M. Mischke (2011) Fundamentals of Biology – Molecular Biology. MIT OCW

CC BY SA