

FUNDAMENTALS OF BIOCHEMISTRY III - INFORMATION PATHWAYS



Henry Jakubowski and Patricia Flatt
LibreTexts

Fundamentals of Biochemistry III - Information Pathways

Henry Jakubowski

College of St. Benedict/St. John's University

This text is disseminated via the Open Education Resource (OER) LibreTexts Project (<https://LibreTexts.org>) and like the thousands of other texts available within this powerful platform, it is freely available for reading, printing, and "consuming."

The LibreTexts mission is to bring together students, faculty, and scholars in a collaborative effort to provide an accessible, and comprehensive platform that empowers our community to develop, curate, adapt, and adopt openly licensed resources and technologies; through these efforts we can reduce the financial burden born from traditional educational resource costs, ensuring education is more accessible for students and communities worldwide.

Most, but not all, pages in the library have licenses that may allow individuals to make changes, save, and print this book. Carefully consult the applicable license(s) before pursuing such effects. Instructors can adopt existing LibreTexts texts or Remix them to quickly build course-specific resources to meet the needs of their students. Unlike traditional textbooks, LibreTexts' web based origins allow powerful integration of advanced features and new technologies to support learning.



LibreTexts is the adaptable, user-friendly non-profit open education resource platform that educators trust for creating, customizing, and sharing accessible, interactive textbooks, adaptive homework, and ancillary materials. We collaborate with individuals and organizations to champion open education initiatives, support institutional publishing programs, drive curriculum development projects, and more.

The LibreTexts libraries are Powered by [NICE CXone Expert](#) and was supported by the Department of Education Open Textbook Pilot Project, the California Education Learning Lab, the UC Davis Office of the Provost, the UC Davis Library, the California State University Affordable Learning Solutions Program, and Merlot. This material is based upon work supported by the National Science Foundation under Grant No. 1246120, 1525057, and 1413739.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation nor the US Department of Education.

Have questions or comments? For information about adoptions or adaptations contact info@LibreTexts.org or visit our main website at <https://LibreTexts.org>.

This text was compiled on 08/26/2025

TABLE OF CONTENTS

[Licensing](#)

[About the Authors](#)

23: Chromosome Structure

- [23.1: Gene Mapping and Chromosomal Karyotypes](#)
- [23.2: DNA Transposable Elements](#)
- [23.3: Chromosome Packaging](#)

24: DNA Metabolism

- [24.1: DNA Replication](#)
- [24.2: DNA Mutations, Damage, and Repair](#)
- [24.3: DNA Recombination](#)

25: RNA Metabolism

- [25.1: DNA-Dependent Synthesis of RNA](#)
- [25.2: RNA Processing](#)
- [25.3: RNA-Dependent Synthesis of RNA and DNA](#)
- [25.4: 26.4 References](#)

26: Protein Metabolism

- [26.1: The Genetic Code](#)
- [26.2: Protein Synthesis](#)
- [26.3: Translational Regulation and Protein Degradation](#)

27: Regulation of Gene Expression

- [27.1: Regulation of Gene Expression in Bacteria](#)
- [27.2: Regulation of Gene Expression in Eukaryotes](#)

[Index](#)

[Glossary](#)

[Detailed Licensing](#)

Licensing

A detailed breakdown of this resource's licensing can be found in [Back Matter/Detailed Licensing](#).

About the Authors

Henry Jakubowski

Dr. Henry Jakubowski is a Professor Emeritus of Chemistry at the College of Saint Benedict/Saint John's University (CSB/SJU) in Minnesota, where he taught for 32 years. Before that, he was a post-doctoral fellow at the Mayo Clinic, where he focused on biophysical techniques to explore protein function. His main expertise in biochemistry is in protein structure/function and enzyme kinetics. His research interests focused on the development and characterization of inhibitors for low molecular weight protein tyrosine phosphatase. While at CSB/SJU, he was active in major curricular reform. It involved replacing all introductory courses in the chemistry curriculum with classes focusing on structure/function relationships that integrated organic, inorganic, and biochemistry content. In addition, he designed novel in-depth chemistry classes in signal transduction and computational biochemistry. He was also actively involved in study abroad, and created a unique research exchange program with Southwest University in China, where he is a visiting Professor of Chemistry. For the last six years, he has become a climate activist and has given many talks and presentations to lay and academic groups. His current efforts focus on introducing climate change into the biochemistry curriculum.

Patricia Flatt

Dr. Patricia Flatt is a Professor of Chemistry at Western Oregon University. She holds a PhD in Biochemistry from Vanderbilt University where her research focused on the response of the p53 signaling pathways to DNA damage. Her current research interests include natural product biosynthesis in microorganisms and their evolutionary origins. She is a strong advocate for the advancement of women and underrepresented minorities pursuing careers in STEM. At WOU, she has served as the PI on a collaborative grant with Oregon State University to develop a STEM Scholars Center for Diversity and Excellence at WOU for the advancement of women and underrepresented minority students pursuing STEM degree paths. She is also serving as a Co-PI on an NSF ADVANCE grant supporting mid career women faculty in STEM, entitled 'ASCEND: Advancing STEM Careers by Empowering Network Development'. She believes that the creation of open education resources (OERs) can help to reduce the barriers to education and improve educational outcomes for underserved populations, as well as promote collaboration and sharing among educators to develop more effective and innovative teaching materials.

CHAPTER OVERVIEW

23: Chromosome Structure

[Return to Fundamentals of Biochemistry](#)

[Search Fundamentals of Biochemistry](#)

[23.1: Gene Mapping and Chromosomal Karyotypes](#)

[23.2: DNA Transposable Elements](#)

[23.3: Chromosome Packaging](#)

This page titled [23: Chromosome Structure](#) is shared under a [not declared](#) license and was authored, remixed, and/or curated by [Henry Jakubowski and Patricia Flatt](#).

File failed to load: <https://cdnjs.cloudflare.com/ajax/libs/mathjax/2.7.5/jax/output/HTML-CSS/jax.js>

23.2: DNA Transposable Elements

Learning Goals (ChatGPT o3-mini)

1. Define Transposable Elements (TEs) and Their Significance

- Explain what TEs are and describe their ability to move within genomes.
- Discuss the prevalence of TEs across different organisms (from bacteria to plants and animals) and how they can constitute large portions of a genome.

2. Historical and Scientific Context

- Summarize Barbara McClintock's discovery of TEs and its initial reception.
- Recognize how later discoveries (e.g., insertion sequences in bacteria) validated and expanded upon McClintock's work.

3. Classifications and Mechanisms of TEs

- Distinguish between the two major classes of TEs: Class I (retrotransposons, "copy-and-paste") and Class II (DNA transposons, "cut-and-paste").
- Identify and describe key orders within Class I (e.g., LTR retrotransposons, LINEs, SINEs, DIRS-like elements, Penelope-like elements) and subclass divisions within Class II (e.g., TIR elements, Helitrons, Mavericks).

4. Structural Features and Molecular Mechanisms

- Recognize characteristic structural elements such as long terminal repeats (LTRs), target site duplications (TSDs), and terminal inverted repeats (TIRs).
- Outline the general molecular mechanisms by which TEs mobilize and integrate into genomes (retrotranscription, excision, rolling-circle replication).

5. Impact on Genome Evolution and Function

- Analyze how TE insertions can disrupt gene function, modulate gene expression, or promote genomic rearrangements (e.g., deletions, duplications, inversions).
- Discuss the role of TEs in increasing genetic diversity and shaping genome architecture over evolutionary time.

6. Host-TE Interactions and Regulation

- Explain the selective pressures on both TEs and host genomes that lead to a balance between TE proliferation and host fitness.
- Describe the cellular defense mechanisms (e.g., DNA methylation, RNA interference, and sequence-specific repressors) that repress excessive TE activity.

7. Domestication and Exaptation of TEs

- Identify examples of TE "domestication" where TE-derived proteins (e.g., RAG proteins, CENP-B, SETMAR) have been repurposed for essential cellular functions.
- Evaluate how TE sequences contribute to the evolution of new regulatory elements and noncoding RNAs.

8. TEs as Sources of Mutation and Genetic Variation

- Summarize how active TE mobilization can lead to insertional mutagenesis in both the germline and somatic cells.
- Discuss the implications of TE-induced mutations in human diseases (e.g., genetic disorders, cancer) and natural population diversity.

9. Horizontal Transfer and Genome Dynamics

- Describe the concept of horizontal transposon transfer and its role in spreading TEs between different species.
- Assess how horizontal transfer contributes to rapid genomic evolution and the dissemination of TE families across taxa.

10. Interpreting TE-Related Data and Diagrams

- Develop skills in reading and interpreting TE classification schematics and genome distribution maps.

- Connect the structural and functional properties of TEs with their broader impact on genome organization and regulation.

These goals provide a framework for mastering the diverse roles that transposable elements play in genome dynamics, evolution, and cellular regulation, and will prepare you to critically evaluate both experimental data and current research in this exciting field.

23.2.1: Introduction

Eukaryotic genomes contain a substantial amount of repeated DNA, and some of these repeated sequences are capable of moving. **Transposable elements** (TEs) are defined as DNA sequences that can move from one location to another in the genome. TEs have been identified in all organisms, prokaryotic and eukaryotic, and can occupy a high proportion of a species' genome. For example, transposable elements comprise approximately 10% of the genomes of several fish species, 12% of the *C. elegans* genome, 37% of the mouse genome, 45% of the human genome, and up to more than 80% of the genomes of some plants, such as maize. From bacteria to humans, transposable elements have accumulated over time and continue to shape genomes through their mobilization.

TEs were discovered by Barbara McClintock during experiments conducted in 1944 on maize. Since they appeared to influence phenotypic traits, she named them **controlling elements**. However, her discovery was met with less than enthusiastic reception by the genetic community. Her presentation at the 1951 Cold Spring Harbor Symposium was not understood and was at least not very well received. She had no better luck with her follow-up publications, and after several years of frustration, decided not to publish on the subject for the next two decades. Not for the first time in the history of science, an unappreciated discovery was brought back to life after another discovery had been made. In this case, it was the discovery of insertion sequences (IS) in bacteria by the Szybalski group in the early 1970s. In the original paper, they wrote: "Genetic elements were found in higher organisms which appear to be readily transposed from one to another site in the genome. Such elements, identifiable by their controlling functions, were described by McClintock in maize. They might be somehow analogous to the presently studied IS insertions". The importance of McClintock's original work was eventually appreciated by the genetic community with numerous awards, including 14 honorary doctoral degrees and a Nobel Prize in 1983 "for her discovery of mobile genetic elements". Her picture is shown in Figure [\[Math Processing Error\]](#).



Figure [\[Math Processing Error\]](#): Barbara McClintock (1902-1992). The photo was taken in her lab at the Department of Genetics, Carnegie Institution at Cold Spring Harbor, New York. This photo was taken when McClintock received the American Association of University Women Achievement Award in 1947 for her work on cytogenetics. She received the Nobel Prize in Physiology or Medicine in 1983 for her discovery of mobile genetic elements. Image from: [Smithsonian Institution](#)

The mobilization of TEs is termed **transposition or retrotransposition**, depending on the nature of the intermediate used for mobilization. There are several ways in which the activity of TEs can positively and negatively impact a genome; for example, TE mobilization can promote gene inactivation, modulate gene expression or induce illegitimate recombination. Thus, TEs have played a significant role in genome evolution. For example, DNA transposons can inactivate or alter the expression of genes by insertion within introns, exons, or regulatory regions. In addition, TEs can participate in genome reorganization by mobilizing non-transposon DNA or by acting as substrates for recombination. This recombination occurs through homology between two

sequences of a transposon located on the same or different chromosomes, which may be the origin of various types of chromosome alterations. Indeed, TEs can contribute to the loss of genomic DNA through internal deletions or other mechanisms.

The reduction in fitness suffered by the host due to transposition ultimately affects the transposon, since host survival is critical to the perpetuation of the transposon. Therefore, strategies have been developed by host and transposable elements to minimize the deleterious impact of transposition and to reach equilibrium. For example, some transposons tend to insert in non-essential regions of the genome, such as heterochromatic regions, where insertions are likely to have a minimal deleterious impact. In addition, they might be active in the germ line or embryonic stage, where most deleterious mutations can be selected against during fecundation or development, allowing only non-deleterious or mildly deleterious insertions to pass to successive generations. New insertions may also occur within an existing genomic insertion to generate an inactive transposon, or can undergo self-regulation by **overproduction-inhibition**. On the other hand, host organisms have developed various mechanisms of defense against high rates of transposon activity, including DNA methylation to reduce TE expression, several RNA interference-mediated mechanisms, primarily in the germ line, or through the inactivation of transposon activity by the action of specific proteins.

In some cases, transposable elements have been “domesticated” by the host to perform a specific function in the cell. A well-known example is RAG proteins, which participate in V(D)J recombination during antibody class switching, and exhibit a high similarity to DNA transposons, from which these proteins appear to be derived. Another example is the centromeric protein CENP-B, which seems to have originated from the pogo-like transposon. The analogous human mariner *Himar1* element has been incorporated into the SETMAR gene, which consists of the histone H3 methylase gene and the Himar1 transposase domain. This gene is involved in the non-homologous end-joining pathway of DNA repair and has been shown to confer resistance to ionizing radiation. From a genome-wide perspective, it has been estimated that approximately 25% of human promoter regions and 4% of human exons contain sequences derived from transposable elements (TEs). Thus, we are likely underestimating the rate of domestication events in mammalian genomes.

The first TE classification system was proposed by Finnegan in 1989, distinguishing two classes of TEs characterized by their transposition intermediates: RNA (class I, or retrotransposons) or DNA (class II, or DNA transposons). The transposition mechanism of class I is commonly called “copy and paste,” and that of class II, “cut and paste.” In 2007, Wicker et al. proposed a hierarchical classification based on the structural characteristics and mode of replication of TEs, as shown in Figure [\[Math Processing Error\]](#).

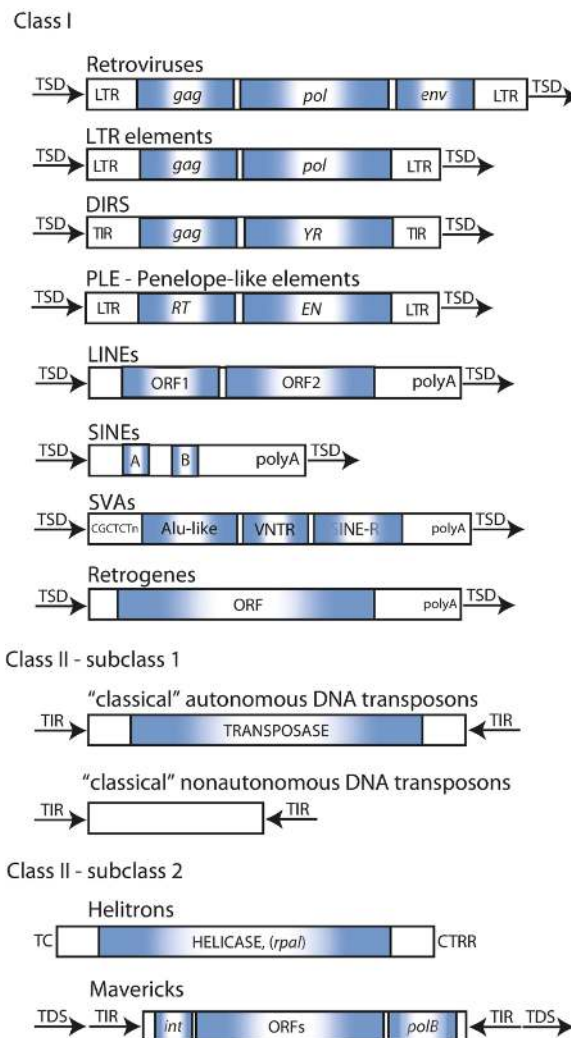


Figure [Math Processing Error]: Structures of eukaryotic mobile elements. See text for detailed discussion Image from Makalowski, W, et. al. (2019)

23.2.2: Class I: Mobile Elements

As mentioned above, class I transposable elements (TEs) transpose through an RNA intermediary. The RNA intermediate is transcribed from genomic DNA and then reverse-transcribed into DNA by a TE-encoded reverse transcriptase (RT), followed by reintegration into a genome. Each replication cycle produces one new copy, and as a result, class I elements are the major contributors to the repetitive fraction in large genomes. Retrotransposons are divided into five orders: LTR retrotransposons, DIRS-like elements, Penelope-like elements (PLEs), LINEs (*long interspersed elements*), and SINEs (*short interspersed elements*). This scheme is based on the mechanistic features, organization, and phylogeny of the reverse transcriptase of these retroelements. Accidentally, the reverse transcriptase coded by an autonomous transposable element (TE) can reverse-transcribe another RNA present in the cell, such as mRNA, and produce a retrocopy of it, which in most cases results in a pseudogene.

The LTR retrotransposons are characterized by the presence of *long terminal repeats* (LTRs) ranging from several hundred to several thousand base pairs. Both exogenous retroviruses and LTR retrotransposons contain a *gag* gene that encodes a viral particle coat and a *pol* gene that encodes a reverse transcriptase, ribonuclease H, and an integrase, which provide the enzymatic machinery for reverse transcription and integration into the host genome. Reverse transcription occurs within the viral or viral-like particle (GAG) in the cytoplasm, and it is a multi-step process. Unlike LTR retrotransposons, exogenous retroviruses contain an *env* gene, which encodes an envelope that facilitates their migration to other cells. Some LTR retrotransposons may contain remnants of an *env* gene, but their insertion capabilities are limited to the genome from which they originated. This would rather suggest that they originated from exogenous retroviruses by losing the *env* gene. However, there is evidence that suggests the contrary, given that LTR retrotransposons can acquire the *env* gene and become infectious entities. Currently, most long terminal repeat (LTR) sequences (85%) in the human genome are found only as isolated LTRs, with the internal sequence likely lost due to homologous

recombination between flanking LTRs. Interestingly, LTR retrotransposons target their reinsertion to specific genomic sites, often around genes, with putative important functional implications for a host gene. It is estimated that 450,000 LTR copies comprise approximately 8% of our genome. LTR retrotransposons inhabiting large genomes, such as maize, wheat, or barley, can contain thousands of families. However, despite the diversity, very few families comprise most of the repetitive fraction in these large genomes. Notable examples are Angela (wheat), BARE1 (barley), Opie (maize), and Retrosor6 (sorghum).

The DIRS order clusters structurally diverged groups of transposons that possess a tyrosine recombinase (YR) gene instead of an integrase (INT) and do not form target site duplications (TSDs). Their termini resemble either split direct repeats (SDR) or inverted repeats. Such features indicate a different integration mechanism than that of other class I mobile elements. DIRS were discovered in the genome of the slime mold (*Dictyostelium discoideum*) in the early 1980s and are present in all major phylogenetic lineages, including vertebrates. It has been shown that they are also common in hydrothermal vent organisms.

Another order, termed *Penelope*-like elements (PLE), has a wide, though patchy, distribution, ranging from amoebae and fungi to vertebrates, with copy numbers of up to thousands per genome. Interestingly, no PLE sequences have been found in mammalian genomes, and it appears that they were lost from the genome of *C. elegans*. Although PLEs with an intact open reading frame (ORF) have been identified in several genomes, including *Ciona* and *Danio*, the only transcriptionally active representative, *Penelope*, is known from *Drosophila virilis*. It causes the hybrid dysgenesis syndrome, characterized by the simultaneous mobilization of several unrelated transposable element families in the progeny of dysgenic crosses. It appears that *Penelope* invaded *D. virilis* relatively recently, and its invasive potential has been demonstrated in *D. melanogaster*. PLEs harbor a single ORF that codes for a protein containing reverse transcriptase (RT) and endonuclease (EN) domains. The PLE RT domain more closely resembles telomerase than the RT from LTRs or LINEs. The EN domain is related to GIY-YIG intron-encoded endonucleases. Some PLE members also have LTR-like sequences, which can be in either a direct or an inverted orientation, and contain a functional intron.

LINEs do not have LTRs; however, they have a poly-A tail at the 3' ends and are flanked by the TSDs. They comprise about 21% of the human genome, and among them, L1, with about 850,000 copies, is the most abundant and best-described LINE family. L1 is the only LINE retrotransposon still active in the human genome. In the human genome, there are two other LINE-like repeats, L2 and L3, distantly related to L1. A contrasting situation has been observed in the malaria mosquito *Anopheles gambiae*, where approximately 100 divergent LINE families comprise only 3% of its genome. LINEs in plants, such as Cin4 in maize and Ta11 in *Arabidopsis thaliana*, appear to be rare compared to LTR retrotransposons. A full copy of mammalian L1 is about 6 kb long and contains a PolIII promoter and two ORFs. The ORF1 codes for a non-sequence-specific RNA binding protein that contains zinc finger, leucine zipper, and coiled-coil motifs. The ORF1p functions as a chaperone for the L1 mRNA. The second ORF encodes an endonuclease, which makes a single-stranded nick in the genomic DNA, and a reverse transcriptase, which uses the nicked DNA to prime reverse transcription of LINE RNA from the 3' end. Reverse transcription is often unfinished, leaving behind fragmented copies of LINE elements; hence, most of the L1-derived repeats are short, with an average size of 900 bp. LINEs are part of the CR1 clade, which has members in various metazoan species, including fruit flies, mosquitoes, zebrafish, pufferfish, turtles, and chickens. Because they encode their own retrotransposition machinery, LINE elements are regarded as autonomous retrotransposons.

SINEs evolved from RNA genes, such as 7SL and tRNA genes. By definition, they are short (up to 1000 base pairs). They do not encode their own retrotranscription machinery and are considered nonautonomous elements; in most cases, they are mobilized by the L1 machinery. The outstanding member of this class from the human genome is the Alu repeat, which contains a cleavage site for the *AluI* restriction enzyme, giving it its name. With over a million copies in the human genome, Alu is probably the most successful transposon in the history of life. Primate-specific Alu and its rodent relative, B1, have a limited phylogenetic distribution, suggesting their relatively recent origins. The mammalian-wide interspersed repeats (MIRs), by contrast, spread before eutherian radiation, and their copies can be found in different mammalian groups, including marsupials and monotremes. SVA elements are unique primate elements due to their composite structure. They are named after their main components: SINE, VNTR (a variable number of tandem repeats), and Alu. Typically, they exhibit the hallmarks of retroposition, i.e., they are flanked by TSDs and terminated by a poly(A) tail. It appears that SVA elements are non-autonomous retrotransposons mobilized by L1 machinery, and they are believed to be transcribed by RNA polymerase II. SVAs are transpositionally active and are responsible for some human diseases. They originated less than 25 million years ago and form the youngest retrotransposon family, with approximately 3,000 copies in the human genome.

Retro(pseudo)genes are a special group of retroposed sequences, which are products of reverse transcription of a spliced (mature) mRNA. Hence, their characteristic features are an absence of promoter sequence and introns, the presence of flanking direct

repeats, and a 3'-end polyadenosine tract. Processed pseudogenes, also known as retropseudogenes, have been generated in vitro at a low frequency in human HeLa cells using mRNA from a reporter gene. The source of the reverse transcription machinery in humans and other vertebrates seems to be active L1 elements. However, not all retroposed messages have to end up as pseudogenes. About 20% of mammalian protein-encoding genes lack introns in their ORFs. It is conceivable that many genes lacking introns arose by retroposition. Some genes are known to be more frequently retroposed than others. For instance, in the human genome, there are over 2000 retropseudogenes of ribosomal proteins. A genome-wide study revealed that the human genome contains approximately 20,000 pseudogenes, 72% of which are likely to have arisen through retroposition. Interestingly, the vast majority (92%) of these are quite recent transpositions that occurred after the divergence of primates and rodents. Some of the retroposed genes may undergo quite complicated evolutionary paths. An example is the RNF13B retrogene, which replaced its parental gene in mammalian genomes. This retrocopy was duplicated in primates, and the evolution of this primate-specific copy was accompanied by the exaptation of two TEs, Alu and L1, and intron gain via changing a part of the coding sequence into an intron, leading to the origin of a functional, primate-specific retrogene with two splicing variants.

23.2.3: Class II: Mobile Elements

Class II elements move by a conservative cut-and-paste mechanism; the excision of the donor element is followed by its reinsertion elsewhere in the genome. DNA transposons are abundant in bacteria, where they are called insertion sequences, but are also present in all phyla. Two subclasses of DNA transposons have been distinguished, based on the number of DNA strands that are cut during transposition.

Classical “cut-and-paste” transposons belong to subclass I, and they are classified as the TIR order. They are characterized by terminal inverted repeats (TIRs) and encode a transposase that binds near the inverted repeats, mediating mobility. This process is not typically replicative, unless the gap caused by excision is repaired using the sister chromatid. When inserted at a new location, the transposon is flanked by small gaps, which, when filled by host enzymes, cause duplication of the sequence at the target site. The length of these TSDs is characteristic of particular transposons. Nine superfamilies belong to the TIR order, including *Tc1-Mariner*, *Merlin*, *Mutator*, and *PiggyBac*. The second-order Crypton consists of a single superfamily of the same name. Originally thought to be limited to fungi, it is now clear that they have a wide distribution, including animals and heterokonts. A heterogeneous, small, nonautonomous group of elements, MITEs, also belong to the TIR order, which in some genomes is amplified to thousands of copies, e.g., *Stowaway* in the rice genome, *Tourist* in most bamboo genomes, or *Galluhop* in the chicken genome.

Subclass II includes two orders of TEs that, just as those from subclass I, do not form RNA intermediates. However, unlike “classical” DNA transposons, they replicate without double-strand cleavage. Helitrons replicate using a rolling-circle mechanism, and their insertion does not result in duplication of the target site. They encode tyrosine recombinase along with some other proteins. Helitrons were first described in plants, but they are also present in other phyla, including fungi and mammals. Mavericks are large transposons that have been found in different eukaryotic lineages, excluding plants. They encode various proteins, including DNA polymerase B and an integrase. Kapitonov and Jurka suggested that their life cycle includes a single-strand excision, followed by extrachromosomal replication and reintegration to a new location.

23.2.4: TEs are not randomly distributed in the genome

As seen in the previous section, TEs are highly diverse and, in principle, every TE sequence in a genome can be affiliated with a (sub)family, superfamily, subclass, and class. This is summarized in Figure [\[Math Processing Error\]](#). However, much like the taxonomy of species, the classification of TEs is in constant flux, perpetually subject to revision due to the discovery of completely novel TE types, the introduction of new levels of granularity in the classification, and the ongoing development of methods and criteria to detect and classify TEs.

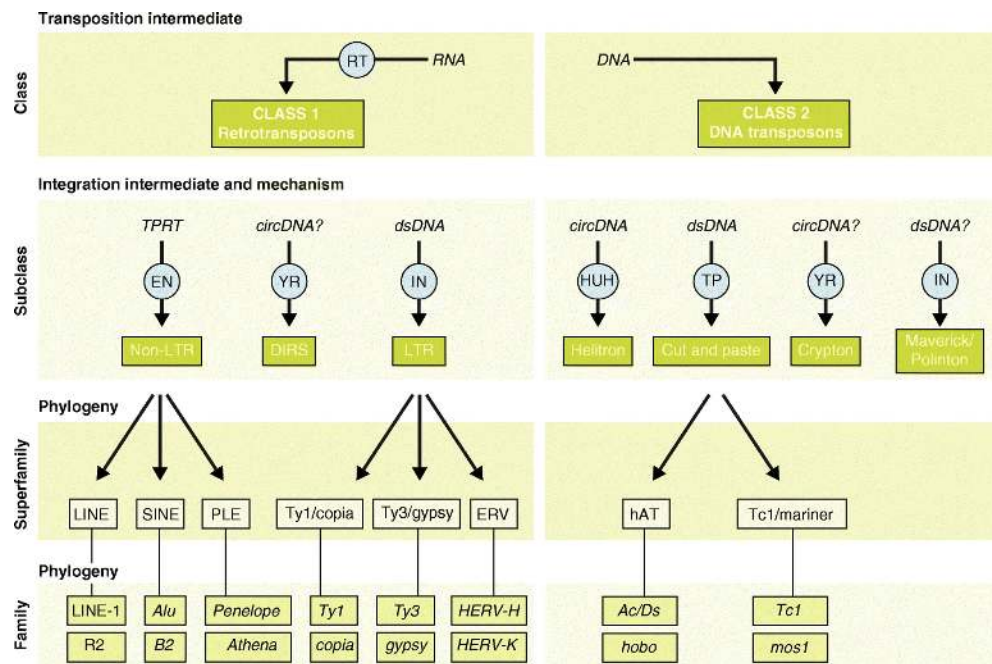


Figure [Math Processing Error]: Classification of Eukaryotic Transposable Elements. This schematic shows the key features and relationships between TE classes, subclasses, superfamilies, and families. Blue circles represent TE-encoded enzymes. circDNA circular DNA intermediate, DIRS Dictyostelium repetitive sequence, dsDNA linear double-stranded DNA intermediate, EN endonuclease, IN integrase, PLEs Penelope-like elements, HUH, Rep/Helicase protein with HUH endonuclease activity, RT reverse transcriptase, TP transposase, TPRT target primed reverse transcription, YR tyrosine recombinase. Image from: Bourque, G., et. al. (2018) *Genetic Biology* 19:199

The genome can be viewed as an ecosystem inhabited by diverse communities of transposable elements (TEs), which seek to propagate and multiply through sophisticated interactions with one another and with other cellular components. These interactions encompass processes familiar to ecologists, such as parasitism, cooperation, and competition. Thus, it is perhaps not surprising that TEs are rarely, if ever, randomly distributed in the genome, as shown in Figure [Math Processing Error]. TEs exhibit various levels of preference for insertion within certain features or compartments of the genome. These are often guided by opposing selective forces, a balancing act that facilitates future propagation while mitigating deleterious effects on host cell function. At the end of the site-selection spectrum, many elements have evolved mechanisms to target specific loci where their insertions are less detrimental to the host but favorable for their propagation. For instance, several retrotransposons in species as diverse as slime mold, budding, and fission yeast have evolved independently, but convergently, the capacity to target the upstream regions of genes transcribed by RNA polymerase III, where they do not appear to affect host gene expression but retain the ability to be transcribed themselves.

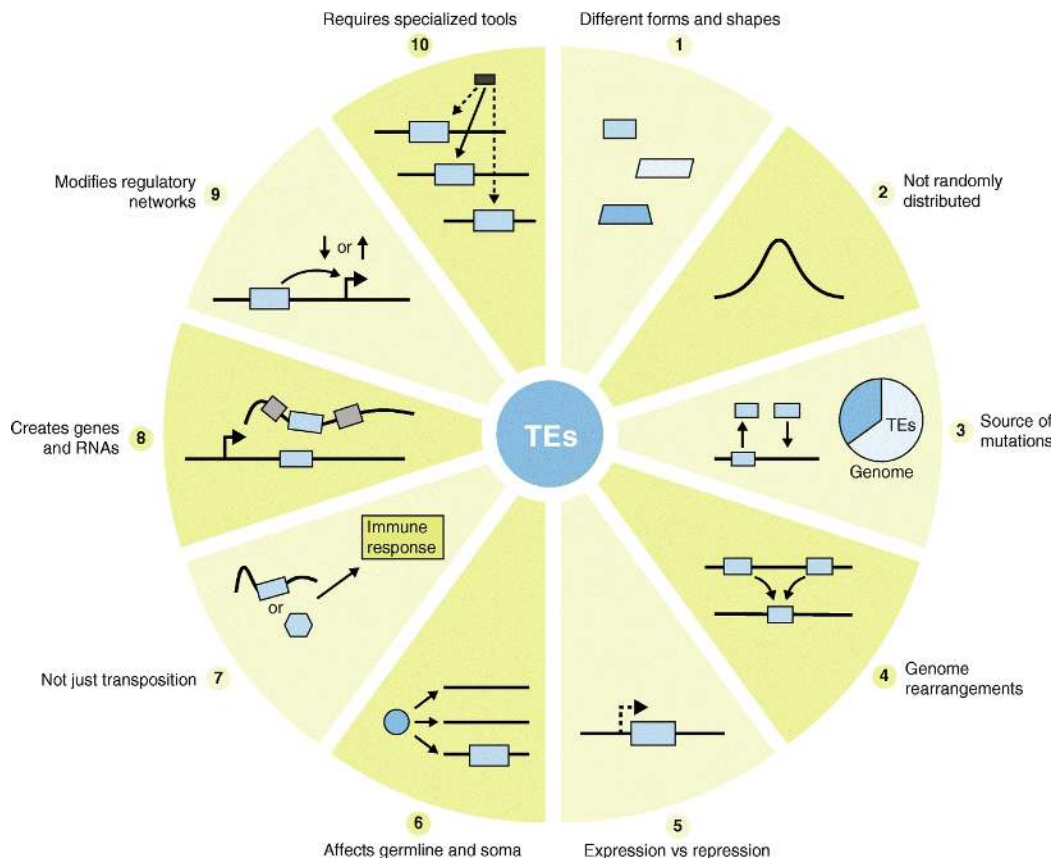


Figure [Math Processing Error]: Ten things you should know about transposable elements (TEs). Examples of how TEs can impact genomes in direct and indirect ways. Blue boxes represent TEs, gray boxes represent canonical exons, and the black box represents a sequencing read. Right-angled arrows represent gene or TE promoters. Image from: Bourque, G., et al. (2018) *Genetic Biology* 19:199

23.2.5: TEs are an extensive source of mutations and genetic polymorphisms

TEs occupy a substantial portion of the genome of a species, including a large fraction of the DNA unique to that species. In maize, where Barbara McClintock did her seminal work, an astonishing 60 to 70% of the genome is comprised of LTR retrotransposons, many of which are unique to this species or its close wild relatives. Still, the less prevalent DNA transposons are currently the most active and mutagenic (Fig. 24.2.4). Similarly, the vast majority of TE insertions in *Drosophila melanogaster* are absent at the orthologous site in its closest relative *D. simulans* (and vice versa). Most are not fixed in the population. Many TE families are still actively transposing, and the process is highly mutagenic; more than half of all known phenotypic mutants of *D. melanogaster* isolated in the laboratory are caused by spontaneous insertions of a wide variety of TEs. Transposition events are also common and mutagenic in laboratory mice, where the ongoing activity of several families of LTR elements is responsible for 10–15% of all inherited mutant phenotypes. This contribution of TEs to genetic diversity may be underestimated, as TEs can be more active when organisms are under stress, such as in their natural environment.

Because TE insertions rarely provide an immediate fitness advantage to their host, those that reach fixation in the population do so largely through genetic drift and are subsequently eroded by point mutations that accumulate neutrally. Over time, these mutations result in transposable elements (TEs) that can no longer encode transposition enzymes and therefore produce new integration events. For instance, our haploid genome contains approximately 500,000 L1 copies, but more than 99.9% of these L1 copies are fixed and no longer mobile due to various forms of mutations and truncations. It is estimated that each person carries a set of approximately 100 active L1 elements, and most of these are young insertions that are still segregating within the human population. Thus, like any other organism, the ‘reference’ human genome sequence does not represent a comprehensive inventory of transposable elements (TEs) in humans. Thousands of ‘non-reference’, unfixed TE insertions have been cataloged through whole-genome sequencing and other targeted approaches. On average, any two human haploid genomes differ by approximately a thousand transposable elements (TEs) insertions, primarily from the L1 or Alu families. The number of TE insertion polymorphisms in a species with much higher TE activity, such as maize, dwarfs the number in humans.

If TEs bring no immediate benefit to their host and are largely decaying neutrally once inserted, how do they persist in evolution? One key to this conundrum is the ability of TEs not only to propagate vertically but also horizontally between individuals and species. There is now a substantial body of evidence supporting the idea that horizontal transposon transfer is a common phenomenon that affects virtually every major type of transposable element (TE) and all branches of the tree of life. While the cellular mechanisms underlying horizontal transposon transfer remain murky, it is increasingly apparent that the intrinsic mobility of TEs and ecological interactions between their host species, including those with pathogens and parasites, facilitate the transmission of elements between widely diverged taxa.

TEs are associated with genome rearrangements and unique chromosome features

Transposition represents a potent mechanism of genome expansion that, over time, is counteracted by the removal of DNA via deletion. The balance between the two processes is a major driver in the evolution of genome size in eukaryotes. Several studies have demonstrated the impact and range of this shuffling and cycling of genomic content on the evolution of plant and animal genomes. Because the insertion and removal of TEs are often imprecise, these processes can indirectly affect surrounding host sequences. Some of these events occur at high enough frequency to result in vast amounts of duplication and reshuffling of host sequences, including genes and regulatory sequences. For example, a single group of DNA transposons, known as MULEs, has been responsible for the capture and reshuffling of approximately 1,000 gene fragments in the rice genome. Such studies have led to the conclusion that the rate at which transposable elements (TEs) transpose, which is in part under host control, is a significant driver of genome evolution.

In addition to rearrangements induced as a byproduct of transposition, TEs can promote genomic structural variation long after they have lost the capacity to mobilize. In particular, recombination events can occur between the highly homologous regions dispersed by related TEs at distant genomic positions and result in large-scale deletions, duplications, and inversions (Fig. 24.2.4). TEs also provide regions of microhomology that predispose to template switching during repair of replication errors leading to another source of structural variants. These non-transposition-based mechanisms for TE-induced or TE-enabled structural variation have contributed substantially to genome evolution. These processes can also complicate the identification of actively transposing elements in population studies that infer the existence of active elements through the detection of non-reference insertions.

TEs also contribute to the specialized features of chromosomes. An intriguing example is found in *Drosophila*, where LINE-like retrotransposons form and maintain telomeres in place of the telomerase enzyme, which was lost during dipteran evolution. This domestication event could be viewed as a replay of what might have happened much earlier in eukaryotic evolution to solve the ‘end problem’ created by the linearization of chromosomes. Indeed, the reverse transcriptase component of telomerase is thought to have originated from an ancient lineage of retroelements. TE sequences and domesticated transposase genes also play structural roles at centromeres.

There is an intricate balance between TE expression and repression

To persist in evolution, TEs must strike a delicate balance between expression and repression (Fig. 24.2.4). Expression should be sufficient to promote amplification, but not so vigorous as to lead to a fitness disadvantage for the host that would offset the benefit to the TE of increased copy numbers. This balancing act may explain why TE-encoded enzymes are naturally suboptimal for transposition and why some transposable elements (TEs) have evolved self-regulatory mechanisms that control their copy numbers. A variety of host factors are also employed to control TE expression, which includes small RNA, chromatin, and DNA modification pathways, as well as sequence-specific repressors such as the recently profiled KRAB zinc-finger proteins. However, many of these silencing mechanisms must be at least partially released to permit the developmental regulation of host gene expression programs, particularly during early embryonic development. For example, genome-wide loss of DNA methylation is necessary to reset imprinted genes in primordial germ cells. This affords TEs an opportunity, as reduced DNA methylation often promotes the expression of TEs. Robust expression of a transposable element (TE) in the germ lineage (but not necessarily in the gametes themselves) is often its own downfall. In one example of a clever trick employed by the host, TE repression is relieved in a companion cell derived from the same meiotic product as flowering plant sperm. However, this companion cell does not contribute genetic material to the next generation. Thus, although TEs transpose in a meiotic product, the events are not inherited. Instead, TE activity in the companion cell may further dampen TE activity in sperm via the import of TE-derived small RNAs.

Another important consequence of the intrinsic expression/repression balance is that the effects of TEs on a host can vary considerably among tissue types and stages of an organism’s life cycle. From the TE’s perspective, an ideal scenario is to be expressed and active in the germline, but not in the soma, where expression would gain the TE no advantage, only disadvantages. This phenomenon is indeed observed among many species, with ciliates representing an extreme example of this division—TEs are actively deleted from the somatic macronucleus but retained in the micronucleus, which serves as the germline. Another example is

the P-elements in *Drosophila*, which are differentially spliced in the germline versus soma. Many organisms, including plants, do not differentiate germ lineage cells early in development; rather, they are specified from somatic cells shortly before meiosis commences. Thus, TEs that transpose in somatic cells in plants have the potential to be inherited, which suggests that the interests of TEs and hosts conflict in many more cells and tissues than in animals with a segregated germline.

TEs are insertional mutagens in both the germline and soma

Like other species, humans contend with a contingent of currently active transposable elements (TEs), where the intrinsic balance between expression and repression is still at play. For us, this includes L1 and other mobile elements that depend on L1-encoded proteins for retrotransposition. These elements are responsible for new germline insertions that can lead to genetic diseases. More than 120 independent TE insertions have been associated with human disease. The rate of de novo germline transposition in humans is approximately one in 21 births for *Alu* and one in 95 births for L1.

Historically, little attention has been given to transposition in somatic cells and its consequences, as somatic transposition may be viewed as an evolutionary dead end for the transposable element (TE) with no long-term consequences for the host species. Yet, there is abundant evidence that TEs are active in somatic cells in many organisms (Fig. 24.2.4). In humans, L1 expression and transposition have been detected in various somatic contexts, including early embryos and specific stem cells. There is also considerable interest in the expression and activity of mobile elements in the mammalian brain, where L1 transposition has been proposed to diversify neuronal cell populations. One challenge in assessing somatic activity has been the development of reliable single-cell insertion site mapping strategies.

Somatic activity has also been observed in human cancers, where tumors can acquire hundreds of new L1 insertions. As for human polymorphisms, somatic activity in human cancers is caused by small numbers of so-called ‘hot’ L1 loci. The activities of these master copies vary depending on the individual, tumor type, and timeframe in the clonal evolution of the tumor. Some of these de novo L1 insertions disrupt critical tumor suppressors and oncogenes and thus drive cancer formation, although the vast majority appear to be ‘passenger’ mutations. Host cells have evolved several mechanisms to regulate transposable elements (TEs). However, as the force of natural selection begins to diminish with age and completely drops in post-reproductive life, TEs may become more active.

TEs can be damaging in ways that do not involve transposition

TEs are best known for their mobility and their ability to transpose to new locations. While the breakage and insertion of DNA associated with transposition represent an obvious source of cell damage, this is not the only, nor perhaps even the most common, mechanism by which TEs can be harmful to their host. Reactivated transposons harm the host in multiple ways. First, derepression of transposon loci, including their own transcription, may interfere with transcription or processing of host mRNAs through a myriad of mechanisms. Genome-wide transcriptional derepression of TEs has been documented during replicative senescence of human cells and several mouse tissues, including the liver, muscle, and brain. De-repression of LTR and L1 promoters can also cause oncogene activation in cancer. Second, TE-encoded proteins, such as the endonuclease activity of L1 ORF2p, can induce DNA breaks and genomic instability. Third, accumulation of RNA transcripts and extrachromosomal DNA copies derived from TEs may trigger an innate immune response leading to autoimmune diseases and sterile inflammation (Fig. 24.2.4). Activation of interferon response is now a well-documented property of transcripts derived from endogenous retroviruses and may give immunotherapies a boost in identifying and attacking cancer cells. The relative contribution of all the above mechanisms in organismal pathologies remains to be determined.

Following transcription (and sometimes splicing) of TEs, the next step in the process involves the translation of the encoded proteins and, for retroelements, reverse transcription of the TEs into cDNA substrates suitable for transposition. Once engaged by a TE-encoded reverse transcriptase protein, the resulting cytosolic DNAs and RNA:DNA hybrids can alert inflammatory pathways. An example of this is seen in patients with Aicardi–Goutières syndrome, where the accumulation of TE-derived cytosolic DNA is due to mutations in pathways that normally block TE processing or degrade TE-derived DNA. Although not all TEs encode functional proteins, some do, including a few endogenous retroviruses capable of producing Gag, Pol, or envelope (Env) proteins. Overexpression of these Env proteins can be cytotoxic and has been linked to at least two neurodegenerative diseases, multiple sclerosis and amyotrophic lateral sclerosis. Small accessory proteins produced by the youngest human endogenous retrovirus (HERV) group, HERV-K (HML-2), may play a role in some cancers, but the evidence remains circumstantial.

Key coding and non-coding RNAs are derived from TEs

Although usually detrimental, there is growing evidence that TE insertions can provide the raw material for the emergence of protein-coding genes and non-coding RNAs, which can take on important and, in some cases essential, cellular function (Fig.

24.2.4). The process of TE gene ‘domestication’ or exaptation over evolutionary time contributes to both deeply conserved functions and more recent, species-specific traits. Most often, the ancestral or a somewhat modified role of a TE-encoded gene is harnessed by the host and conserved, while the rest of the TE sequence, and hence its ability to autonomously transpose, has been lost. Spectacular examples of deeply conserved TE-derived genes are *Rag1* and *Rag2*, which catalyze V(D)J somatic recombination in the vertebrate immune system. Both genes, and probably the DNA signals they recognize, were derived from an ancestral DNA transposon around 500 million years ago. Indeed, DNA transposases have been co-opted multiple times to form new cellular genes.

The *gag* and *env* genes of LTR retrotransposons or endogenous retroviruses (ERVs) have been domesticated numerous times to perform various functions in placental development, contribute to host defense against exogenous retroviruses, influence brain development, and play other diverse roles. One of the most intriguing examples of TE domestication is the repeated, independent capture of ERV *env* genes, termed *syncytins*, which appear to function in placentation by facilitating cell–cell fusion and the formation of syncytiotrophoblasts. Notably, one or more *syncytin* genes have been found in virtually every placental mammalian lineage where they have been sought, strongly suggesting that ERVs have played essential roles in the evolution and extreme phenotypic variability of the mammalian placenta. Another example of a viral-like activity re-purposed for host cell function is provided by the neuronal *Arc* gene, which arose from the *gag* gene of an LTR retrotransposon domesticated in the common ancestor of tetrapod vertebrates. Genetic and biochemical studies of murine *Arc* show that it is involved in memory and synaptic plasticity and has preserved most of the ancestral activities of *Gag*, including the packaging and intercellular trafficking of its own RNA. Remarkably, flies appear to have independently evolved a similar system of trans-synaptic RNA delivery involving a *gag*-like protein derived from a similar yet distinct lineage of LTR retrotransposons. Thus, the biochemical activities of TE-derived proteins have been repeatedly co-opted during evolution to foster the emergence of convergent cellular innovations in different organisms.

TEs can donate their own genes to the host, but they can also add exons and rearrange and duplicate existing host genes. In humans, intronic *Alu* elements are particularly prone to be captured as alternative exons through cryptic splice sites residing within their sequences. L1 and SVA (SINE/VNTR/*Alu*) elements also contribute to exon shuffling through transduction events of adjacent host sequences during their mobilization. The reverse transcriptase activity of retroelements is also responsible for the trans-duplication of cellular mRNAs to create ‘processed’ retrogenes in a wide range of organisms. The L1 enzymatic machinery is thought to be involved in the generation of tens of thousands of retrogene copies in mammalian genomes, many of which remain transcribed and some of which have acquired new cellular functions. This is a process that is still actively shaping our genomes; it has been estimated that one in every 6,000 humans carries a novel retrogene insertion.

TEs also make substantial contributions to the non-protein-coding functions of the cell. They are major components of thousands of long non-coding RNAs in human and mouse genomes, often transcriptionally driven by retroviral long terminal repeats (LTRs). Some of these TE-driven lncRNAs appear to play crucial roles in maintaining stem cell pluripotency and other developmental processes. Numerous studies have shown that TE sequences embedded within lncRNAs and mRNAs can directly modulate RNA stability, processing, or localization, with significant regulatory consequences. Furthermore, TE-derived microRNAs and other small RNAs processed from transposable elements (TEs) can also adopt regulatory roles, serving host cell functions. The myriad of mechanisms by which TEs contribute to coding and non-coding RNAs illustrate the multi-faceted interactions between these elements and their host.

TEs contribute cis-regulatory DNA elements and modify transcriptional networks

Cis-regulatory networks coordinate the transcription of multiple genes that function in concert to orchestrate entire pathways and complex biological processes. In line with Barbara McClintock’s insightful predictions, there is now mounting evidence that TEs have been a rich source of material for the modulation of eukaryotic gene expression (Fig. 24.2.4). Indeed, TEs can disperse vast amounts of promoters and enhancers, transcription factor binding sites, insulator sequences, and repressive elements. The varying coat colors of agouti mice provide a striking example of a host gene controlling coat color whose expression can be altered by the methylation levels of a transposable element (TE) upstream of its promoter. In the oil palm, the methylation level of a transposable element (TE) that sits within a gene important for flowering ultimately controls whether the plants bear oil-rich fruit.

As TE families typically populate a genome as a multitude of related copies, it has long been postulated that they have the potential to donate the same cis-regulatory module to ‘wire’ batteries of genes dispersed throughout the genome. An increasing number of studies support this model, suggesting that TEs have provided the building blocks for the assembly and remodeling of cis-regulatory networks during evolution. These networks underlie processes as diverse as pregnancy, stem cell pluripotency, neocortex development, innate immunity in mammals, and the response to abiotic stress in maize. Indeed, TE sequences harbor all the necessary features of a ‘classical’ gene regulatory network. They are bound by diverse sets of transcription factors that integrate

multiple inputs (activation/repression), respond to signals in both *cis* and *trans*, and are capable of coordinately regulating gene expression. In this context, TEs are highly suitable agents to modify biological processes by creating novel *cis*-regulatory circuits and fine-tuning pre-existing networks.

Outlook

As potent insertional mutagens, TEs can have both positive and negative effects on host fitness. Still, it is likely that the majority of TE copies in any given species—and especially those such as humans with a small effective population size—have reached fixation through genetic drift alone and are now largely neutral to their host. When can we say that TEs have been co-opted for cellular function? The publication of the initial ENCODE paper, which asserted ‘function for 80% of the genome’, was the subject of much debate and controversy. Technically speaking, ENCODE assigned only ‘biochemical’ activity to this large fraction of the genome. Yet critics objected to the grand proclamations in the popular press (The Washington Post Headline: “Junk DNA concept debunked by new analysis of the human genome”) and to the ENCODE consortium’s failure to prevent this misinterpretation. To these critics, ignoring evolutionary definitions of function was a major misstep.

This debate can be easily extended to include TEs. TEs make up the vast majority of what is often referred to as ‘junk DNA’. Today, the term is mainly used—and abused—by the media, but it has deep roots in evolutionary biology. Regardless of the semantics, what evidence is needed to assign a TE with a function? Many TEs encode a wide range of biochemical activities that normally benefit their own propagation. For example, TEs often contain promoter or enhancer elements that hijack cellular RNA polymerases for transcription and autonomous elements encode proteins with various biochemical and enzymatic activities, all of which are necessary for the transposon to replicate. Do these activities make them functional?

The vast differences in TEs between species make standard approaches to establishing their regulatory roles particularly challenging. For example, intriguing studies on the impact of HERVs, particularly HERV-H, in stem cells and pluripotency must be interpreted using novel paradigms that do not invoke deep evolutionary conservation to imply function, as these particular ERVs are absent outside of the great apes. Evolutionary constraints can be measured at shorter time scales, including the population level, but this remains a statistically challenging task, especially for non-coding sequences. Natural loss-of-function alleles may exist in the human population, and their effect on fitness can be studied if their impact is apparent; however, these are quite rare and do not allow for systematic studies. It is possible to engineer genetic knockouts of a particular human transposable element (TE) locus to test its regulatory role; however, these are typically restricted to *in vitro* systems, especially when the orthologous TE does not exist in the model species. In this context, studying the impact of TEs in model species with powerful genome engineering tools and vast collections of mutants and other genetic resources, such as plants, fungi, and insects, will also continue to be extremely valuable.

Finally, a growing consensus is urging more rigor when assigning cellular function to TEs, particularly for the fitness benefit of the host. Indeed, a TE displaying biochemical activity (such as those bound by transcription factors or lying within open chromatin regions) cannot be equated to a TE that shows evidence of purifying selection at the sequence level or, when genetically altered, results in a deleterious or dysfunctional phenotype. Recent advances in editing and manipulating the genome and the epigenome en masse, yet with precision, including repetitive elements, offer the promise for a systematic assessment of the functional significance of TEs.

23.2.6: Summary

Transposable elements (TEs) are mobile DNA sequences capable of moving within and between genomes, a discovery that revolutionized our understanding of genome plasticity. First described by Barbara McClintock in maize and later validated by studies on bacterial insertion sequences, TEs are now known to constitute a significant fraction of genomes across all domains of life—from approximately 10% in some fish to over 80% in large plant genomes like maize.

TEs are classified into two major classes based on their transposition mechanisms. **Class I retrotransposons** (the “copy-and-paste” elements) mobilize via an RNA intermediate, which is reverse transcribed into DNA before integration. These include LTR retrotransposons (related to retroviruses), LINEs, SINEs, and other less common orders. In contrast, **Class II DNA transposons** (the “cut-and-paste” elements) move directly as DNA, typically excising from one location and reinserting at another. Additionally, novel subclasses such as Helitrons (which replicate via rolling-circle mechanisms) and Mavericks add further diversity.

TEs impact genome structure and function in multiple ways. They can disrupt gene coding regions or regulatory sequences upon insertion, causing mutations or altering gene expression. They also provide substrates for homologous recombination, leading to large-scale genomic rearrangements such as deletions, duplications, and inversions. Over evolutionary time, many TEs become inactivated by accumulating mutations, yet their remnants—often termed “junk DNA”—can be exapted to serve new functions. For

example, several key cellular proteins, such as the RAG recombinases involved in immune receptor diversity and centromere protein CENP-B, have evolved from TE sequences.

The host genome employs various mechanisms to control TE activity and minimize deleterious effects. Epigenetic modifications (like DNA methylation) and RNA interference pathways, including small RNA-mediated silencing, restrict TE expression. In some cases, TE activity is differentially regulated between the germline and somatic tissues, as somatic TE activity is generally more detrimental to cell function.

Furthermore, TEs contribute to genetic diversity not only through their mutagenic potential but also by donating sequences that form regulatory elements or new exons via exon shuffling. Horizontal transfer between species further spreads TEs, providing an evolutionary force that shapes genomes over both short and long timescales.

In summary, transposable elements are far more than genomic parasites; they are dynamic components of the genome that drive evolution, contribute to genetic variation, and even become co-opted for essential cellular functions. Understanding TEs requires an integration of biochemistry, genetics, and evolutionary biology, and underscores the complex interplay between mobile genetic elements and host genome stability.

23.2.7: References

1. Munoz-Lopez, M. and Garcia-Perez, J.L. (2010) DNA Transposons: Nature and Applications in Genomics. *Curr Genomics* 11(2):115-128. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2874221/>
2. Makalowski, W., Gotea, V., Pande, A. and Makalowska, I. (2019) Transposable Elements: Classification, Identification, and Their Use As a Tool For Comparative Genomics. In: Anisimova M. (eds) *Evolutionary Genomics. Methods in Molecular Biology*, vol 1910. Humana, New York, NY. https://doi.org/10.1007/978-1-4939-9074-0_6
3. Bourque, G., Burns, K.H., Gehring, M., Borbunova, V., Seluanov, A., Hammell, M., Imbeault, M., Izvak, Z., Levin, H.L., Macfarlan, T.S., Mager, D.L., Feschotte, C. (2018) Ten things you should know about transposable elements. *Genome Biology* 19: 199. Available at: <https://genomebiology.biomedcentral.com/articles/10.1186/s13059-018-1577-z#Fig1>

This page titled [23.2: DNA Transposable Elements](#) is shared under a [not declared](#) license and was authored, remixed, and/or curated by [Henry Jakubowski and Patricia Flatt](#).

- [Current page](#) by [Henry Jakubowski and Patricia Flatt](#) has no license indicated.
- [11.2: Diffusion Across a Membrane - Passive and Facilitated Diffusion](#) by [Henry Jakubowski and Patricia Flatt](#) has no license indicated.

CHAPTER OVERVIEW

24: DNA Metabolism

[Return to Fundamentals of Biochemistry](#)

[Search Fundamentals of Biochemistry](#)

[24.1: DNA Replication](#)

[24.2: DNA Mutations, Damage, and Repair](#)

[24.3: DNA Recombination](#)

This page titled [24: DNA Metabolism](#) is shared under a [not declared](#) license and was authored, remixed, and/or curated by [Henry Jakubowski and Patricia Flatt](#).

CHAPTER OVERVIEW

25: RNA Metabolism

[Return to Fundamentals of Biochemistry](#)

[Search Fundamentals of Biochemistry](#)

[25.1: DNA-Dependent Synthesis of RNA](#)

[25.2: RNA Processing](#)

[25.3: RNA-Dependent Synthesis of RNA and DNA](#)

[25.4: 26.4 References](#)

This page titled [25: RNA Metabolism](#) is shared under a [not declared](#) license and was authored, remixed, and/or curated by [Henry Jakubowski and Patricia Flatt](#).

25.2: RNA Processing

Learning Goals (ChatGPT o3-mini)

- **Compare and Contrast RNA Processing in Prokaryotes and Eukaryotes:**
 - Describe why prokaryotic mRNA undergoes minimal processing compared to eukaryotic mRNA.
 - Explain how transcription–translation coupling in bacteria (e.g., role of the Shine-Dalgarno sequence, NusG, and NusA) contrasts with the compartmentalized nature of eukaryotic gene expression.
- **Understand the Mechanisms of 3' End Formation in Eukaryotic mRNA:**
 - Outline the steps involved in cleavage and polyadenylation (CPA), including recognition of polyadenylation signals (PAS) and the role of auxiliary sequence elements.
 - Identify and describe the functions of the key protein complexes (e.g., CPSF, CstF, CFI, CFII) that constitute the cleavage/polyadenylation apparatus.
- **Analyze the Structural and Functional Features of the Cleavage/Polyadenylation Machinery:**
 - Compare the subunit composition and roles of the yeast CPF and human CPSF complexes.
 - Explain how interactions among protein subcomplexes (including RNA-binding motifs and conserved residues) regulate endonuclease and polyadenylation activities.
- **Master the Process of 5' Cap Formation and Its Functional Implications:**
 - Describe the sequential enzymatic steps of 5' cap formation (triphosphate removal, guanylyltransferase action, and methyl transfer) and the molecular features of the 7-methylguanylate cap.
 - Explain how the cap and cap-binding complexes (CBC and later eIF4E/eIF4G) protect mRNA from degradation, promote nuclear export, and enhance translation.
- **Detail the Molecular Basis of Pre-mRNA Splicing:**
 - Outline the sequential assembly of the spliceosome, including the roles of U1, U2, U4/U5/U6 snRNPs, and other splicing factors in recognizing exon/intron boundaries.
 - Describe the structural transitions from early spliceosomal complexes (E, A) to the pre-catalytic (B) and catalytic (B* and C) spliceosome, including the two transesterification reactions that remove introns and ligate exons.
- **Evaluate the Regulation and Impact of Alternative Splicing:**
 - Explain how cis-acting elements (ESE, ISE, ESS, ISS) and trans-acting factors (SR proteins, hnRNPs) govern the choice of splice sites.
 - Discuss how alternative splicing contributes to proteomic diversity and the regulation of gene expression in different cellular contexts.
- **Integrate Concepts of mRNA Processing with Overall Gene Expression:**
 - Connect how post-transcriptional modifications (capping, polyadenylation, splicing) affect mRNA stability, localization, and translation efficiency.
 - Assess the broader implications of RNA processing defects on cellular function and potential disease states.

These goals provide a roadmap for mastering the detailed mechanisms and regulatory pathways underlying mRNA processing in both prokaryotic and eukaryotic systems.

Post-transcriptional modifications of rRNA and tRNA will be topics of Chapter 27 as their structure and function in protein synthesis will be a focal point. Thus, this section will focus on post-transcriptional modifications of mRNA. We'll spend most of our time on eukaryotic RNA processing.

25.2.1: Prokaryotic RNA Processing

First, let's take a brief look at a fascinating feature of transcription in bacterial cells. Bacterial cells do not have extensive post-transcriptional modifications of mRNA primarily because transcription and translation are coupled processes. Bacterial cells lack the physical barrier of a nucleus, which allows transcription and translation machinery to function at the same time, enabling the

concurrent translation of an mRNA while it is being transcribed (Fig 26.2.1). As mRNA is synthesized in prokaryotes, a ribosome binding motif called the Shine-Dalgarno sequence, located in the 5' untranslated region of the mRNA emerges early, allowing the ribosome to bind and translation to occur. In addition, the protein N-utilization Substance, better known as **NusG**, plays a critical role. NusG has three separate domains, and the functions of two of them are known. The NusG N-terminal domain (NusG-NTD) can bind to RNAP, whereas the C-terminal domain (NusG-CTD) can combine with the NusE (RpsJ) component of ribosomes. These two functions of NusG enable transcription to be coupled with translation. NusG CTD can also bind to Rho to terminate transcription, as shown in Figure [Math Processing Error].

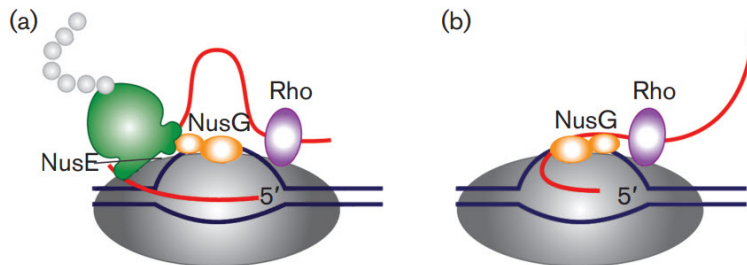


Figure [Math Processing Error]: **The roles of NusG in transcription/translation coupling.** (a) Composition of an active RNAP complex. RNAP is shown in dark grey, DNA in blue, and nascent RNA in red. The ribosome is shown in green with the nascent polypeptide chain in light grey; the bulge in the small subunit denotes the location of NusE (RpsJ). NusG is shown in orange: its shape denotes two functional sections. The larger section denotes the N-terminal domain, which binds to RNAP. The smaller section denotes the C-terminal domain, which interacts with NusE in situ. Rho is shown in purple. (b) After the translation is completed, NusG remains bound to RNAP and may also bind to Rho through the C-terminal domain leading to the termination of transcription. Figure from: Cortes, T., and Cox, R.A. (2015) *Microbiology* 161:719-728.

Another protein, NusA, slows RNAP, in contrast to NusG, which increases its processivity, as reflected by the length of the RNA made before RNAP falls off.

25.2.2: Eukaryotic RNA Processing

In multicellular organisms, almost every cell contains the same genome, yet complex spatial and temporal diversity is observed in gene transcripts. This is achieved through multiple levels of processing, leading from gene to protein, with RNA processing being an essential stage. Following the transcription of a gene by RNA polymerases to produce a primary mRNA transcript, further processing is required to make a stable and functional mature RNA product. This involves various processing steps, including RNA cleavage at specific sites, intron removal, called **splicing**, which substantially increases the transcript repertoire, and the addition of a 5'CAP. Another crucial feature of the RNA processing of most genes is the generation of 3' ends through an initial endonucleolytic cleavage, followed in most cases by the addition of a poly(A) tail, a process termed 3' end **cleavage and polyadenylation (CPA)**.

25.2.3: Cleavage and 3'-Polyadenylation (CPA)

Polyadenylation is a required step for the correct termination of nearly all mRNA transcripts. Except for replication-dependent histone genes, metazoan protein-encoding mRNAs contain a uniform 3' end consisting of a stretch of adenosines. In addition to determining the correct transcript length at transcription termination, the poly(A) tail helps to ensure the translocation of the nascent RNA molecule from the nucleus to the cytoplasm, enhances translation efficiency, acts as a signal feature for RNA degradation, and thereby contributes to the production efficiency of a protein.

Cleavage and polyadenylation (CPA) are carried out by the **cleavage/polyadenylation apparatus (CPA)**, a multi-subunit 3' end processing complex, which involves over 80 proteins, comprised of four core protein subcomplexes, as shown in Figure [Math Processing Error]. These consist of

1. **cleavage and polyadenylation specificity factor (CPSF)**, comprised of proteins CPSF1-4, factor interacting with PAPOLA and CPSF1 (FIP1L1), and WD repeat domain 33 (WDR33) (shown in green below);
2. cleavage stimulation factor (CstF), a trimer of CSTF1-3 (shown in red below);
3. cleavage factor I (CFI), a tetramer of two small nudix hydrolase 21 (NUDT21) subunits, and two large subunits of CPSF7 and/or CPSF6 (shown in orange in Figure 26.2.2 A); Note: Nudix are named for **nucleoside diphosphates hydroxylases**.

4. cleavage factor II (CFII), composed of cleavage factor polyribonucleotide kinase subunit 1 (CLP1) and PCF11 cleavage and polyadenylation factor subunit (PCF11) (shown in yellow in Figure 26.2.2 A). Additional factors include symplekin, the **poly(A) polymerase (PAP)**, and the nuclear poly(A) binding proteins such as poly(A) binding protein nuclear 1 (PABPN1).

CPA is initiated by this complex, which recognizes specific sequences within the nascent pre-mRNA transcripts, termed **polyadenylation signals (PAS)**. The PAS sequence normally consists of either a canonical 6-base sequence, the **AATAAA hexamer**, or a close variant usually differing by a single nucleotide (e.g., ATTAAA, TATAAA). It is located 10 to 35 nucleotides upstream of the cleavage site (CS) usually consisting of a CA dinucleotide. The PAS is also determined by surrounding auxiliary elements, such as upstream U-rich elements (USE), as well as downstream U-rich and G-rich sequences (DSE).

As soon as the nascent RNA molecule emerges from RNA polymerase II (RNA Pol II), the CPSF complex is recruited to the PAS **AATAAA hexamer** through numerous interactions. Upon successful assembly of this macromolecular machinery, CPSF3 performs endonucleolytic cleavage, followed by the non-templated addition of approximately 50-100 nucleotides.

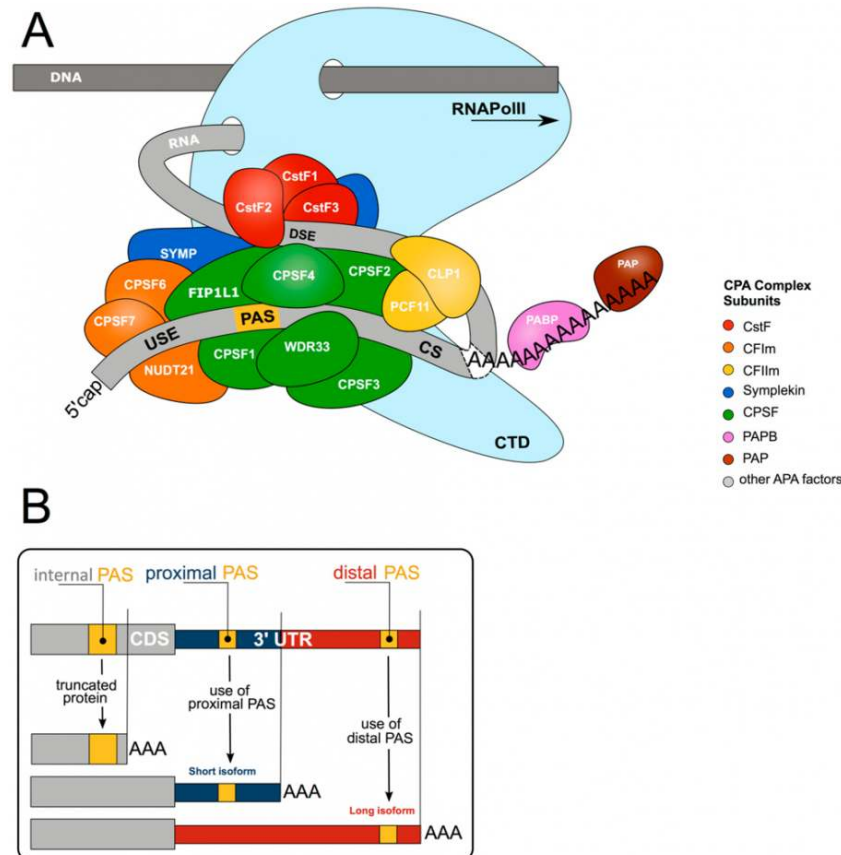


Figure [Math Processing Error]: The human core 3' end RNA processing machinery and impact on alternative polyadenylation. Nourse, J., et. al. (2020) *Biomolecules* 10(6):915

Panel (A) shows that the core 3' end processing machinery consists of complexes composed of multiple trans-acting proteins interacting with RNA via multiple cis-elements (USE = upstream sequence element; PAS = poly(A) signal; CS = cleavage site; DSE = downstream sequence element; CTD = C-terminal domain). Upon co-transcriptional assembly of these complexes, RNA cleavage and polyadenylation occur to form the 3' end of the nascent RNA molecule.

Panel (B) shows that more than 70% of all genes harbor more than one polyadenylation signal (PAS). This gives rise to transcript isoforms differing at the mRNA 3' end. While alternative polyadenylation (APA) in 3'UTR changes the properties of the mRNA (stability, localization, translation), internal PAS usage (in introns or the coding sequence (CDS)) changes the C-termini of the encoded protein, resulting in different functional or regulatory properties.

Alternative polyadenylation (APA) occurs when more than one polyadenylation signal (PAS) is present within a pre-mRNA, providing an additional level of complexity in CPA-mediated RNA processing (Figure 26.2.2B). Early studies revealed that a

significant portion of genes undergo alternative polyadenylation (APA), and with the advent of next-generation RNA sequencing technologies, the large-scale regulation of genes has become apparent, with approximately 70% of the transcriptome exhibiting APA regulation. As APA determines 3'UTR content and thus the regulatory features available to the mRNA, changes in the APA profile of a gene can have enormous impacts on expression.

For those trying to understand the structure and mechanism of the **cleavage/polyadenylation apparatus (CPA)**, it is especially frustrating that different names are given to the constituents that comprise it, especially when comparing the proteins from different organisms. Hence, it is useful to see multiple representations of the complex. Figure [Math Processing Error] shows a different cartoon representation of the cleavage and polyadenylation reactions. Note again the number of colored subcomplexes within the CPA, as well as the different abbreviations shown for the individual proteins. This cartoon diagram is useful in visualizing the different steps involved.

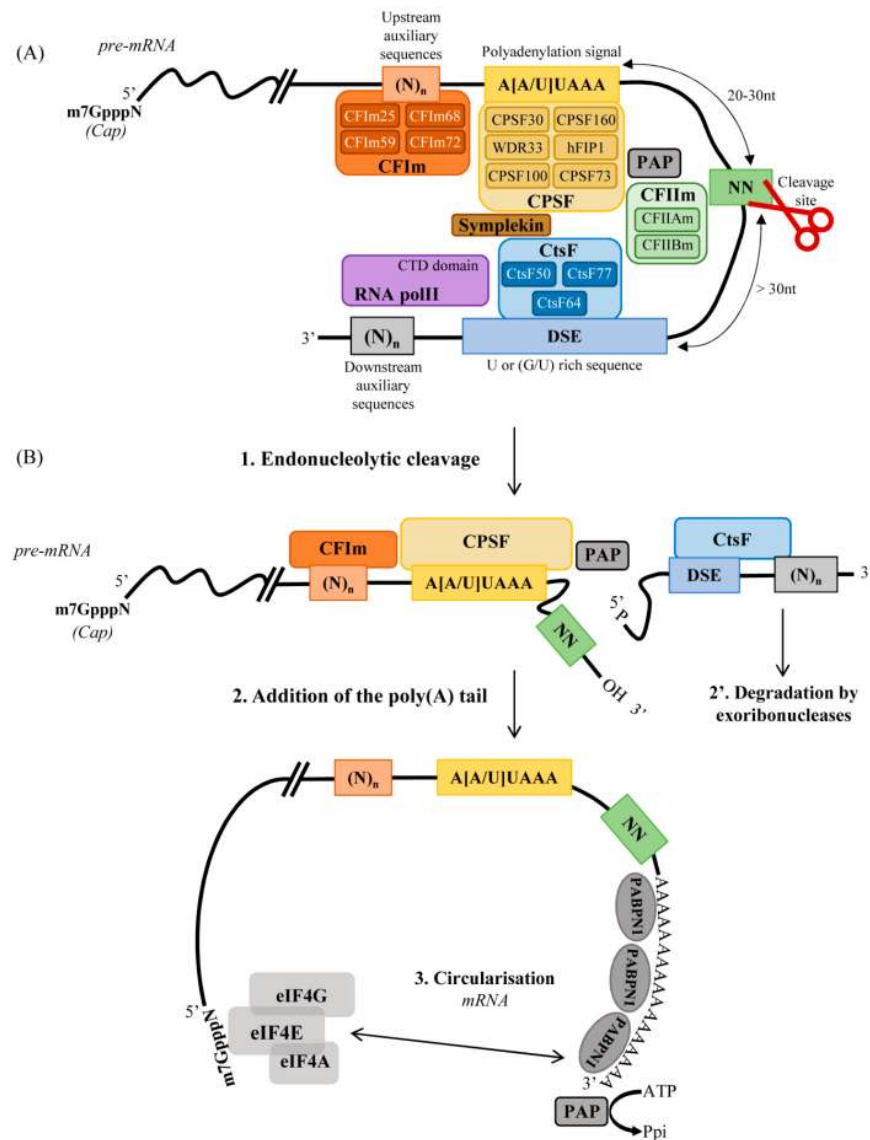


Figure [Math Processing Error]: Cis-regulatory sequence elements and protein factors involved in cleavage and polyadenylation. Marsollier, A.-C.; Joubert, R.; Mariot, V.; Dumonceaux, J. Targeting the Polyadenylation Signal of Pre-mRNA: A New Gene Silencing Approach for Facioscapulohumeral Dystrophy. *Int. J. Mol. Sci.* **2018**, *19*, 1347. <https://doi.org/10.3390/ijms19051347>. [Creative Commons Attribution License](#)

Panel (A) shows that the specificity and efficiency of 3' end processing are determined by the binding of more than 80 RNA-binding proteins to regulatory *cis*-acting RNA sequence elements including the polyadenylation signal (PAS) A[A/U]UAAA; the cleavage site (represented by NN) and the downstream sequence element (DSE). Auxiliary sequences can be found near the

polyadenylation signal or the DSE. The core processing complex, which is sufficient for the cleavage and polyadenylation, is composed of approximately 20 proteins, distributed in 8 complexes: the cleavage and polyadenylation specificity factor (CPSF), the cleavage stimulation factor (CstF); the mammalian cleavage factors I (CFIm) and the mammalian cleavage factors II (CFIIm); the single protein poly(A) polymerase (PAP); the single protein poly(A)-binding protein nuclear 1 (PABPN1); the single protein RNA polymerase II large subunit (Pol II); and the symplekin. Subunits of the different factors are indicated.

Panel (B) shows how CPSF and CstF are co-transcriptionally recruited to the poly(A) signal and the DSE, respectively, causing an endonucleolytic cleavage of the pre-mRNA between the PAS and the DSE at the cleavage site. Two fragments are generated: one fragment with a free 5'phosphate group, which is rapidly degraded by exoribonucleases, and one fragment with a free 3'hydroxyl group on which PAP will add 250 adenines. The newly synthesized poly(A) tail is covered by PABPN1, allowing mRNA circularization and stabilization.

Now let's look at the structure of some of the complexes of the **cleavage/polyadenylation apparatus (CPA)**.

In yeast, 3' processing is carried out by the cleavage and polyadenylation factor (CPF), also known as CPSF in humans. Upon endonuclease cleavage, the RNA bound to RNA polymerase II is cleaved into two pieces, as shown in the figure above. The main mRNA now has a 3'-OH, which is the site of polyadenylation. The minor cleavage fragment has a 5'-phosphate which gets degraded by the exonuclease Rat1, which, as it cleaves the minor product, helps displace RNA polymerase II and helps to stop transcription.

In yeast, the CPF consists of 14 subunits, which are organized into polymerase, nuclease, and phosphatase subcomplexes, also known as "modules". The polymerase module, as the name implies, has the poly(A) polymerase, Pap1. Table [\[Math Processing Error\]](#) below shows some components of the polymerase module of both yeast CPF and human CPSF.

yeast Polymerase Module of CPF	Human mammalian polyadenylation specificity factor (mPSF) or CPSF
Cft1	CPSF160
Pfs2	WDR33
Yth1 - RNA binding subunit	CPSF30 - RNA binding subunit
Fip1 - Pap1 binding subunit	FIP1 - Pap1 binding subunit

Table [\[Math Processing Error\]](#) Some components of the polymerase module of both yeast CPF and human CPSF.

The **nuclease module** has an endonuclease (Ysh1) and a Mpe1 protein, which facilitate the cleavage site selection and polyadenylation. Table [\[Math Processing Error\]](#) below shows some components of the polymerase module in yeast and humans

yeast	human
endonuclease Ysh1	endonuclease CPSF73
pseudo-nuclease Cft2	pseudo-nuclease CPSF100
multidomain protein Mpe1	multidomain protein RBBP6

Table [\[Math Processing Error\]](#): Some components of the polymerase module in yeast and humans

Part of the Cft2 called the **yeast polymerase module interacting motif (yPIM)**, as its name implies, interacts with the polymerase module, in part through the interaction of key and conserved aromatic residues in it (F537, Y549, and F558) with a hydrophobic binding site in Cft1 and Pfs1. These interactions are key in activating and regulating the endonuclease and polyadenylation activities, thereby controlling the termination of transcription.

Figure [\[Math Processing Error\]](#) shows an [interactive iCn3D model](#) of the yeast cleavage and polyadenylation specificity factor (CPF) polymerase module in complex with Mpe1, the yPIM of Cft2 and the pre-cleaved CYC1 RNA (7ZGR)



Figure [Math Processing Error]: Yeast cleavage and polyadenylation specificity factor (CPF) polymerase module in complex with Mpe1, the yPIM of Cft2 and the pre-cleaved CYC1 RNA (7ZGR). (Copyright; author via source). Click the image for a popup or use this external link: <https://structure.ncbi.nlm.nih.gov/1...ePGK2jiiuum7A7>

The different parts of the complex are colored as shown below.

- Yth1 (mRNA 3' processing protein): magenta
- CFT1: green
- MPE1: orange
- Polyadenylation subunit 2 (Pfs2): yellow
- cleavage factor 2 protein (CF2P): cyan
- precleaved RNA CYC1: gray spacefill (backbone) with CPK-colored bases

25.2.4: 5'-CAP Formation

In eukaryotes, the 5' cap, found on the 5' end of the eventual mRNA molecule, consists of a guanine nucleotide connected to the mRNA via an unusual 5' to 5' triphosphate linkage, as shown in Figure [Math Processing Error]. This guanosine is methylated on the 7 position directly after capping *in vivo* by a methyltransferase. It is referred to as the 7-methylguanylate cap, abbreviated m⁷G.

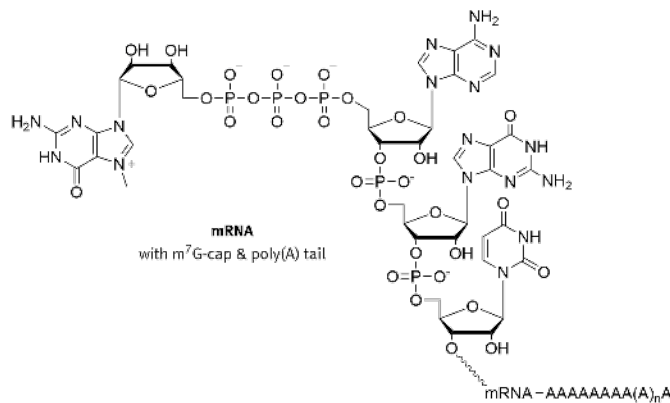


Figure [Math Processing Error]: Structure of the 7-methylguanylate CAP. Brisbane

In multicellular eukaryotes and some viruses, further modifications exist, including the methylation of the 2' hydroxy-groups of the first two ribose sugars of the 5' end of the mRNA. Cap-1 has a methylated 2'-hydroxy group on the first ribose sugar, while cap-2 has methylated 2'-hydroxy groups on the first two ribose sugars. The 5' cap is chemically similar to the 3' end of an RNA molecule (the 5' carbon of the cap ribose is bonded, and the 3'-OH is unbonded). This provides significant resistance to 5' exonucleases.

Small nuclear RNAs (snRNAs) contain unique 5'-caps. Sm-class snRNAs are found with 5'-trimethylguanosine caps, while Lsm-class snRNAs are found with 5'-monomethylphosphate caps. In bacteria, and potentially also in higher organisms, some RNAs are capped with NAD⁺, NADH, or 3'-dephospho-coenzyme A. In all organisms, mRNA molecules can be decapped in a process known as messenger **RNA decapping**.

For capping with 7-methylguanylate, the **capping enzyme complex (CEC)** binds to RNA polymerase II before transcription initiation. As soon as the 5' end of the new transcript emerges from RNA polymerase II, the CEC initiates the capping process (this type of mechanism ensures capping, similar to polyadenylation). The enzymes for capping can only bind to RNA polymerase II that is engaging in mRNA transcription, ensuring the specificity of the m⁷G cap almost entirely to mRNA.

The 5' cap has four main functions:

1. Regulation of nuclear export
2. Prevention of degradation by exonucleases
3. Promotion of translation (see ribosome and translation)
4. Promotion of 5' proximal intron excision

In addition to the polyA tail, the nuclear export of RNA is regulated by the cap-binding complex (CBC), which binds to 7-methylguanylate-capped RNA, as shown in Figure [Math Processing Error]. The nuclear pore complex then recognizes the CBC, and the mRNA is exported out of the nucleus. Once in the cytoplasm after the pioneer round of translation, the CBC is replaced by the translation factors eIF4E and eIF4G of the eIF4F complex. This complex is then recognized by other translation initiation machinery, including the ribosome, aiding in translation efficiency.

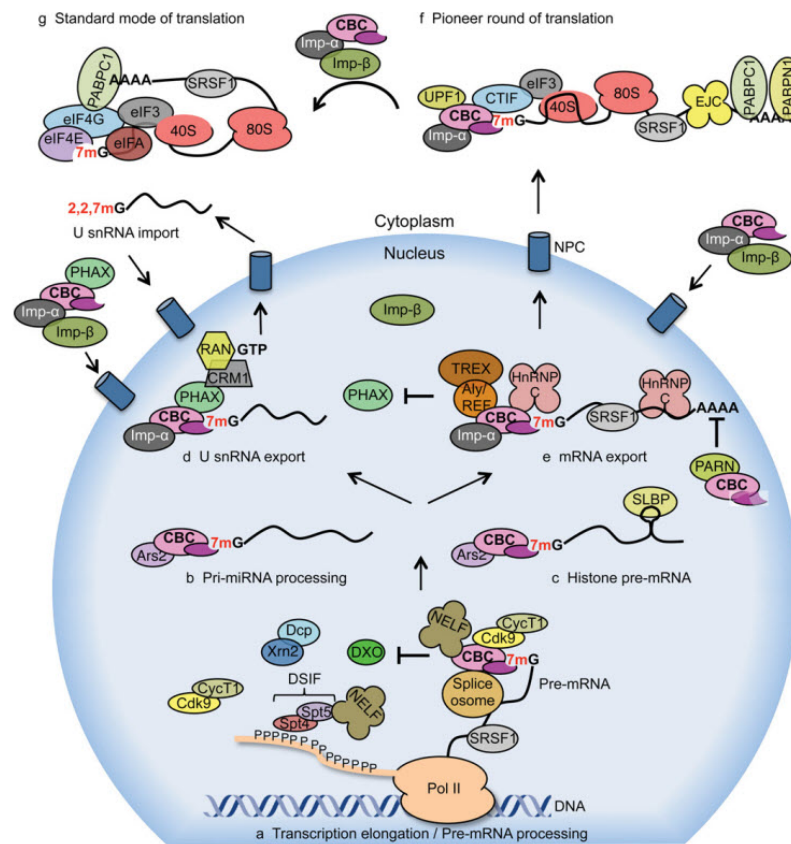


Figure [Math Processing Error]: Importance of the 5-CAP during the lifespan of a mRNA Transcript. [Gonatopoulos-Pournatzis, T. and Crowing V. \(2014\) Biochemical Journal 457\(2\):231-42.](#)

Panel (a) CBC is required for pre-mRNA processing. The co-transcriptional binding of CBC to 7mG prevents the decapping activities of pre-mRNA degradation complexes [DXO (decapping exoribonuclease) and Dcp (decapping mRNA) Xrn2 (5'–3' exoribonuclease 2)] and promotes pre-mRNA processing. CBC recruits P-TEFb [Cdk9/Cyclin T1 (CycT1)] to transcription initiation sites of specific genes, promoting phosphorylation of the RNA pol II CTD at Ser2 residues. This results in the recruitment of splicing factors, including SRSF1, which regulates both constitutive and alternative splicing events. Furthermore, CBC interacts with splicing machinery components, resulting in spliceosomal assembly. CBC interacts with NELF and promotes pre-mRNA processing of replication-dependent histone transcripts.

Panel (b) CBC forms a complex with Ars2 and promotes miRNA biogenesis by mediating pri-miRNA processing.

Panel (c) CBC/Ars2 promotes pre-mRNA processing of replication-dependent histone transcripts.

Panel (d) CBC promotes the export of U snRNA. CBC interacts with PHAX, which recruits export factors including CRM1 and RAN·GTP.

Panel (e) CBC promotes the export of mRNA. For the export of transcripts over 300 nucleotides, hnRNP C interacts with CBC and inhibits the interaction between CBC and PHAX, allowing the CBC to interact with TREX and the transcript to be translocated to the cytoplasm. CBC interacts with the PARN deadenylase and inhibits its activity, protecting mRNAs from degradation.

Panel (f) CBC mediates the pioneer round of translation. Cbp80 interacts with CTIF, which recruits the 40S ribosomal subunit via eIF3 to the 5' end of the mRNA for translation initiation. Upon binding of importin-β (Imp-β) to importin-α (Imp-α), mRNA is released from CBC and binds to eIF4E for the initiation of the standard mode of translation. CBC-bound mRNP components not found in eIF4E-bound mRNPs are CTIF, exon junction complex (EJC), and PABPN1.

Panel (g) The standard mode of translation is mediated by eIF4E cap-binding protein. eIF4E is a component of the eIF4F complex, which promotes translation initiation.

Capping with 7-methylguanylate prevents 5' degradation in two ways. First, degradation of the mRNA by 5' exonucleases is prevented by functionally looking like a 3' end. Second, the CBC and eIF4E/eIF4G block the access of decapping enzymes to the

cap. This increases the half-life of the mRNA, which is essential in eukaryotes, as the export and translation processes take a significant amount of time.

The mechanism that promotes 5' proximal intron excision during splicing is not well understood; however, the 7-methylguanylate cap appears to loop around and interact with the spliceosome, potentially playing a role in the splicing process.

The decapping of a 7-methylguanylate-capped mRNA is catalyzed by the decapping complex made up of at least Dcp1 and Dcp2, which must compete with eIF4E to bind the cap. Thus the 7-methylguanylate cap is a marker of an actively translating mRNA and is used by cells to regulate mRNA half-lives in response to new stimuli. During the decay process, mRNAs may be sent to P-bodies. P-bodies are granular foci within the cytoplasm that contain high levels of exonuclease activity.

25.2.4.1: Triphosphatase and guanylyltransferase

In capping the new mRNA, three different enzymes act sequentially:

- A phosphatase cleaves a terminal phosphate from the 5' end, which has three phosphates at the start, leaving two phosphates (the yeast triphosphatase is called Cet1);
- a GMP is added to the remaining diphosphate on the 5'-end to form a triphosphate in a reverse direction as shown in Figure 5 above (in yeast, the RNA guanylyltransferase is called Ceg1);
- a methyl group is added to the N7 of the guanine base by a methylase, also called a methyl transferase

These three enzymes are localized to a part of RNA polymerase that is highly phosphorylated, positioning them at the correct location for Cap formation. In yeast, the

Figure [Math Processing Error] shows an interactive iCn3D model of the *Saccharomyces cerevisiae* Cet1 (the triphosphatase)-Ceg1 (the guanylyltransferase) mRNA Capping Apparatus (3KYH)

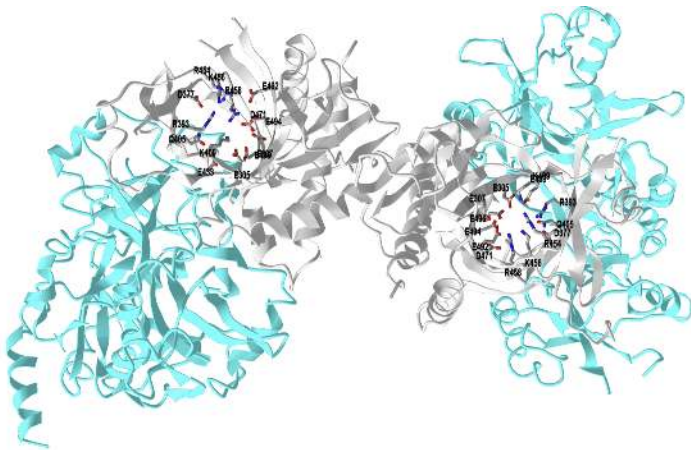


Figure [Math Processing Error]: *Saccharomyces cerevisiae* Cet1-Ceg1 mRNA Capping Apparatus (3KYH) . (Copyright; author via source). Click the image for a popup or use this external link: <https://structure.ncbi.nlm.nih.gov/...UWQ32mQrcF1eN6>

The two enzymes exist as a heterotetramer of two homodimers. Two beta (Cet1-triphosphatase) subunits are shown in gray and the two alpha (Ceg1-guanylyltransferase) subunits are shown in cyan. The active site residues are shown in CPK-colored sticks and labeled. The two enzymes interact with yeast RNAP II primarily through Ceg1. Specifically, the Ceg1 oligonucleotide domain interacts with a motif, WAQKW (247-251), on Cet1. A conformational change in the flexible linker following that motif enables capping to occur.

25.2.4.2: Methyl transferase

The next step is the methyl transfer to the N7 of guanine in the 5' end of the cap. Figure [\[Math Processing Error\]](#) shows an [interactive iCn3D model](#) of the Structure of a bacterial mRNA cap (Guanine-n7) methyltransferase (1RI1)

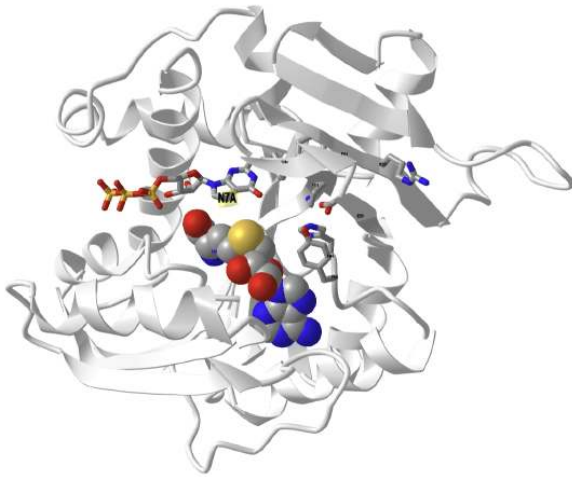


Figure [\[Math Processing Error\]](#): Structure of a bacterial mRNA cap (Guanine-n7) methyltransferase (1RI1) with bound S-adenosyl-homocysteine. (Copyright; author via source). Click the image for a popup or use this external link: <https://structure.ncbi.nlm.nih.gov/i...YfaJ6W93zSDaX7>

7-Methyl-guanosine-5'-triphosphate-5'-guanosine (GTG) is shown in sticks. The N7 nitrogen of guanine is labeled. S-Adenosyl-L-homocysteine, the leaving group after methylation by S-adenosyl-L-methionine (SAM), is shown in spacefill. Key amino acid side chains in the active site are labeled (in small letters). The structure is most consistent with an in-line methyl transfer from SAM to the attacking nucleophile, the N7 of guanine. Specificity in most capping methylases is achieved through noncovalent interactions, primarily base stacking, involving two aromatic amino acid groups (an example of π - π stacking) or one aromatic and one nonpolar side chain. In the bacterial example shown above, the interactions include hydrophobic and hydrogen bonding interactions using Y284, F24, P175, E225, H144, and Y145.

25.2.4.3: Decapping

"What goes up must come down!" If the mRNA is capped during synthesis, there must be an enzyme to decap it. Figure [\[Math Processing Error\]](#) shows an [interactive iCn3D model](#) of the yeast mRNA decapping enzyme Dcp1-Dcp2 complex in the ATP bound closed conformation (2QKM)

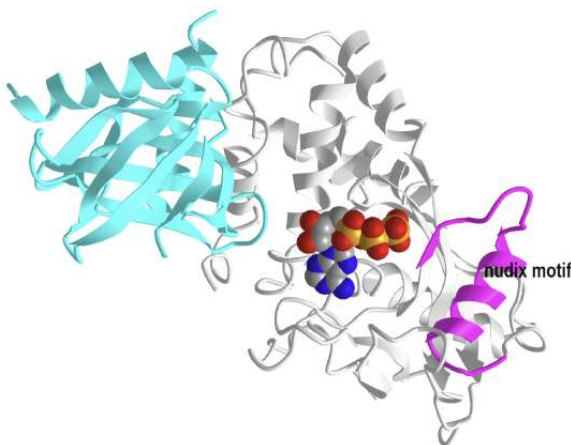


Figure [\[Math Processing Error\]](#): Yeast mRNA decapping enzyme Dcp1-Dcp2 complex in the ATP bound closed conformation (2QKM). (Copyright; author via source). Click the image for a popup or use this external link: <https://structure.ncbi.nlm.nih.gov/i...hBPokeNeMfcGUA>

The rendering is colored as shown below:

- a chain in cyan
- b chain in gray
- nudix motif in magenta
- ATP in spacefill

Dcp2p, like many other enzymes, exhibits an open (inactive) and closed (active) conformation, indicating that a conformational change between these two states regulates decapping. The ATP binding site demarcates the active site. The Dcp1 protein probably functions to stabilize the closed state.

25.2.5: mRNA Splicing

Eukaryotic genes that encode polypeptides are composed of coding sequences called **exons** (*ex-on* signifies that they are *expressed*) and intervening sequences called **introns** (*int-ron* denotes their *intervening* role). Transcribed RNA sequences corresponding to introns do not encode regions of the functional polypeptide and are removed from the pre-mRNA during processing. All of the intron-encoded RNA sequences must be completely and precisely removed from a pre-mRNA before protein synthesis so that the exon-encoded RNA sequences are properly joined together to code for a functional polypeptide. If the process errs by even a single nucleotide, the sequences of the rejoined exons would be shifted, and the resulting polypeptide would be nonfunctional. The process of removing intron-encoded RNA sequences and reconnecting those encoded by exons is called **RNA splicing**. Intron-encoded RNA sequences are removed from the pre-RNA while it is still in the nucleus. Although they are not translated, introns appear to have various functions, including regulating gene expression and facilitating mRNA transport. Upon completion of these modifications, the **mature transcript**, which encodes a polypeptide, is transported out of the nucleus and destined for the cytoplasm for translation. Introns can be spliced out in different ways, resulting in various exons being included or excluded from the final mRNA product. This process is known as **alternative splicing**. The advantage of alternative splicing is that different types of mRNA transcripts can be generated, all of which are derived from the same DNA sequence. In recent years, it has been shown that some archaea can also splice their pre-mRNA.

The spliceosome catalyzes the **splicing reaction**. This is a macromolecular complex formed by five small nuclear ribonucleoproteins (snRNPs), termed U1, U2, U4, U5, and U6, and approximately 200 proteins, as shown in Figure [Math Processing Error]. Each of these snRNPs contains snRNAs that can interact with each other through intrastrand hydrogen bonding, which helps localize the snRNPs to the complexes. The assembly of the spliceosome on pre-mRNA includes the binding of U1 snRNP, U2 snRNP, the pre-formed U4/U6-U5 triple snRNP, and the Prp19 complex. This assembly occurs through the recognition of several sequence elements on the pre-mRNA that define the exon/intron boundaries, which include the 5' and 3' splice sites (SS), the associated 3' sequences for intron excision, the polypyrimidine (Py) tract, and the branch point sequence (BPS). The assembly of the spliceosome during the process is depicted in Figure [Math Processing Error].

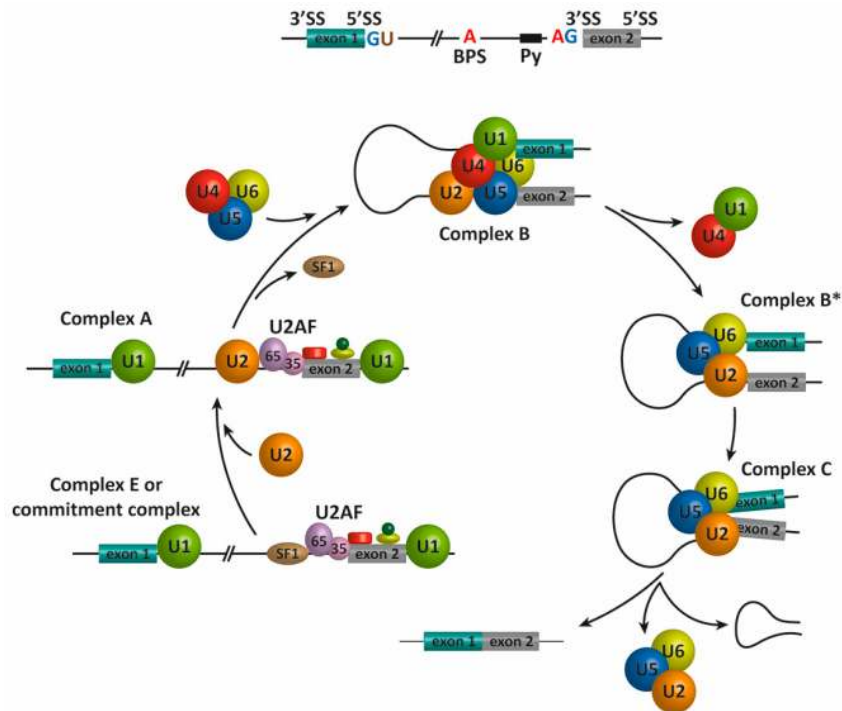


Figure [Math Processing Error]: Schematic representation of the spliceosome assembly and pre-mRNA splicing. Suñé-Pou, M., et. al. (2017) *Genes* 8(3):87 .

In the first step of the splicing process, the 5' splice site (GU, 5' SS) is bound by the U1 snRNP, and the splicing factors SF1/BBP and U2AF cooperatively recognize the branch point sequence (BPS), the polypyrimidine (Py) tract, and the 3' splice site (AG, 3' SS) to assemble complex E. The binding of the U2 snRNP to the BPS results in the pre-spliceosomal complex A. Subsequent steps lead to the binding of the U4/U5–U6 tri-snRNP and the formation of complex B. Complex C is assembled after rearrangements that detach the U1 and U4 snRNPs to generate complex B*. Complex C is responsible for the two transesterification reactions at the SS. Additional rearrangements result in the excision of the intron, which is removed as a lariat RNA, and ligation of the exons. The U2, U5, and U6 snRNPs are then released from the complex and recycled for subsequent rounds of splicing.

There are different ways that pre-spliceosomes convert to full spliceosomes. The interactions of the small ribonucleoproteins U1 and U2, which bind at the 5' and 3' ends of the introns, respectively, are key. When multiple introns exist, a few different pathways occur. The interactions can be upstream (cross-introns) and downstream (cross-exons), as shown in Figure [Math Processing Error].

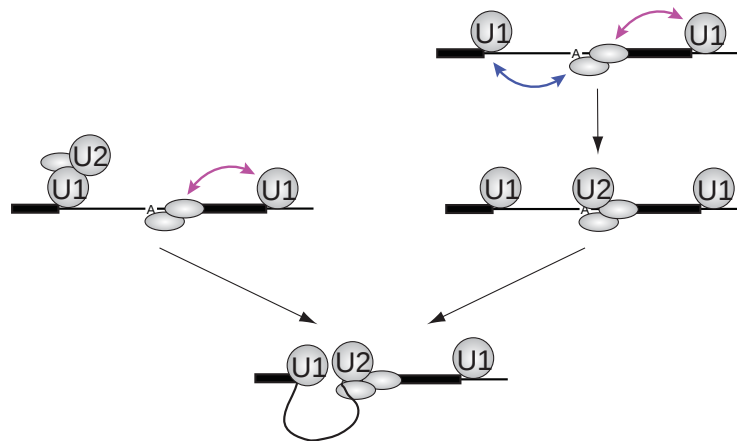


Figure [Math Processing Error]: Implications of synergistic U2 recruitment for the mechanism of exon and intron recognition. Braun et al. (2018) . Synergistic assembly of human pre-spliceosomes across introns and exons. *eLife* 7:e37751.

<https://doi.org/10.7554/eLife.37751>. Creative Commons Attribution License

The left-hand side of the figure illustrates how the interaction between U1 and U2 can occur across a single intron to form the mature spliceosome, highlighting the U1-U2 interaction. The colored arrow indicates possible interactions that either accelerate or inhibit the U1-U2 mature interactions.

The right-hand side shows the formation of an intermediate cross-exon U2-U1 pair in which the exons that flank two introns are first delineated, which then leads to the physical U1-U2 interaction in the mature spliceosome when cross-intron interactions occur.

Now let's move to show the actual structures of two of the spliceosome structures shown in Figure 10 above. Figure *[Math Processing Error]* shows an [interactive iCn3D model](#) of the human fully-assembled precatalytic spliceosome (pre-B complex) (6QX9)

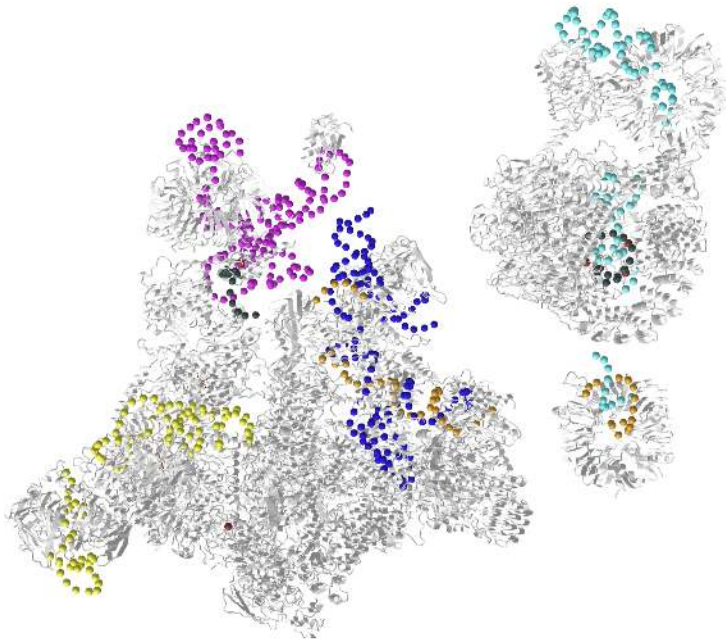


Figure *[Math Processing Error]*: Human fully-assembled precatalytic spliceosome (pre-B complex) (6QX9). (Copyright; author via source). Click the image for a popup or use this external link: <https://structure.ncbi.nlm.nih.gov/i...zLLvMh82m5u9M8>

The full structure of the spliceosome is almost too complex to comprehend, even when displayed in a molecular model. It's better grasped by the cartoon figure shown previously. To avoid unnecessary visual complexity, the protein subunits in the iCn3D model of the human spliceosome are shown in a gray cartoon, and only the RNA molecules in the U_n small ribonucleoproteins are shown in color, as described below. The resolution is such that each nucleotide in the RNA polymers is shown just as a single-colored sphere, not connected to the next nucleotide in the RNA.

- U1 snRNA - magenta
- U6 snRNA - orange
- U5 snRNA - yellow
- U2 snRNA - cyan
- U4 snRNA - blue
- AdML pre-mRNA - black

Since it has U1, U2, U4, U5, and U6, it most closely resembles Complex B (or an immediate precursor pre-B of it) as shown in Figure 10. This form occurs before U1 snRNP dissociates. A helix from the 5'-single-stranded U1 snRNA inserts into a helicase in the complex between two RecA domains, which bind ATP and unfold the nucleic acids. This allows the 5' single-stranded RNA to form interactions with an ACAGAGA sequence (mobile loop) in U6. These conformational changes enable the eventual separation of the U4 and U6 snRNAs, thereby freeing the snRNA in U6 from the catalytic site. The B complex itself does not have a functioning active site. After the dissociation of U1 and U4 snRNPs and the binding of another 20 or so proteins, the active spliceosome is formed.

Now, let's examine the final catalytic structure, Complex C, which comprises only three of the snRNPs: U2, U5, and U6. Figure *[Math Processing Error]* shows an [interactive iCn3D model](#) of the Human C Complex Spliceosome (7A5P)

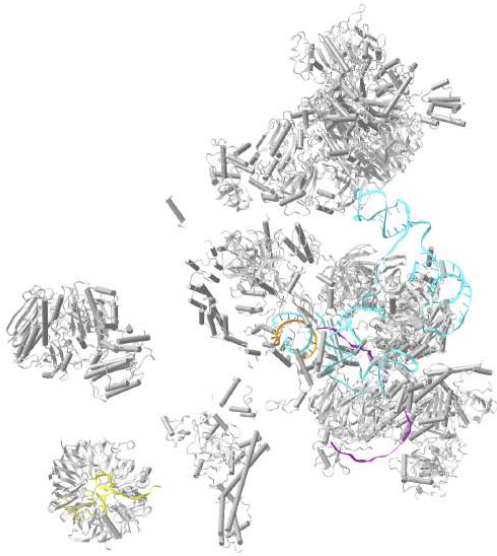


Figure [Math Processing Error]: Human C Complex Spliceosome (7A5P). (Copyright; author via source). Click the image for a popup or use this external link: <https://structure.ncbi.nlm.nih.gov/...5co1nofAVEnyR6>

The color coding is as follows:

- protein - gray
- U6 snRNA - orange
- U5 snRNA - yellow
- U2 snRNA - cyan
- AdML pre-mRNA - magenta

The interactions of the snRNAs are key for structure and catalysis in the spliceosome, and in the figures above, the interactions are hard to discern. Figure [Math Processing Error] shows the interactions among U2, U5, and U6 snRNA in the human and yeast spliceosomes.

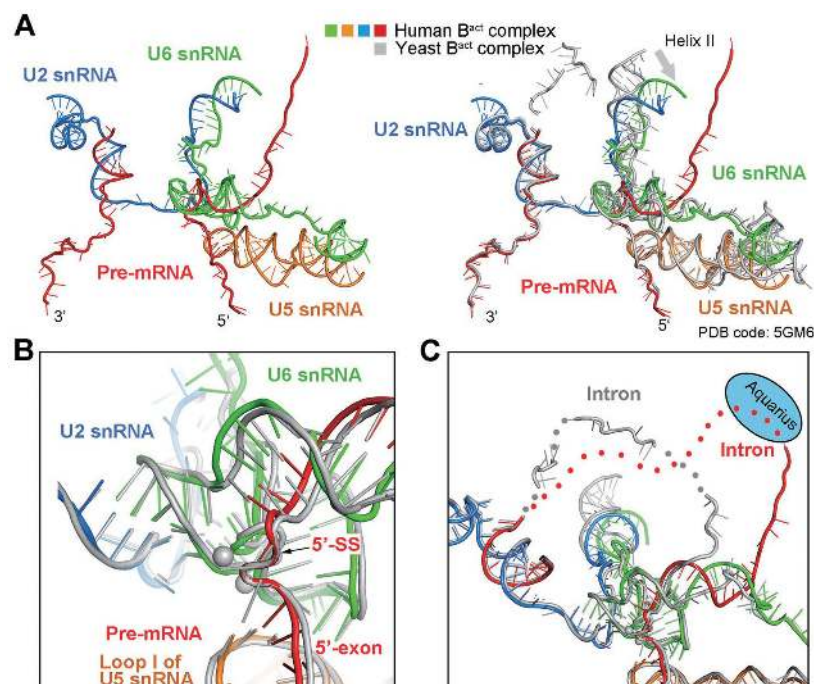


Figure [Math Processing Error]: The RNA elements and the splicing active site of the human mature B^{act} complex. Zhang, X., Yan, C., Zhan, X. *et al.* Structure of the human-activated spliceosome in three conformational states. *Cell Res* **28**, 307–322 (2018). <https://doi.org/10.1038/cr.2018.14>. Creative Commons Attribution 4.0 Unported License. <http://creativecommons.org/licenses/by/4.0/>

Panel (A) shows the structure of the RNA elements in the core of the human mature B^{act} complex. The color-coded RNA elements of the human B^{act} complex are shown in the left panel, and their superposition with those of the *S. cerevisiae* B^{act} complex is displayed in the right panel. All yeast RNA elements are colored grey. Helix II of the U2/U6 duplex in the human B^{act} complex is bent relative to that in the yeast complex.

Panel (B) shows a structural overlay of the active site RNA elements between the human and *S. cerevisiae* B^{act} complexes.

Panel (C) shows that the U6/intron duplex in the human B^{act} complex is considerably longer than that in the *S. cerevisiae* B^{act} complex.

In mammals, the first catalytic step of the splicing reaction begins when the U1 snRNP binds the 5' SS of the intron (defined by the consensus sequence AGGURAGU), and the splicing factors SF1 and U2AF cooperatively recognize the BPS, Py, and 3' SS to assembled complex E or the commitment complex. Subsequently, U2 snRNP and additional proteins are recruited to the pre-mRNA BPS to form the pre-spliceosome or complex A. The binding of the U4/U6-U5 tri-snRNP forms the pre-catalytic spliceosome or complex B. After RNA-RNA and RNA-protein rearrangements at the heart of the spliceosome, U1 and U4 are released to form the activated complex B or complex B*. This complex is responsible for executing the first catalytic step, through which the phosphodiester bond at the 5' SS of the intron is modified by the 2'-hydroxyl of adenosine of the BPS to form a free 5' exon and a branched intron, as shown in Figure [Math Processing Error]. The reaction of the 2'-hydroxyl from the branch point adenosine nucleotide is known as a **transesterification reaction**. During this process, additional rearrangements occur to generate the catalytic spliceosome or complex C (see Figure 10 above), which is responsible for catalyzing the second transesterification reaction leading to intron excision and exon–exon ligation. The resulting intron structure is referred to as a **lariat structure**. After the second catalytic step, the U2, U5, and U6 snRNPs are released from the post-spliceosomal complex and recycled for additional rounds of splicing.

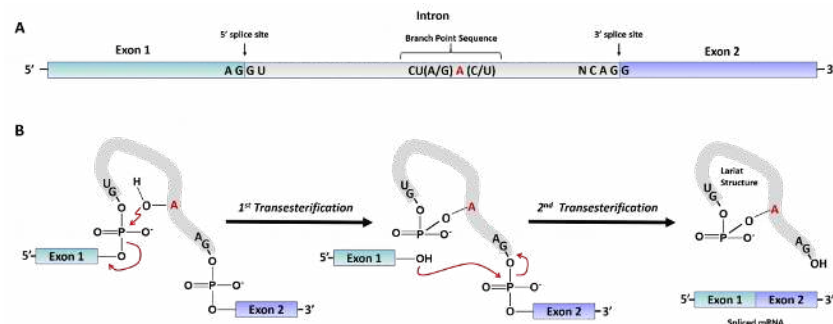


Figure [Math Processing Error]: Transesterification Reactions Involved in mRNA Splicing.

Panel (A) shows a schematic diagram of the pre-mRNA with exons and introns indicated. Key sequences are required for splicing at the 5' and 3' intron locations, and for the recognition and positioning of the branch point Adenosine residue for the first transesterification reaction. (

Panel (B) shows a schematic of the two transesterification reactions required for intron removal. The branch point 2'-OH residue mediates attack on the 5'-phosphate of the intron guanosine residue located at the 5'-splice site. This releases the 3' hydroxyl of Exon 1, which subsequently mediates the attack of the 5' phosphate of the first guanosine residue in Exon 2. The 3' hydroxyl of the intron guanine residue is released, forming the Lariat structure, and Exon 1 is ligated to Exon 2.

Alternative Splicing (AS) offers an additional mechanism for regulating protein production and function. Alternate splicing options are determined by the expression of or exposure to *trans* elements present within unique cellular locations and environments. Additional sequence elements within the mRNA, known as **exonic** and **intronic** splicing silencers or enhancers (ESS, ISS, ESE, and ISE, respectively), participate in the regulation of alternative splicing (AS). Specific RNA-binding proteins, including heterogeneous nuclear ribonucleoproteins (hnRNPs) and serine/arginine-rich (SR) proteins, recognize these sequences to either positively or negatively regulate alternative splicing (AS), as shown in Figure [Math Processing Error]. These regulators, together with an ever-increasing number of additional auxiliary factors, provide the basis for the specificity of this pre-mRNA processing event in different cellular locations within the body.

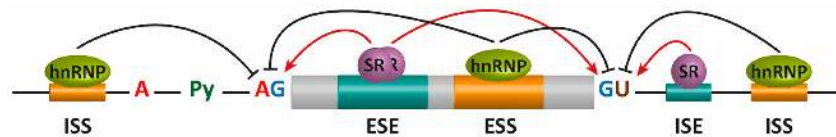


Figure [Math Processing Error]: Alternative splicing (AS) regulation by *cis* mRNA elements and *trans*-acting factors. Suñé-Pou, M., et. al. (2017) *Genes* 8(3):87

The core *cis* sequence elements that define the exon-intron boundaries, including the 5' and 3' splice sites (SS), the GU-AG motif, the polypyrimidine (Py) tract, and the branch point sequence (BPS), are poorly conserved. Additional enhancer and silencer elements in exons and introns (ESE: exonic splicing enhancers; ESI: exonic splicing silencers; ISE: intronic splicing enhancers; ISI: intronic splicing silencers) contribute to the specificity of alternative splicing (AS) regulation. *Trans*-acting splicing factors, such as SR family proteins and heterogeneous nuclear ribonucleoprotein particles (hnRNPs), bind to enhancers and silencers and interact with spliceosomal components. In general, SR proteins bound to enhancers facilitate exon definition, and hnRNPs inhibit this process. These *trans*-acting elements are expressed differentially within different locations or under different environmental stimuli to regulate AS.

There are several different types of AS events, which can be classified into four main subgroups. The first type is exon skipping, which is the major AS event in higher eukaryotes. In this type of event, a cassette exon is removed from the pre-mRNA as shown in Figure [Math Processing Error] (panel A). The second and third types are alternative 3' and 5' SS selections (panels b and c). These types of AS events occur when the spliceosome recognizes two or more splice sites at one end of an exon. The fourth type is intron retention (panel d), in which an intron remains in the mature mRNA transcript. This AS event is much more common in plants, fungi, and protozoa than in vertebrates. Other events that affect the transcript isoform outcome include mutually exclusive exons (panel e), alternative promoter usage (panel f), and alternative polyadenylation (panel g).

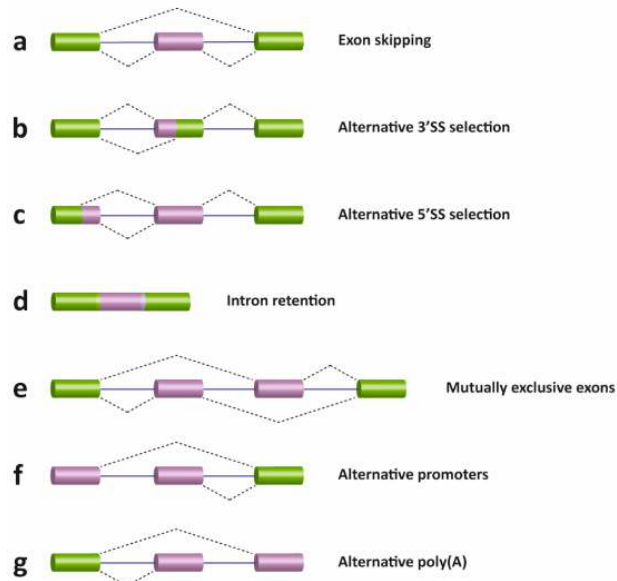


Figure [Math Processing Error]: Schematic representation of different types of alternative transcriptional or splicing events, with exons (boxes) and introns (lines). Constitutive exons are shown in green, and alternatively spliced exons are in purple. Dashed lines indicate the AS event. Exon skipping (a); alternative 3' (b) and 5' SS selection (c); intron retention (d); mutually exclusive exons (e); alternative promoter usage (f); and alternative polyadenylation (g) events are shown. Like alternative splicing (AS), the usage of alternative promoters and polyadenylation sites allows a single gene to encode multiple mRNA transcripts. Suñé-Pou, M., et. al. (2017) *Genes* 8(3):87

Here is a detailed and incredible narrated animation of splicing and the spliceosome.



25.2.6: Summary

This chapter provides a comprehensive exploration of the post-transcriptional modifications that convert a primary mRNA transcript into a mature, functional molecule, with a focus on eukaryotic RNA processing and a brief comparison to prokaryotic mechanisms.

Key Points:

- **Prokaryotic RNA Processing:**

- **Coupled Transcription and Translation:**

- In bacteria, mRNA processing is minimal because transcription and translation occur simultaneously. The early emergence of the Shine-Dalgarno sequence in the 5' untranslated region facilitates immediate ribosome binding.

- **Role of Nus Factors:**

- Proteins such as NusG and NusA modulate transcription by linking RNA polymerase with the ribosome. NusG, for instance, binds to both RNAP and the ribosomal component NusE, thereby coordinating transcription with translation and even participating in transcription termination through interactions with Rho.

- **Eukaryotic RNA Processing:**

- **Cleavage and Polyadenylation (CPA):**

- Almost all eukaryotic mRNAs undergo cleavage at a specific site dictated by polyadenylation signals (commonly the AATAAA hexamer) and surrounding auxiliary elements. This is followed by the addition of a poly(A) tail by a large, multi-subunit cleavage/polyadenylation apparatus (involving complexes like CPSF, CstF, CFI, and CFII). The process not only defines the 3' end of the mRNA but also affects its stability, export, and translation.

- **5' Cap Formation:**

- As soon as the nascent mRNA emerges from RNA polymerase II, a unique 7-methylguanylate cap is added. This process is carried out in three steps:

1. Removal of a terminal phosphate by a triphosphatase.
2. Addition of GMP by a guanylyltransferase.
3. Methylation of the added guanine by a methyltransferase.

The cap plays crucial roles in protecting the mRNA from degradation, promoting nuclear export, facilitating translation initiation, and even aiding in splicing.

- **Splicing and Alternative Splicing:**

- Pre-mRNAs in eukaryotes are segmented into exons (coding regions) and introns (non-coding regions). The spliceosome—a dynamic assembly of five snRNPs (U1, U2, U4, U5, and U6) and numerous proteins—mediates the removal of introns via two successive transesterification reactions. Alternative splicing, regulated by cis-acting elements (like ESEs and ESSs) and trans-acting factors (including SR proteins and hnRNPs), allows a single gene to produce multiple mRNA isoforms, thereby enhancing proteomic diversity.

- **Decapping:**

- To regulate mRNA stability, decapping enzymes (e.g., the Dcp1-Dcp2 complex) remove the 5' cap, marking the mRNA for degradation by exonucleases.

Structural Insights:

- **Molecular Complexes:**

The chapter highlights several structural models:

- **CPA Machinery:**

Detailed structural representations illustrate how the cleavage/polyadenylation complex is organized and how it interacts with nascent RNA.

- **Spliceosome Dynamics:**

Interactive models of various spliceosomal complexes (pre-B, B, and C complexes) reveal the intricate rearrangements that underpin the splicing process.

- **Capping Enzymes:**

Structural depictions of the capping apparatus, including the heterotetrameric complexes involved in 5' cap formation, emphasize the coordinated nature of this modification.

By detailing both the enzymatic and structural components of mRNA processing, this chapter underscores the precision and regulation required to produce stable and translationally competent mRNA molecules in eukaryotes. In contrast to the streamlined process in prokaryotes, the extensive modifications in eukaryotic systems not only ensure proper mRNA maturation but also add layers of regulation that contribute to gene expression diversity and cellular complexity.

This summary serves as a guide to the major themes and molecular mechanisms discussed, providing a solid foundation for understanding how post-transcriptional modifications impact protein synthesis and gene regulation in complex organisms.

25.2.7: References

1. Parker, N., Schneegurt, M., Thi Tu, A.-H., Lister, P., Forster, B.M. (2019) Microbiology. *Openstax*. Available at: <https://opentextbc.ca/microbiologyopenstax/>
2. Palazzo, A., and Lee, E.S. (2015) Non-coding RNA: what is function and what is junk? *Frontiers in Genetics* 6:2 Available at: <file:///C:/Users/flatt/AppData/Local/Temp/fgene-06-00002.pdf>
3. Wikipedia contributors. (2020, July 9). RNA. In *Wikipedia, The Free Encyclopedia*. Retrieved 15:30, August 6, 2020, from <https://en.Wikipedia.org/w/index.php?title=RNA&oldid=966784317>
4. Burenina, O.Y., Oretskaya, T.S., and Kubareva, E.A. (2017) Non-Coding RNAs As Transcriptional Regulators in Eukaryotes. *Acta Naturae* 9(4):13-25. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5762824/>
5. Khatter, H., Vorlander, M.K., and Muller C.W. (2017) RNA polymerase I and III: similar yet unique. *Current Opinion in Structural Biology* 47:88-94. Available at: <https://www.sciencedirect.com/science/article/pii/S0959440X17300313>
6. Wikipedia contributors. (2020, May 8). Sigma factor. In *Wikipedia, The Free Encyclopedia*. Retrieved 17:50, August 7, 2020, from https://en.Wikipedia.org/w/index.php?title=Sigma_factor&oldid=955570499
7. Bae, B., Felkistov, A., Lass-Napiokowska, A., Landick, R., and Darst, S.A. (2015) Structure of a bacterial RNA polymerase holoenzyme open protomer complex. *eLife* 4:e08504. Available at: <https://elifesciences.org/articles/08504>
8. Petrenko, N., Jin, Y., Dong, L., Wong, K.H., and Struhl, K. (2019) Requirements for RNA polymerase II preinitiation complex formation in vivo. *eLife* 8:e43654. Available at: <https://elifesciences.org/articles/43654>
9. Gupta, K., Sari-Ak, D., Haffke, M., Trowitzsch, S., and Berger, I. (2016) Zooming in on transcription preinitiation. *J Mol Biol.* 428(12):2581-2591. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4906157/>
10. Wikipedia contributors. (2020, April 17). TATA-binding protein. In *Wikipedia, The Free Encyclopedia*. Retrieved 14:54, August 8, 2020, from https://en.Wikipedia.org/w/index.php?title=TATA-binding_protein&oldid=951583592
11. Patel, A.B., Greber, B.J., and Nogales, E. (2020) Recent insights into the structure of TFIID, its assembly, and its binding to core promoter. *Curr Op Struct Bio* 61:17-24. Available at: <https://www.sciencedirect.com/science/article/pii/S0959440X19301113#fig0010>
12. Ruff, E.F., Record, Jr., M.T., Artsimovitch, I., (2015) Initial events in bacterial transcription initiation. *Biomolecules* 5(2):1035-1062. Available at: <https://www.mdpi.com/2218-273X/5/2/1035/html>
13. Kireeva, M., Opron, K., Seibold, S., Domesq, C., Cukier, R.I., Coulombe, B., Kashlev, M., and Burton, Z. (2012) Molecular dynamics and mutational analysis of the catalytic and translocation cycle of RNA polymerase. *BMC Biophysics* 5(1):11. Available at: https://www.researchgate.net/publication/225281979_Molecular_dynamics_and_mutational_analysis_of_the_catalytic_and_translocation_cycle_of_RNA_polymerase/figures?lo=1

14. Washburn, R.S., and Gottesman, M.E. (2015) Regulation of transcription elongation and termination. *Biomolecules* 5(2):1063-1078. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4496710/pdf/biomolecules-05-01063.pdf>
15. Zenkin, N., and Yuzenkova, Y. (2015) New insights into the functions of transcription factors that bind the RNA polymerase secondary channel. *Biomolecules* 5(3):1195-1209. Available at: <https://www.mdpi.com/2218-273X/5/3/1195/html>
16. Gocheva, V., LeGall, A., Boudvillain, M., Margeat, E., and Nollmann, M. (2015) Direct observation of the translocation mechanism of transcription termination factor Rho. *Nuc Acids Res* 43(1):10.1093. Available at: https://www.researchgate.net/publication/272162172_Direct_observation_of_the_translocation_mechanism_of_transcription_termination_factor_Rho
17. Miki, T.S., Carl, S.H., and Großhans, H. (2017) Two distinct transcription termination modes dictated by promoters. *Genes & Dev* 31:1-10. Available at: https://www.researchgate.net/publication/320350041_Two_distinct_transcription_termination_modes_dictated_by_promoters
18. Gurumurthy, A., Shen, Y., Gunn, E.M., Bungert, J. (2018) Phase separation and transcription regulation: Are Super-Enhancers and Locus Control Regions primary sites of transcription complex assembly? *BioEssays* 1800164. Available at: https://www.researchgate.net/publication/329331157_Phase_Separation_and_Transcription_Regulation_Are_Super-Enhancers_and_Locus_Control_Regions_Primary_Sites_of_Transcription_Complex_Assembly
19. Suñé-Pou, M., Prieto-Sánchez, Boyero-Corral, S., Moreno-Castro, C., El Yousfi, Y., Suñé-Negre, J.M., Hernández-Munain, C., and Suñé, C. (2017) Targeting splicing in the treatment of human disease. *Genes* 8(3):87. Available at: <https://www.mdpi.com/2073-4425/8/3/87/html>
20. Schaughency, P., Merran, J., and Corden J.L. (2014) Genome-wide mapping of yeast RNA polymerase II termination. *PLOS Genetics* 10(10):e1004632 Available at: <https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1004632>
21. Nourse, J., Spada, S., and Danckwardt, S. (2020) Emerging roles of RNA 3'-end cleavage and polyadenylation in pathogenesis, diagnosis, and therapy of human disorders. *Biomolecules* 10(6):915. Available at: <https://www.mdpi.com/2218-273X/10/6/915/html>
22. Wikipedia contributors. (2020, July 30). Five-prime cap. In *Wikipedia, The Free Encyclopedia*. Retrieved 05:53, August 11, 2020, from https://en.Wikipedia.org/w/index.php?title=Five-prime_cap&oldid=970240533
23. Cortes, T. and Cox, R.A. (2015) Transcription and translation of the rpsJ, rplN and rRNA operons of the tubercle bacillus. *Microbiology* (2015) 161:719-728. Available at: https://www.microbiologyresearch.org/docserver/fulltext/micro/161/4/719_mic000037.pdf?expires=1597159574&id=id&accname=guest&checksum=6FFC9C066EF41C7799FAE843CE94C49F
24. Hein, P.P. and Landick, R. (2010) The bridge helix coordinates movements of modules in RNA polymerase. *BMC Biology* 8:141. Available at: <https://bmcbiol.biomedcentral.com/articles/10.1186/1741-7007-8-141>
25. Gonatopoulos-Pournatzis, T., and Cowling, V.H. (2014) Cap-binding complex (CBC). *Biochem. J.* 457:231-242. Available at: https://www.researchgate.net/publication/259392894_Cap-binding_complex_CBC

This page titled [25.2: RNA Processing](#) is shared under a [not declared](#) license and was authored, remixed, and/or curated by [Henry Jakubowski and Patricia Flatt](#).

- **Current page** by [Henry Jakubowski and Patricia Flatt](#) has no license indicated.
- **11.2: Diffusion Across a Membrane - Passive and Facilitated Diffusion** by [Henry Jakubowski and Patricia Flatt](#) has no license indicated.

25.3: RNA-Dependent Synthesis of RNA and DNA

Learning Goals (ChatGPT o3-mini)

- **Classify RNA Viruses and Their Genomic Architectures:**
 - Identify and differentiate between the various types of RNA virus genomes (dsRNA, ssRNA positive-sense, and ssRNA negative-sense).
 - Explain the taxonomic organization of RNA viruses based on their genome architecture and replication strategies.
- **Understand Viral Replication Strategies:**
 - Describe how different RNA viruses replicate their genomes, including the translation of positive-sense RNA viruses, transcription by negative-sense RNA viruses, and the unique replication of dsRNA viruses.
 - Compare the replication processes of +RNA viruses, -RNA viruses, dsRNA viruses, and retroviruses (including the roles of RNA-dependent RNA polymerase and reverse transcriptase).
- **Examine the Role and Structure of RNA-Dependent RNA Polymerases (RdRps):**
 - Explain the importance of RdRp in the life cycle of RNA viruses and its conservation across different viral families.
 - Identify the conserved motifs (A–G) within the RdRp domain and discuss their roles in nucleotide binding, catalysis, and the overall “cupped right hand” structural arrangement (fingers, palm, and thumb subdomains).
- **Analyze the Mechanism of RNA Synthesis:**
 - Outline the step-by-step process of RNA chain elongation, including NTP binding, active site closure, phosphodiester bond formation, and translocation.
 - Discuss the differences between primer-independent (de novo) and primer-dependent initiation mechanisms in viral RNA synthesis.
- **Evaluate Structural Dynamics and Conformational Changes:**
 - Interpret how conformational changes in the RdRp, especially within the palm and thumb domains, facilitate nucleotide incorporation and translocation.
 - Use structural models (such as those of Foot-and-mouth disease virus or poliovirus polymerase complexes) to explain how specific residues and metal ions contribute to catalysis.
- **Assess the Impact of Viral RdRp Error Rates:**
 - Explain why RNA viruses exhibit high mutation rates due to the lack of proofreading activity in RdRps.
 - Discuss how this error-prone replication mechanism influences viral evolution, adaptability, and the challenges it poses for treatment and vaccine development.
- **Relate Structure to Function in Specific Viral Models:**
 - Describe the replication cycle of a model RNA virus (e.g., poliovirus), emphasizing the role of its RdRp (3Dpol) in generating progeny genomes.
 - Analyze interactive structural models to identify key structural features (such as the pinky finger, primer grip, and NTP channel) and correlate them with their functional roles during replication.

These learning goals provide a framework to deepen understanding of RNA virus biology and the sophisticated enzymatic mechanisms that drive viral replication, setting the stage for further exploration into antiviral strategies and the molecular basis of viral pathogenesis.

25.3.1: RNA Viruses

Infections with RNA viruses place a constant burden on our healthcare systems and economy. Over the past century, this has been particularly true for infections with the Human immunodeficiency virus 1 (HIV-1), Influenza A virus (IAV), Rotavirus (RotaV), West Nile virus (WNV), Dengue virus (DV), Measles virus (MV), and the Porcine reproductive and respiratory syndrome virus (PRRSV). However, emergent RNA viruses can also have considerable consequences, such as the Severe Acute Respiratory

Syndrome-related coronavirus (SARS-CoV) in 2003, the Middle East Respiratory Syndrome coronavirus (MERS-CoV), and most recently, the Severe Acute Respiratory Syndrome-related coronavirus-2 (SARS-CoV-2) in December 2019.

Eukaryotes and bacteria can be infected with a wide variety of RNA viruses. On average, these pathogens exhibit little sequence similarity and employ distinct replication and transcription strategies. RNA virus genomes can consist of double-stranded RNA (dsRNA) or single-stranded (ssRNA) as shown in Figure [Math Processing Error] (Panel a). In turn, the ssRNA viruses can be classified into positive-sense (+RNA) and negative-sense (−RNA) viruses, depending on the translatability of their genetic material. As summarized for the four model RNA pathogens in Panel b, all RNA viruses use dedicated replication and transcription strategies to amplify their genetic material. The common denominator of these strategies is a conserved RNA-dependent polymerase domain. Figure [Math Processing Error] describes variations of the replication process in the different types of RNA viruses.

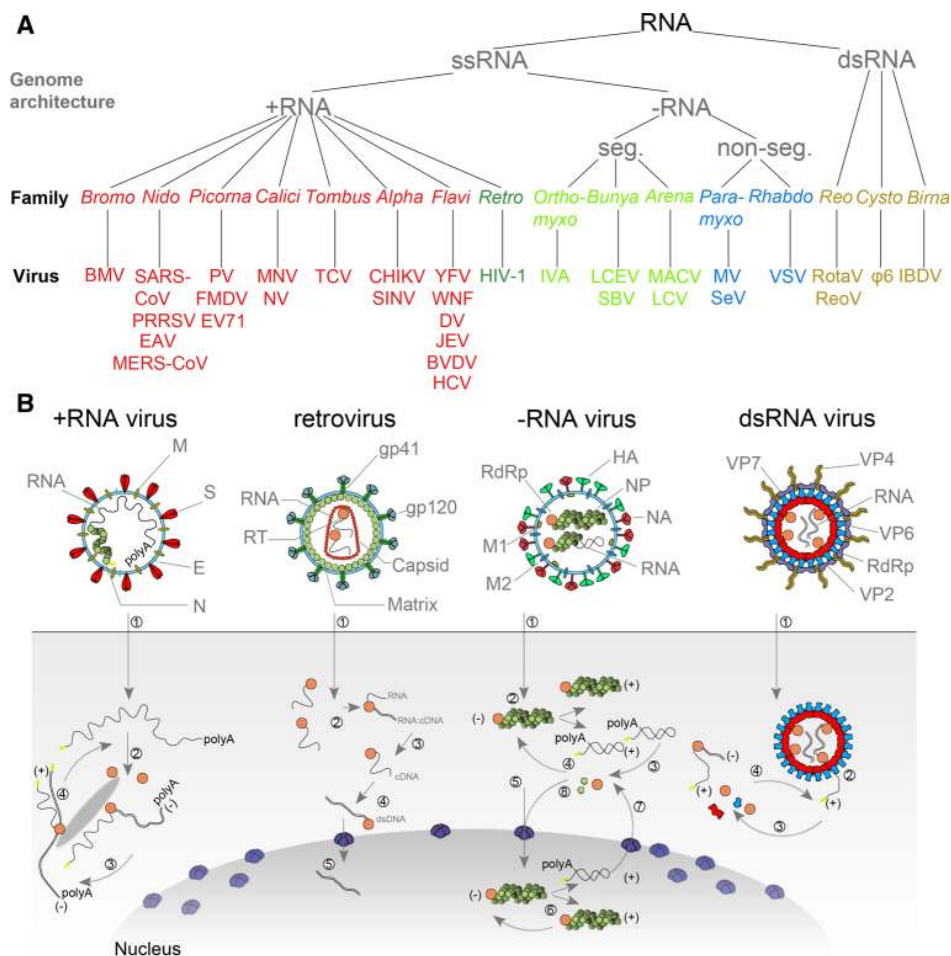


Figure [Math Processing Error]: Taxonomy and Replication Strategies of RNA Viruses. te Velhuis, A.J.W. (2014) *Cell & Mol Life Sci* 71:4403-4420

Panel A shows a simplified taxonomy of the genome architecture of representative RNA viruses.

Panel B (+RNA virus) shows infection with a +RNA virus, as exemplified here with a CoV-like virion, which releases a single-stranded RNA genome into the cytoplasm (1). (2) Translation of the 5'-terminal open-reading frame of the genome produces the viral replicase. (3) This multi-enzyme complex includes **RNA-dependent RNA polymerase (RdRp)** activity (orange) and associates with intracellular membranes before RNA synthesis commences. Newly synthesized −RNAs are subsequently used to produce new +RNAs (4), which are typically capped (yellow) and polyadenylated (polyA). **(Retrovirus)** HIV-1 genomes are packaged as single-stranded RNA (ssRNA) in virions. When the ssRNA is released (1), a cDNA copy is synthesized by the reverse transcriptase enzyme (RT) (2). The RNA is next degraded by the intrinsic RNase H activity in the RT (3) and the single-stranded cDNA is converted to dsDNA (4). The dsDNA is imported into the nucleus (5) for integration into the host's genetic material. **(−RNA virus)** (1) As illustrated here with an IAV-like particle, infection with an −RNA virus releases a viral RNA genome that is

associated with a viral polymerase (*orange*) and nucleoprotein (*green*). (2) In the case of non-segmented RNA viruses, these complexes support transcription to produce viral mRNAs or cRNAs. (3) Viral mRNAs are next translated, and new viral proteins complex with cRNAs to synthesize new vRNAs. (5) The vRNA-containing complexes of some segmented-RNA viruses are imported into the nucleus of the host cell, where (6) the RdRp produces mRNAs or cRNAs. (7) mRNAs are transported to the cytoplasm, while new viral proteins bind cRNAs to form cRNPs for RNA synthesis. (**dsRNA virus**) Fully duplexed RNA genomes lack a cap and polyA elements. (1) The RdRp (*orange*), therefore, transcribes the viral genome inside the capsid of the virion (*blue* and *red*), so viral mRNAs can be (2) released into the cytoplasm as illustrated here with a rotavirus-like virion. In the cytoplasm, the mRNA is translated (3) or replicated by newly synthesized viral RdRps.

Viral RNA-dependent RNA Polymerases (RdRPs) were discovered in the early 1960s through studies on mengovirus and polioviruses, when it was observed that these viruses were not sensitive to actinomycin D. This drug inhibits cellular DNA-directed RNA synthesis. This lack of sensitivity suggested that there is a virus-specific enzyme that can copy RNA from an RNA template, but not from a DNA template. RdRps exhibit some structural similarity to telomerase enzymes, suggesting a potential ancestral relationship between telomerase and RdRps.

The most famous example of RdRp is the poliovirus. The viral genome is composed of RNA, which enters the cell through receptor-mediated endocytosis. From there, the RNA can act as a template for the immediate synthesis of complementary RNA. The complementary strand is then itself able to act as a template for the production of new viral genomes, which are further packaged and released from the cell, ready to infect more host cells. The advantage of this method of replication is that it eliminates the need for a DNA stage; replication is quick and straightforward. The disadvantage is that there is no 'backup' DNA copy (Fig. 26.3.1b; + RNA virus).

Reverse Transcriptase (RT) enzymes, on the other hand, that are common to retroviruses, convert the + RNA strand from the virus into a cDNA copy (Fig. 26.3.1b; retroviruses). The RT enzyme then degrades the RNA, and the single-stranded cDNA is converted into double-stranded DNA (dsDNA). The dsDNA is then integrated into the host's genome, where it can remain in a dormant state. Upon reactivation, + RNA will be manufactured, along with viral proteins used in the assembly of the infectious viral particles.

25.3.2: Structure and Mechanism of RNA Viral Polymerases

The RNA-dependent polymerase domain is a member of the superfamily of template-dependent nucleic acid polymerases, typically consisting of fewer than 400 amino acids. Its sequence is highly variable on average, with some regions showing less than ~10% conservation. Strong amino acid conservation can be observed, however, in regions that are directly involved in nucleotide selection or catalysis, such as the strictly conserved glycine and aspartates in the center of the domain. The prototypic RNA virus RNA polymerase domain harbors seven such regions or motifs, which are arranged in the order G, F1–3, A, B, C, D, and E from amino- to carboxy-terminus. Each of the seven motifs in the RNA polymerase domain adopts a specific and conserved fold, as shown in Figure [\[Math Processing Error\]](#).

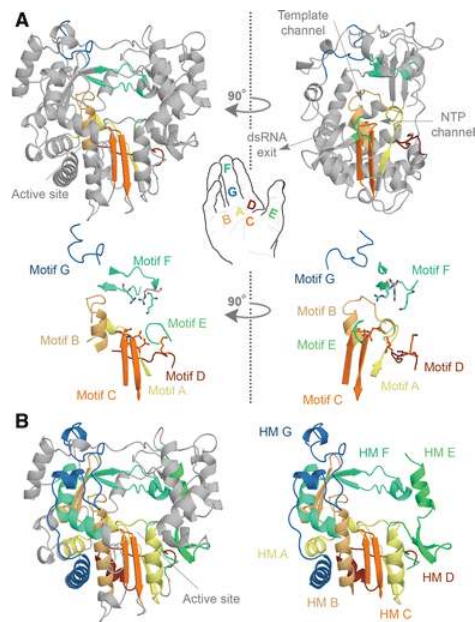


Figure [Math Processing Error]: Conserved structural elements in the RNA virus polymerases. te Velhuis, A.J.W. (2014) *Cell & Mol Life Sci* 71:4403-4420.

Panel A shows the structure of the FMDV RdRp. The motifs A, B, C, D, E, F, and G are color-coded *yellow, gold, orange, red, light green, aquamarine, and blue*. Overall, the polymerase structure adopts a shape that resembles a cupped right hand. Herein, motifs A–E lie on the palm, while motifs F and G are part of the fingers. In the side view of the enzyme, the locations of the template and NTP channels are indicated.

Panel B shows conserved structural elements of the FMDV RdRp. Homomorphs A–G were mapped and color-coded *yellow, gold, orange, red, light green, aquamarine, and blue*, respectively. Images A and B are based on PDB accession 2E9R.

Figure [Math Processing Error] shows an [interactive iCn3D model](#) of the Foot-and-mouth disease virus RNA-polymerase in complex with a template-primer RNA, PPi, and UTP (2E9Z)

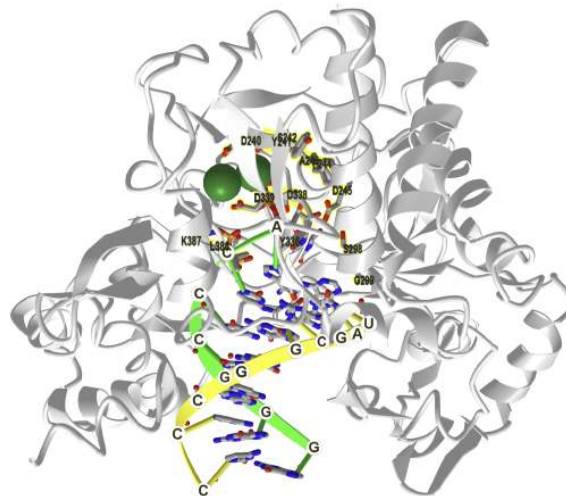


Figure [Math Processing Error]: Foot-and-mouth disease virus RNA-polymerase in complex with a template-primer RNA, PPi, and UTP (2E9Z). (Copyright; author via source). Click the image for a popup or use this external link: <https://structure.ncbi.nlm.nih.gov/i...rZJwYWKHqcAd76>

Active site region side chains are shown as sticks and labeled. The template stand is shown in yellow, and the primer is in green. Note the conserved amino acids Tyr-336, the catalytic Asp-338, and Lys-387 in motifs C and E, respectively. The primer 3'OH forms an H bond with active site Asp 338 (on motif C). In some structures, Asp 338 is bound to a metal ion that interacts with the PPP of the incoming NTP. Two metal ions are involved in catalysis. Tyr 336 interacts with the primer nucleotide while Lys 387 and

Arg 388 interact with the primer backbone. These three residues are highly conserved. The acceptor base of the template strands is adjacent to the NTP binding site. There is no proofreading activity in the replication of viral RNA.

RNA-dependent RNA polymerases generally have a groove on top of the enzyme where RNA enters. It exists in the front. Nucleotides enter at a rear channel. This is illustrated in Figure [Math Processing Error] (Panel b). Following the convention for cellular DNA-dependent DNA polymerases (DdDp), the seven motifs and homomorphs are grouped into three subdomains. These subdomains are called fingers, palm, and thumb based on the polymerase domain's likeness to a cupped right hand, as shown in Figure [Math Processing Error] (Panel A). The finger subdomain loops that interconnect the fingers with the thumb in the RNA-dependent RNA polymerases (RdRps) of +RNA and dsRNA viruses create an overall “closed-hand” conformation that is unique to RNA-dependent RNA polymerases (RdRps) and generally not seen in crystal structures of DNA-dependent DNA polymerases (DdDps) or reverse transcriptases (RTs).

The three subdomains of the RNA-dependent polymerases work together to facilitate the binding of RNA and nucleotides (NTPs). The thumb subdomain contains various residues that are involved in RNA binding, and these generally pack into the minor groove of the template RNA. In some polymerases, the thumb also contains sequences that protrude into the template channel to help stabilize the initiating nucleotide triphosphates (NTPs) on the single-stranded RNA (ssRNA) template. Crucially, these protrusions are also able to undergo relatively large conformational rearrangements to facilitate the translocation of the template following the first condensation reaction. The other residues of the thumb subdomain contribute to the formation of the NTP tunnel, which sits at an $\sim 110^\circ$ angle with the template channel as shown in Figure [Math Processing Error], Panel A. The cavity is lined with positively charged amino acids, although charge interactions are likely insufficient to guide NTPs into the interior. Indeed, molecular dynamics (MD) simulations have demonstrated that the residues of the NTP channel can also explore a relatively large volume of space, which may enable the RdRp to actively “pump” NTPs down the channel.

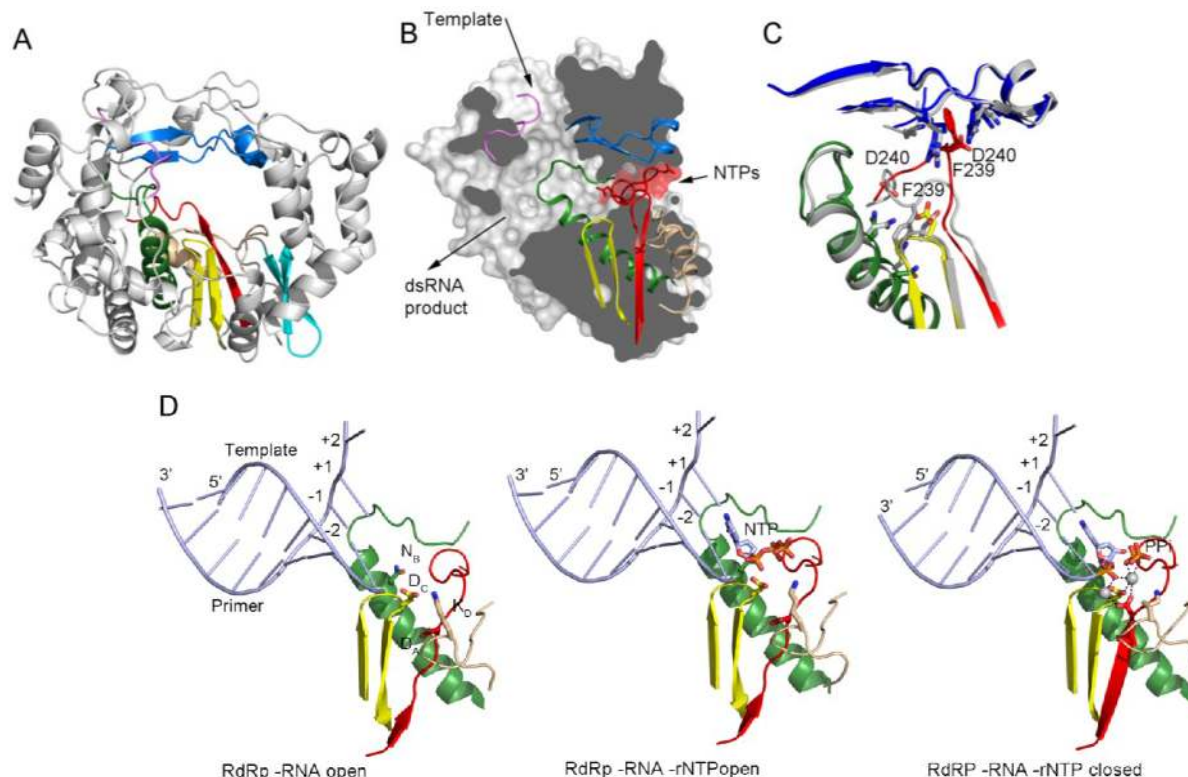


Figure [Math Processing Error]: The Picornavirus (PV) Replication-Elongation Complexes for RNA-Dependent RNA Polymerase. Ferrer-Orta, C., et al. (2015) *Viruses* 7(8):4438-4460

Panel (A) shows a ribbon representation of a typical picornaviral RdRp (model from the cardiovirus EMCV 3D^{pol}, PDB id. 4NZ0). The seven conserved motifs are indicated in different colors: motif A, red; motif B, green; motif C, yellow; motif D, sand; motif E, cyan; motif F, blue; motif G, pink.

Panel (B) shows a lateral view of a surface representation of the enzyme (grey) that has been cut to expose the three channels that are the entry and exit sites of the different substrates and reaction products. The structural elements that support motifs A–G are also shown as ribbons. This panel also shows the organization of the palm sub-domain, with motif A displayed in two alternative

conformations: the standard conformation (PDB ID 4NZ0), found in the apo-form of most crystallized 3Dpol proteins, and the altered conformation found in the tetragonal crystal form of the EMCV enzyme (PDB ID 4NYZ). The alterations affect mainly Asp240, the amino acid in charge of incoming ribonucleotide triphosphate (rNTP) selection, and the neighboring Phe239 that moves ~ 10 Å away from its position in the enzyme catalytic cavity directed towards the entrance of the nucleotide channel, approaching motif F;

Panel (C) shows a close-up of the structural superimposition of the two alternative conformations of the EMCV motif A;

Panel (D) shows the PV replication-elongation complexes. Sequential structures illustrating the movement of the different palm residues from a binary PV 3D^{pol}-RNA open complex (left) to an open 3D^{pol}-RNA-rNTP ternary complex (middle) where the incoming rNTP is positioned in the active site for catalysis and, a closed ternary complex (right) after nucleotide incorporation and pyrophosphate (PPi) release. The residues D_A (involved in rNTP selection through an interaction with the 2' hydroxyl group), D_C (the catalytic aspartate of motif C), K_D (the general acid residue of motif D that can coordinate the export of the PPi group), and N_B (a conserved Asn of motif B, interacting with D_A) have been highlighted as sticks. The different structures correspond to the 3D^{pol}-RNA (PDB id. 3OL6), 3D^{pol}-RNA-CTP open complex (PDB id. 3OLB) and 3D^{pol}-RNA-CTP closed complex (PDB id. 3OL7) structures of PV elongation complexes, respectively

The finger subdomain residues mainly pack into the major groove of the RNA template. Furthermore, upon entry of the template, they can unstack the single strand at position +3, as shown in Figure [Math Processing Error] (Panel A). The non-conserved structural elements of the finger subdomain also play a role in RNA binding. In particular, the fingertips of dsRNA and some +RNA virus RdRps allow the polymerase to “hold” the template without extensive conformational changes. The variations and extensions in the fingers subdomain have also been shown to play roles in protein–protein interactions, phosphorylation, oligomerization, and nuclear import. In contrast, the HIV-1 RT lacks such extensions and adopts a more “open-hand” or U-shaped binding cleft. As a consequence, the RT structure must rearrange its subdomains to coordinate the binding of the template, nascent strand, and incoming dNTPs.

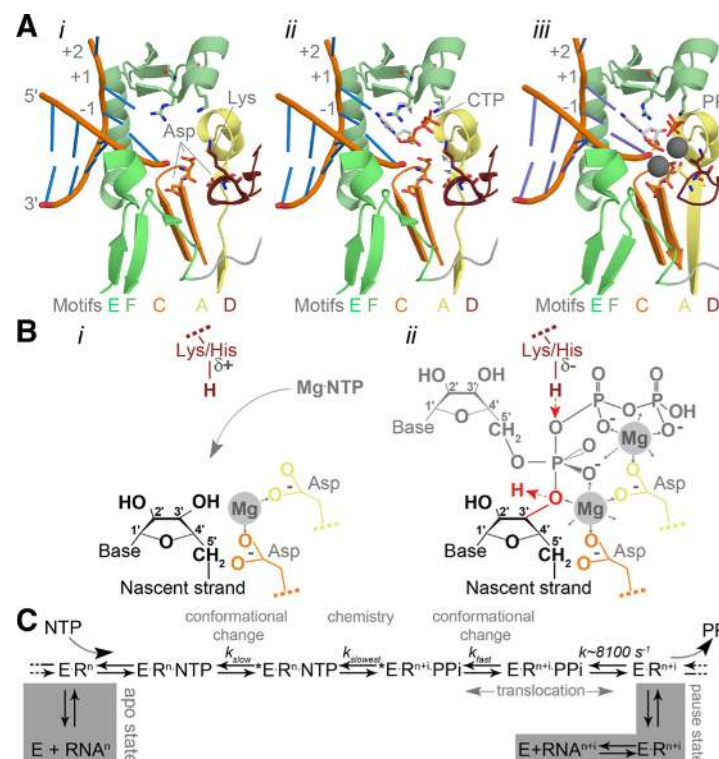


Figure [Math Processing Error]: Catalysis in the RNA virus polymerase active site. te Velthuis, A.J.W. (2014) *Cell & Mol Life Sci* 71:4403-4420.

Panel a shows the structure of the PV active site as it moves from a native state or elongation complex (i) to an open complex (ii), and a closed complex (iii). The closed complex depicted here shows the active site after catalysis has occurred. Highlighted are the

metal-binding aspartates of motifs A and C, and the lysine of motif D that acts as a general acid. Color coding by motif as in Fig. 26.3.2. Image based on PDB accessions 3OI6, 3OLB, and 3OL7.

Panel b shows a schematic presentation of the RdRp active site. The aspartates (Asp) of motif A (yellow) and C (orange) bind divalent metal ions (marked Mg and shaded grey), which are used to coordinate the formation of a new phosphodiester bond at the 3'-OH (red in panel ii) of the nascent strand (yellow). The general acid (red Lys/His in panel ii) is positioned near the β -phosphate of the incoming NTP to protonate the PP_i leaving group.

Panel c shows a simplified schematic of the kinetic steps of RNA polymerases. Asterisk indicates a closed complex.

The NTP and template entry channels meet at the palm subdomain - Figure [Math Processing Error], Panel A. This is a structure that is comprised of a central, partially formed three-stranded β -sheet, which is also present in RNA-recognition motifs (RRMs). The relative movement of these strands within the RRM is crucial for catalysis and depends on NTP binding. Only when a correct NTP binds can motif A and motif C align and the RRM fold be completed. This catalytically competent conformation of the active site is often referred to as the closed complex (not to be confused with the “closed-hand”, which refers to the overall structure of the RdRp) - Figure [Math Processing Error], Panel d.

The polymerase reaction creates new phosphodiester bonds between nucleotide triphosphates (NTPs) using RNA as a template. The NTP substrates involved in this reaction are coordinated by two metal ions, which are bound by the conserved aspartates of motifs A and C - Figure [Math Processing Error] - panel a, i. They also position the NTP's triphosphate optimally for attack by the sugar moiety of the nascent strand once its 3' carbon has lost a proton- Figure [Math Processing Error], panel b. The N-terminal aspartate of motif C thus uses a metal ion to fix the α -phosphate of the incoming NTP and reduce the pK_a of the nascent RNA's 3'-OH to facilitate the attack - Figure [Math Processing Error], panel b, ii. The amino-terminal carboxylate of motif A, on the other hand, stabilizes the β - and γ -phosphates with its divalent metal as well as the pentavalent (phosphorane) intermediate that forms during catalysis - Figure [Math Processing Error], panel b, ii. Structural and biochemical analyses have shown that the formation of this transient structure is dependent on the attack of the NTP's α -phosphate by the 3'-OH, which is often the rate-limiting catalytic step in NTP condensation - Figure [Math Processing Error], panel b.

Motif D's lysine or histidine assists the N-terminal aspartate of motif A in coordinating the β -phosphate of the incoming NTP, analogous to the trigger loop in DdDps. However, when the positively charged side chain of motif D approaches the β -phosphate, it can also protonate the PP_i leaving group - Figure [Math Processing Error], panel b, ii. This second protonation step in the active site is not essential for the polymerase reaction, since catalysis can still take place when motif D's lysine has been mutated to a residue with a different pK_a. The polymerase reaction rate will nevertheless be 50- to 1,000-fold higher when a lysine or histidine is present. Recent data even suggest that motif D can coordinate the export of the PP_i group from the active site once catalysis has occurred, thereby triggering the translocation of the RNA.

Overall, the RNA replication process can be summarized with this four-step mechanism:

1. Nucleoside triphosphate (NTP) binding - initially, the RdRp presents with a vacant active site in which an NTP binds, complementary to the corresponding nucleotide on the template strand. Correct NTP binding causes the RdRp to undergo a conformational change.
2. Active site closure - the conformational change, initiated by the correct NTP binding, restricts active site access and produces a catalytically competent state.
3. Phosphodiester bond formation - two Mg²⁺ ions are present in the catalytically active state and arrange themselves in such a way around the newly synthesized RNA chain that the substrate NTP can undergo a phosphatidyl transfer and form a phosphodiester bond with the newly synthesized chain. With the use of these Mg²⁺ ions, the active site is no longer catalytically stable, and the RdRp complex changes to an open conformation.
4. Translocation - once the active site is open, the RNA template strand can move by one position through the RdRp protein complex and continue chain elongation by binding a new NTP, unless otherwise specified by the template.

Figure [Math Processing Error] shows a more detailed representation of the elongation catalytic cycle.

Elongation catalytic cycle

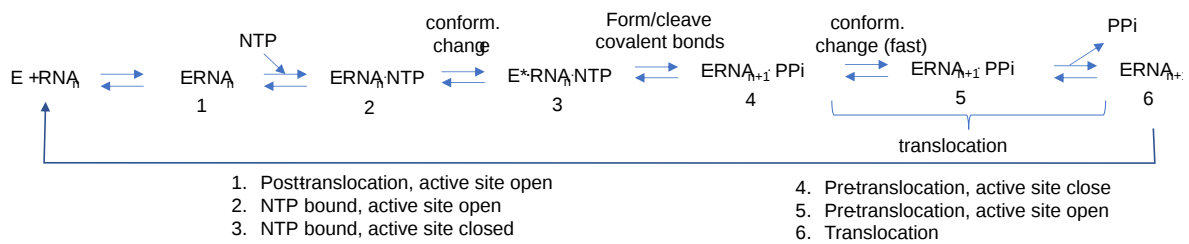


Figure [Math Processing Error]: Elongation catalytic cycle of RNA-dependent RNA polymerases

RNA synthesis can be performed using a primer-independent (*de novo*) or a primer-dependent mechanism that utilizes a viral protein genome-linked (VPg) primer. The *de novo* initiation consists of the addition of a nucleoside triphosphate (NTP) to the 3'-OH of the first initiating NTP. During the following so-called elongation phase, this nucleotidyl transfer reaction is repeated with subsequent NTPs to generate the complementary RNA product. The termination of the nascent RNA chain produced by RdRp is not completely known. However, it has been shown that RdRp termination is sequence-independent.

One feature of RNA-dependent RNA polymerase replication is the immense error rate during transcription. RdRps and RTs are known to have a lack of fidelity on the order of 10^4 nucleotides, which is thought to be a direct result of their insufficient proofreading abilities. This high rate of variation is favored in viral genomes, as it enables the pathogen to overcome defenses developed by hosts trying to avoid infection, thereby facilitating evolutionary growth.

Let's look at another RdRp from the poliovirus. The virus has a single-stranded, positive-sense RNA genome, which makes it infectious in itself but much more infectious when replicated. It is translated into a polyprotein, which is cleaved into about 12 separate proteins. One protein, 3Dpol, is an RNA-dependent RNA polymerase that transcribes the infecting +RNA strand into a -RNA strand, which then serves as a template for the synthesis of additional +RNA strands.

Figure [Math Processing Error] shows an [interactive iCn3D model](https://structure.ncbi.nlm.nih.gov/3D/1f5xp9f7B6Qo4SA) of the Poliovirus polymerase elongation complex with 2',3'-dideoxy-CTP (3OLB). The model shows the palm, finger, and thumb domains.

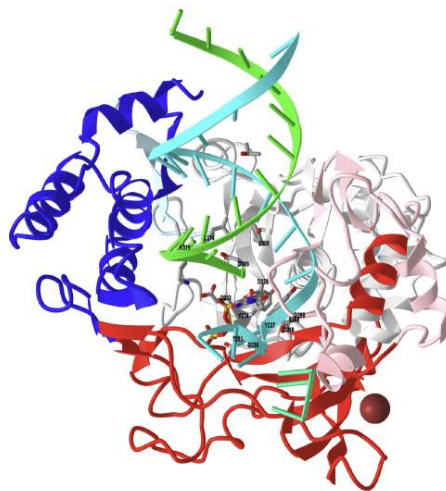


Figure [Math Processing Error]: Poliovirus polymerase elongation complex with 2',3'-dideoxy-CTP (3OLB). (Copyright; author via source). Click the image for a popup or use this external link: <https://structure.ncbi.nlm.nih.gov/3D/1f5xp9f7B6Qo4SA>

Color coding is as follows:

- **red**: finger domain
- **pink**: pinky finger part of the finger domain
- **blue**: thumb domain
- **light blue**: primer grip at the beginning of the thumb domain
- **cyan**: RNA template

- **green:** RNA product
- **gray:** palm domain

Structural studies suggest that the pinky finger is involved in the initiation, while nucleotide binding and catalysis use the palm domain. The thumb domain appears to affect the translocation step. In most nucleases, the finger binds and positions NTPs in the active site. After catalysis, the reverse motion opens the active site, allowing translocation by effectively ratcheting the template by one base pair, driven by the release of PPi. These conformational changes can't be made so easily in +RNA-stranded viral RdRps.

25.3.3: Summary

This chapter delves into the biology and replication mechanisms of RNA viruses, providing an in-depth look at their diversity, genomic architectures, and the molecular machinery that drives their replication. Key points include:

Diversity and Classification of RNA Viruses

- **Genomic Variability:**
RNA viruses display a wide range of genome architectures, including double-stranded RNA (dsRNA) and single-stranded RNA (ssRNA). The ssRNA viruses are further classified as positive-sense (+RNA) or negative-sense (−RNA) based on whether their genomes can be directly translated.
- **Taxonomy and Pathogenesis:**
From established pathogens like HIV-1, Influenza A, and Dengue virus to emergent viruses such as SARS-CoV-2, RNA viruses impose significant health and economic burdens worldwide.

Viral Replication Strategies

- **Replication Modalities:**
Despite their diverse genomes, RNA viruses share a common strategy: the use of an RNA-dependent polymerase (RdRp) to replicate their genomes.
 - **+RNA Viruses:** Their genomes serve directly as mRNA to synthesize viral proteins and the replicase complex.
 - **−RNA Viruses:** They carry a viral polymerase within their virion to transcribe their genomes into mRNA.
 - **dsRNA Viruses:** Their replication occurs within the confines of the viral capsid, ensuring that newly synthesized mRNAs are promptly released into the cytoplasm.
 - **Retroviruses:** These convert their RNA genome into cDNA using reverse transcriptase, integrating it into the host genome.

RNA-Dependent RNA Polymerases (RdRps)

- **Structural Conservation:**
RdRps are central to RNA virus replication and are characterized by a “cupped right hand” architecture with fingers, palm, and thumb subdomains. Seven conserved motifs (A–G) play critical roles in nucleotide binding and catalysis.
- **Catalytic Mechanism:**
The RdRp undergoes a series of conformational changes during RNA synthesis:
 1. **NTP Binding:** The correct nucleoside triphosphate binds to the vacant active site.
 2. **Active Site Closure:** Conformational changes align conserved motifs, forming a catalytically competent complex.
 3. **Phosphodiester Bond Formation:** Two metal ions (typically Mg^{2+}) coordinate the NTP and the 3′-OH of the nascent strand, facilitating bond formation.
 4. **Translocation:** After catalysis, the active site reopens to allow the template to shift, positioning the polymerase for the next nucleotide addition.

Structural Dynamics and Error-Prone Replication

- **Conformational Changes:**
Structural studies, including interactive models of the Foot-and-mouth disease virus and poliovirus polymerases, reveal how distinct subdomains and specific residues (e.g., in the pinky finger, primer grip, and NTP channel) orchestrate nucleotide selection and chain elongation.
- **Mutation Rates:**
The inherent lack of proofreading by RdRps results in a high error rate during replication. While this promotes rapid viral evolution and adaptability, it also presents challenges in developing effective antiviral strategies.

Conclusion

This chapter highlights the intricate interplay between structure and function in RNA virus replication. By exploring the diverse replication strategies and the conserved mechanisms of RdRps, students gain insight into how RNA viruses efficiently hijack host machinery and rapidly evolve. These molecular insights are critical for understanding viral pathogenesis and for the development of targeted antiviral therapies.

This summary provides a concise yet comprehensive overview designed to support advanced biochemistry majors in grasping the complexities of RNA virus replication and the structural basis underlying their enzymatic activities.

25.3.4: References

1. te Velthuis, A.J.W. (2014) Common and unique features of viral RNA-dependent polymerases. *Cell. & Mol Life Sci* 71:4403-4420. Available at: <https://link.springer.com/article/10.1007/s00018-014-1695-z#Sec1>
2. Wikipedia contributors. (2021, August 2). RNA-dependent RNA polymerase. In *Wikipedia, The Free Encyclopedia*. Retrieved 02:47, August 8, 2021, from https://en.Wikipedia.org/w/index.php?title=RNA-dependent_RNA_polymerase&oldid=1036813791
3. Ferrer-Orta, C., Ferrero, D., and Verdaguer, N. (2015) RNA-Dependent RNA Polymerases of Picornaviruses: From the Structure to Regulatory Mechanisms (2015 *Viruses* 7(8):4438-4460. Available at: <https://www.mdpi.com/1999-4915/7/8/2829/html>

This page titled [25.3: RNA-Dependent Synthesis of RNA and DNA](#) is shared under a [not declared](#) license and was authored, remixed, and/or curated by [Henry Jakubowski and Patricia Flatt](#).

- [Current page](#) by [Henry Jakubowski and Patricia Flatt](#) has no license indicated.
- [11.2: Diffusion Across a Membrane - Passive and Facilitated Diffusion](#) by [Henry Jakubowski and Patricia Flatt](#) has no license indicated.

25.4: 26.4 References

<h210.6refs">26.4.6 References

1. Parker, N., Schneegurt, M., Thi Tu, A-H., Lister, P., Forster, B.M. (2019) Microbiology. *Openstax*. Available at: <https://opentextbc.ca/microbiologyopenstax/>
2. Palazzo, A., and Lee, E.S. (2015) Non-coding RNA: what is function and what is junk? *Frontiers in Genetics* 6:2 Available at: <file:///C:/Users/flatt/AppData/Local/Temp/fgene-06-00002.pdf>
3. Wikipedia contributors. (2020, July 9). RNA. In *Wikipedia, The Free Encyclopedia*. Retrieved 15:30, August 6, 2020, from <https://en.Wikipedia.org/w/index.php?title=RNA&oldid=966784317>
4. Burenina, O.Y., Oretskaya, T.S., and Kubareva, E.A. (2017) Non-Coding RNAs As Transcriptional Regulators in Eukaryotes. *Acta Naturae* 9(4):13-25. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5762824/>
5. Khatter, H., Vorlander, M.K., and Muller C.W. (2017) RNA polymerase I and III: similar yet unique. *Current Opinion in Structural Biology* 47:88-94. Available at: <https://www.sciencedirect.com/science/article/pii/S0959440X17300313>
6. Wikipedia contributors. (2020, May 8). Sigma factor. In *Wikipedia, The Free Encyclopedia*. Retrieved 17:50, August 7, 2020, from https://en.Wikipedia.org/w/index.php?title=Sigma_factor&oldid=955570499
7. Bae, B., Felkistov, A., Lass-Napiokowska, A., Landick, R., and Darst, S.A. (2015) Structure of a bacterial RNA polymerase holoenzyme open protomer complex. *eLife* 4:e08504. Available at: <https://elifesciences.org/articles/08504>
8. Petrenko, N., Jin, Y., Dong, L., Wong, K.H., and Struhl, K. (2019) Requirements for RNA polymerase II preinitiation complex formation in vivo. *eLife* 8:e43654. Available at: <https://elifesciences.org/articles/43654>
9. Gupta, K., Sari-Ak, D., Haffke, M., Trowitzsch, S., and Berger, I. (2016) Zooming in on transcription preinitiation. *J Mol Biol.* 428(12):2581-2591. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4906157/>
10. Wikipedia contributors. (2020, April 17). TATA-binding protein. In *Wikipedia, The Free Encyclopedia*. Retrieved 14:54, August 8, 2020, from https://en.Wikipedia.org/w/index.php?title=TATA-binding_protein&oldid=951583592
11. Patel, A.B., Greber, B.J., and Nogales, E. (2020) Recent insights into the structure of TFIID, its assembly, and its binding to core promoter. *Curr Op Struct Bio* 61:17-24. Available at: <https://www.sciencedirect.com/science/article/pii/S0959440X19301113#fig0010>
12. Ruff, E.F., Record, Jr., M.T., Artsimovitch, I., (2015) Initial events in bacterial transcription initiation. *Biomolecules* 5(2):1035-1062. Available at: <https://www.mdpi.com/2218-273X/5/2/1035/htm>
13. Kireeva, M., Opron, K., Seibold, S., Domecq, C., Cukier, R.I., Coulombe, B., Kashlev, M., and Burton, Z. (2102) Molecular dynamics and mutational analysis of the catalytic and translocation cycle of RNA polymerase. *BMC Biophysics* 5(1):11. Available at: https://www.researchgate.net/publication/225281979_Molecular_dynamics_and_mutational_analysis_of_the_catalytic_and_translocation_cycle_of_RNA_polymerase/figures?lo=1
14. Washburn, R.S., and Gottesman, M.E. (2015) Regulation of transcription elongation and termination. *Biomolecules* 5(2):1063-1078. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4496710/pdf/biomolecules-05-01063.pdf>
15. Zenkin, N., and Yuzenkova, Y. (2015) New insights into the functions of transcription factors that bind the RNA polymerase secondary channel. *Biomolecules* 5(3):1195-1209. Available at: <https://www.mdpi.com/2218-273X/5/3/1195/htm>
16. Gocheva, V., LeGall, A., Boudvillain, M., Margeat, E., and Nollmann, M. (2015) Direct observation of the translocation mechanism of transcription termination factor Rho. *Nuc Acids Res* 43(1):10.1093. Available at: https://www.researchgate.net/publication/272162172_Direct_observation_of_the_translocation_mechanism_of_transcription_termination_factor_Rho
17. Miki, T.S., Carl, S.H., and Großhans, H. (2017) Two distinct transcription termination modes dictated by promoters. *Genes & Dev* 31:1-10. Available at: https://www.researchgate.net/publication/320350041_Two_distinct_transcription_termination_modes_dictated_by_promoters
18. Gurumurthy, A., Shen, Y., Gunn, E.M., Bungert, J. (2018) Phase separation and transcription regulation: Are Super-Enhancers and Locus Control Regions primary sites of transcription complex assembly? *BioEssays* 1800164. Available at: https://www.researchgate.net/publication/329331157_Phase_Separation_and_Transcription_Regulation_Are_Super-Enhancers_and_Locus_Control_Regions_Primary_Sites_of_Transcription_Complex_Assembly
19. Suñé-Pou, M., Prieto-Sánchez, Boyero-Corral, S., Moreno-Castro, C., El Yousfi, Y., Suñé-Negre, J.M., Hernández-Munain, C., and Suñé, C. (2017) Targeting splicing in the treatment of human disease. *Genes* 8(3):87. Available at: <https://www.mdpi.com/2073-4425/8/3/87/htm>

20. Schaughency, P., Merran, J., and Corden J.L. (2014) Genome-wide mapping of yeast RNA polymerase II termination. *PLOS Genetics* 10(10):e1004632 Available at: <https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1004632>
21. Nourse, J., Spada, S., and Danckwardt, S. (2020) Emerging roles of RNA 3'-end cleavage and polyadenylation in pathogenesis, diagnosis, and therapy of human disorders. *Biomolecules* 10(6):915. Available at: <https://www.mdpi.com/2218-273X/10/6/915/htm>
22. Wikipedia contributors. (2020, July 30). Five-prime cap. In *Wikipedia, The Free Encyclopedia*. Retrieved 05:53, August 11, 2020, from https://en.Wikipedia.org/w/index.php?title=Five-prime_cap&oldid=970240533
23. Cortes, T. and Cox, R.A. (2015) Transcription and translation of the rpsJ, rplN and rRNA operons of the tubercle bacillus. *Microbiology* (2015) 161:719-728. Available at: https://www.microbiologyresearch.org/docserver/fulltext/micro/161/4/719_mic000037.pdf?expires=1597159574&id=id&accname=guest&checksum=6FFC9C066EF41C7799FAE843CE94C49F
24. Hein, P.P. and Landick, R. (2010) The bridge helix coordinates movements of modules in RNA polymerase. *BMC Biology* 8:141. Available at: <https://bmcbiol.biomedcentral.com/articles/10.1186/1741-7007-8-141>
25. Gonatopoulos-Pournatzis, T., and Cowling, V.H. (2014) Cap-binding complex (CBC). *Biochem. J.* 457:231-242. Available at: https://www.researchgate.net/publication/259392894_Cap-binding_complex_CBC

This page titled [25.4: 26.4 References](#) is shared under a [not declared](#) license and was authored, remixed, and/or curated by [Henry Jakubowski and Patricia Flatt](#).

CHAPTER OVERVIEW

26: Protein Metabolism

[Return to Fundamentals of Biochemistry](#)

[Search Fundamentals of Biochemistry](#)

[26.1: The Genetic Code](#)

[26.2: Protein Synthesis](#)

[26.3: Translational Regulation and Protein Degradation](#)

This page titled [26: Protein Metabolism](#) is shared under a [not declared](#) license and was authored, remixed, and/or curated by [Henry Jakubowski](#) and [Patricia Flatt](#).

26.1: The Genetic Code

Learning Goals (ChatGPT o3-mini)

- **Compare Prokaryotic and Eukaryotic Translation Systems:**
 - Describe how the spatial and temporal separation of transcription and translation in eukaryotes differs from the coupled system in prokaryotes.
 - Explain the implications of cellular compartmentalization on the regulation and efficiency of protein synthesis.
- **Understand the Chemical Basis of Protein Synthesis:**
 - Detail the dehydration (condensation) reaction that forms peptide bonds between amino acids.
 - Illustrate the directionality of protein synthesis (N- to C-terminal) and the role of the ribosome as a ribozyme in catalyzing peptide bond formation.
- **Interpret mRNA Reading and the Genetic Code:**
 - Define codons and the concept of a reading frame; explain how the ribosome reads mRNA in non-overlapping, triplet codons from the 5' to 3' direction.
 - Discuss codon redundancy (degeneracy) and the wobble hypothesis, including how these features minimize the impact of mutations on protein function.
 - Analyze the organization of the genetic code and its universality across species, noting variations such as those in mitochondrial genomes.
- **Examine tRNA Structure and Function:**
 - Outline the canonical cloverleaf secondary structure and L-shaped tertiary structure of tRNAs, highlighting the role of each domain (acceptor stem, D-arm, anticodon arm, variable loop, TΨC-arm).
 - Describe the maturation process of tRNAs, including leader and trailer removal, intron splicing, and the post-transcriptional addition of the CCA sequence.
 - Discuss the significance of post-transcriptional modifications in tRNAs, particularly in the anticodon loop and core regions, and their roles in maintaining translation fidelity.
- **Explore Aminoacylation and Aminoacyl-tRNA Synthetases:**
 - Explain the two-step aminoacylation reaction, including amino acid activation and the transfer of the amino acid to the tRNA.
 - Compare and contrast the structural and mechanistic differences between Class I and Class II aminoacyl-tRNA synthetases.
 - Describe the editing mechanisms (pre-transfer and post-transfer) that ensure high fidelity in tRNA charging.
- **Understand Ribosome Structure and Function:**
 - Identify the key components of the ribosome (rRNA and ribosomal proteins) and describe the organization of the A, P, and E sites.
 - Explain how the ribosome's structure supports its function in mRNA decoding and peptide bond formation.
 - Compare the structures of prokaryotic (70S) and eukaryotic (80S) ribosomes, including differences in subunit composition and rRNA processing.
 - Describe the process of ribosome assembly and the formation of polysomes, noting how these structures contribute to efficient protein synthesis.

Each of these goals is intended to build a comprehensive understanding of translation from the molecular to the cellular level, highlighting the intricate interplay of RNA, protein, and enzymatic activity that underpins the synthesis of proteins in all living organisms.

26.1.1: Overview of Translation

Within this chapter, we will cover the details of prokaryotic and eukaryotic translation. *Translation* is the process of converting the information housed in mRNA into the protein sequence. Essentially, you are translating the language of nucleotides into the

language of amino acids. Recall that prokaryotic and eukaryotic transcription and translation systems differ in large part due to the compartmentalization of larger eukaryotic cells. Due to this compartmentalization, transcription and translation are separated spatially and temporally within the cell. Transcription occurs within the nucleus of eukaryotes, and translation occurs within the cytoplasm. Prokaryotes lack compartmentalization and, as a result, have evolved a coupled transcription/translation system in which both processes occur simultaneously. Both are illustrated in Figure [Math Processing Error].

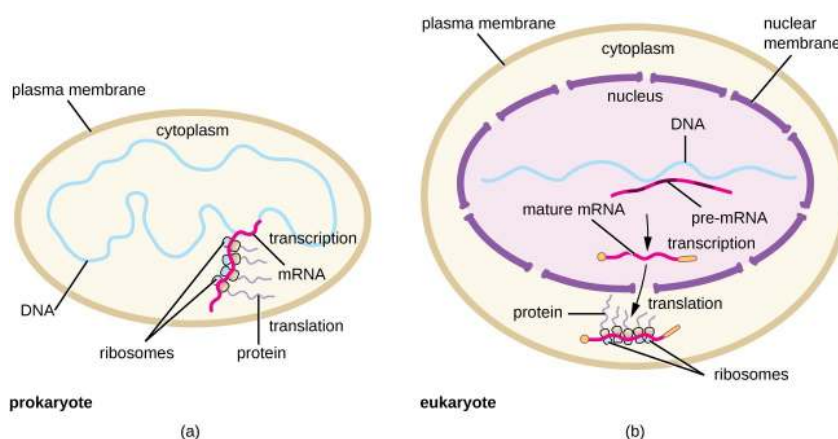


Figure [Math Processing Error]: Cellular Location of Transcription and Translation in Prokaryotes and Eukaryotes. Baccei, A., and Rice, M. Lumen Learning

Panel (a) shows that prokaryotes lack cellular compartmentalization and show coupled transcription-translation processing;

Panel (b) shows that eukaryotes have a high degree of compartmentalization and separate the processes of transcription, which are in the nucleus of the cell, from the processes of translation, which are localized in the cytoplasm.

Recall that peptide formation is a dehydration reaction that combines the carboxylic acid of the upstream amino acid with the amine functional group of the downstream amino acid to form an amide linkage, as shown in Figure [Math Processing Error]. Water is the by-product. The ribosome, a large complex of peptides and rRNA molecules, serves as the enzyme that mediates this reaction. It requires a mature mRNA to serve as the template and directionally performs peptide bond synthesis from the N to the C-terminal of the growing peptide/protein. This is known as N-to-C-synthesis. Note that the overall protonation state shown is unlikely, as under conditions where the carboxyl groups are protonated, so would the amines. This representation makes it easier to highlight the departing water

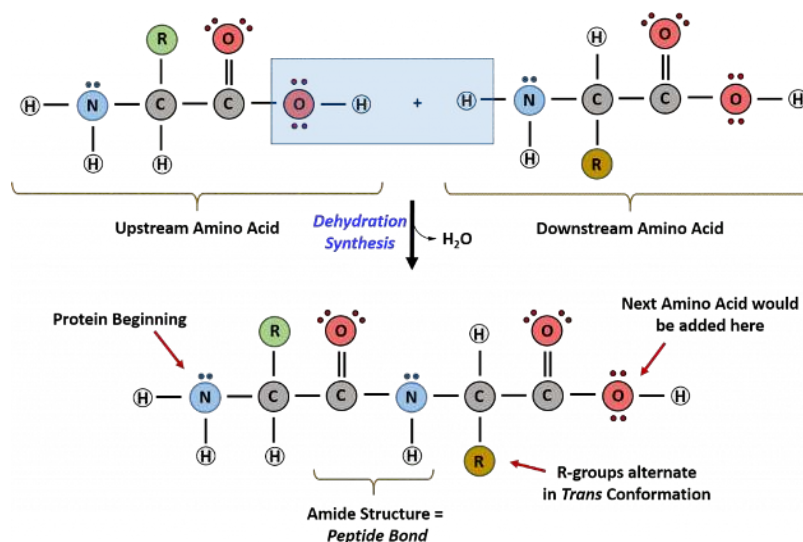


Figure [Math Processing Error]: Formation of the Peptide Bond. The addition of two amino acids to form a peptide requires dehydration synthesis. The carboxylic acid of the upstream amino acid is joined with the amine functional group of the downstream amino acid to form the amide linkage. Within the ribosome, this reaction is highly directional and only occurs in the N-to-C orientation.

To maintain proper protein function, the error rate of translation is approximately 10^{-4} or 1 error in every 10,000 amino acids encoded. The fidelity of protein synthesis is maintained by the ribosome's ability to match the code from the template mRNA

strand with the appropriate amino acid. The ribosome reads template mRNA in groups of three nucleotides, called a *codon*, as shown in Figure [Math Processing Error].

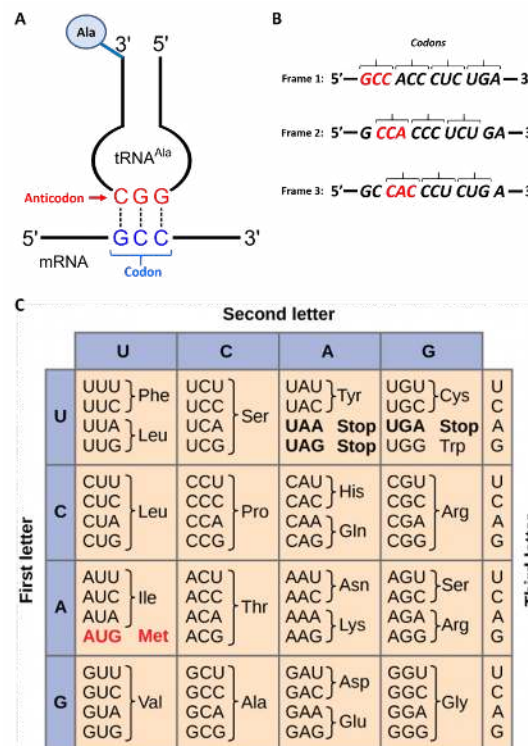


Figure [Math Processing Error]: Reading the mRNA Template. modified from: (A) Yikrazuul and (C) Zedalis, J., et al. Openstax

The template is non-overlapping and reads in discrete groups of three. This is known as the *reading frame* of the mRNA, and it is always read from the 5' to 3' direction. Thus, for each mRNA, there are three potential reading frames (panel B). Only one reading frame will be the correct one for protein synthesis. The ribosome must recognize and align the correct reading frame of the mRNA such that the correct codon sequences can be read. Small, distinct tRNA molecules are tethered with specific amino acids and contain specific anticodons that complement mRNA codon sequences. The tRNA molecules can cycle on and off the ribosome structure to hybridize with the correct codon sequences and chaperone the proper amino acid for peptide bond formation. The ribosome then serves as a ribozyme, mediating the peptidyl transferase activity to form the peptide bond. The mRNA is then shifted to reveal the next codon within the sequence, and the process is repeated until the entire protein has formed. Panel C displays the codon chart for all possible combinations of three nucleotides. 64 possible codon combinations are possible using the four nucleotide possibilities, but only 20 amino acids are encoded during protein synthesis. Each codon is specific for a single amino acid. There is *very little ambiguity* within the code.

However, there is *redundancy* within the code; i.e., many amino acids have more than one codon that encodes for that specific amino acid. To account for this redundancy, many tRNA molecules can recognize more than one codon using a single anticodon. This is known as *degeneracy*. *Degeneracy* usually occurs at the third position of the codon and is known as the *wobble base* position. *Degeneracy* helps minimize the effects of mutations within the coding sequence, as mutations in the *wobble base* position often lead to a *silent mutation* – i.e., the mutation will still encode for the same amino acid.

In addition, if comparing the polarity of amino acids encoded by the different codons, neighboring codons typically encode for amino acids with similar polarity, as shown in Figure [Math Processing Error]. This also helps minimize the effects of mutations by converting one amino acid within the sequence to one with a similar polarity. This type of mutation is more likely to cause less disturbance to the 3-dimensional structure of the resulting protein and retain biological function.

The Genetic Code is universal for almost all species alive on the planet, providing support for a single origin of life. Most deviations in the code occur within the mitochondria of eukaryotic species, as shown in Figure [\[Math Processing Error\]](#).

	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	
	UUC		UCC		UAC	Tyr	UGC		
	UUA	Leu	UCA		UAA	Term	UGA	Term	→ All animals Trp
	UUG		UCG	UAG	Term	UGG	Trp		
Most metazoans except Echinoderms, Planaria, Cnidarians, Poriferans	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg	
	CUC		CAC			CGC			
	CUA		CAA		Gln	CGA			
	CUG		CAG			CGG			
	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	
AUC	ACC	AAC			AGC				
AUA	Ile	AAA	Lys		AGA	→ Most invertebrates Ser Tunicates (Urochordates) Gly			
AUG	Met	ACG	AAG		Lys		AGG		
	GUU		GCU	Ala	GAU		GGU	Vertebrates Term	
	GUC	GCC	GAC		Asp	GGC			
	GUA	GCA	GAA		Glu	GGA			
	GUG	GCG	GAG			GGG			
					Asn				

Radopholus similis (a nematode)
Planaria

↑ Tyr

↓ Asn

Echinoderms Platyhelminths

<https://bio.libretexts.org/@go/page/15201>

The life of a transfer RNA (tRNA) molecule starts with a series of important maturation steps that can vary in their sequential order from case to case. Leader and trailer sequences are removed by a set of endo- and exonucleases, and in several tRNA precursors, splicing reactions excise intronic sequences. Furthermore, in many organisms, the sequence CCA, which represents the site of amino acid attachment, is not encoded but has to be added post-transcriptionally by CCA-adding enzymes. While all primary tRNA transcripts are composed of the four standard RNA bases A, C, G, and U, many of these nucleotides are modified, altering their properties in very different ways. Currently, 93 post-transcriptional modifications are known, and the variety of their functions is at least similarly diverse and not fully understood. The complexity of such modifications ranges from simple methylations at the bases or the ribose to more complex and extensive base hypermodifications, whose synthesis often requires a cascade of enzymatic reactions. Modifications can alter a tRNA's shape in subtle ways, but can also lead to massive structural rearrangements. Additionally, they ensure efficient translation by maintaining the anticodon loop structure and promoting accurate codon-anticodon interactions.

After maturation, tRNAs have multiple interaction partners in their life cycle, ranging from aminoacyl-tRNA-synthetases that are responsible for amino acid attachment to translation factors, ribosomes, and mRNAs. Apart from synthetases, these interaction partners do not specifically act on a single individual tRNA transcript or isoacceptor, but rather on all tRNAs, similar to the above-mentioned CCA-adding enzyme. Thus, despite a high sequence variation, a cell's tRNAs show a well-conserved cloverleaf-like secondary structure that was originally discovered in 1965. The similar structure of all tRNA molecules allows them to bind to common protein synthesis machinery, such as the ribosome and CCA-adding enzymes. The cloverleaf consists of five parts: the acceptor stem (containing the tRNA's 5'- and 3'-ends), the D-arm, the anticodon arm, the variable loop, and the TΨC-arm (T-arm). At the 3'-terminus, the tRNA carries the CCA sequence, required for aminoacylation, tRNA positioning in the ribosome, and translation termination. In a conserved network of tertiary interactions, mostly between D- and T-loop, tRNAs fold into an L-shaped three-dimensional structure, which was first solved by Kim et al. in 1974, as shown in Figure [\[Math Processing Error\]](#) (Panel B). The longest possible distance separates the anticodon and the amino acid-accepting CCA-ends from each other. The conserved structure of a tRNA is essential for its recognition by the ribosome, other RNAs, and proteins, and, consequently, for its proper functionality. For example, the CCA-adding enzyme utilizes the acceptor domain for substrate recognition, whereas aminoacyl-tRNA synthetases employ several recognition elements, including the anticodon, acceptor stem, and discriminator position.

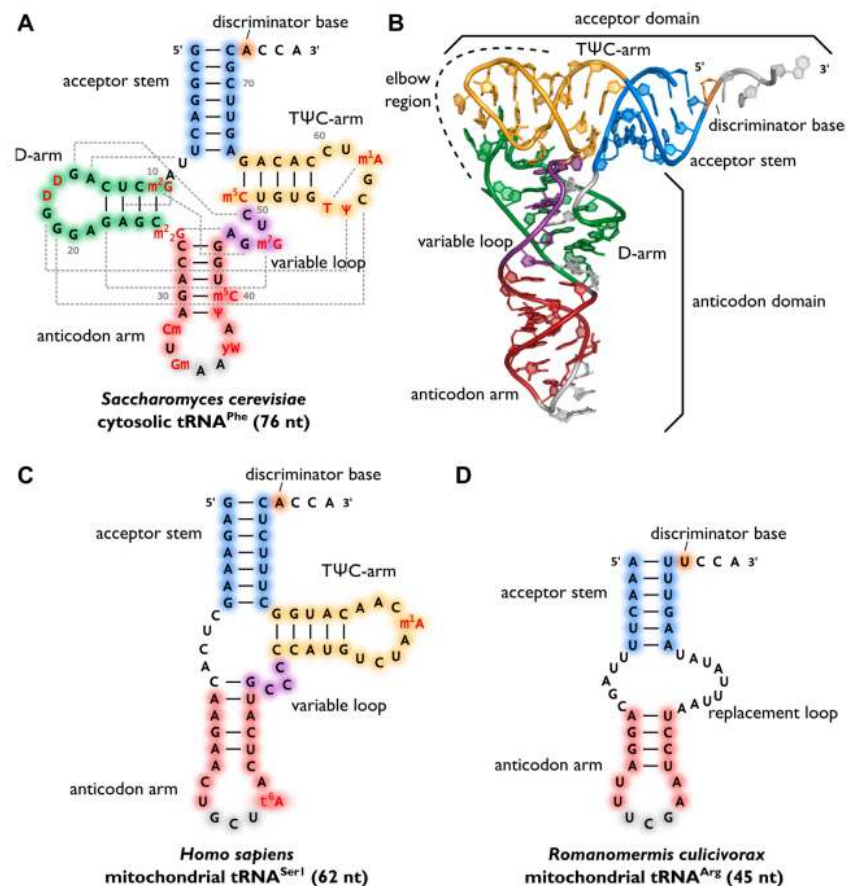


Figure [Math Processing Error]: Variability of transfer RNA (tRNA) structures. Lorenz, C., et. al. (2017) Biomolecules 7(2):35

Panel (A) shows the canonical cloverleaf secondary structure of cytosolic tRNA^{Phe} from *S. cerevisiae*, with acceptor stem (blue), D-arm (green), anticodon arm (red), variable loop (purple), and TΨC-arm (yellow). The anticodon is labeled in grey, the discriminator base in orange, and post-transcriptional modifications in red. Grey dashed lines indicate tertiary interactions based on structural data, and the length of the RNA is indicated in parentheses.

Panel (B) shows the L-shaped tertiary structure of the cytosolic tRNA^{Phe} from *S. cerevisiae*—Protein Data Bank entry (PDB): 1EHZ. The acceptor domain is composed of a stacked T-arm and acceptor stem, whereas D- and anticodon arm form the anticodon domain. The region where both domains come together and interact with each other via tertiary base pairing is also called the elbow region.

Panel(C) shows the secondary structure of human mitochondrial tRNA^{Ser1}, which lacks the whole D-arm;

Panel (D) shows the secondary structure of the mitochondrial tRNA^{Arg} from the nematode *Romanomermis culicivorax*, which lacks both D- and T-arm. Instead, we find a so-called replacement loop. It represents the shortest tRNA found *in vivo*.

Surprisingly, not all tRNAs fold into the canonical cloverleaf structure. Many mitochondrial tRNAs are reduced in length and sometimes completely lack the D- or T-arm as shown in Figure [Math Processing Error], Panel C. In the mitochondria of nematodes, this situation is carried to an extreme, as tRNAs lacking one or even both arms seem to be the rule (Panel D).

Figure [Math Processing Error] shows an [interactive iCn3D model](#) of the yeast phenylalanine tRNA (1EHZ).

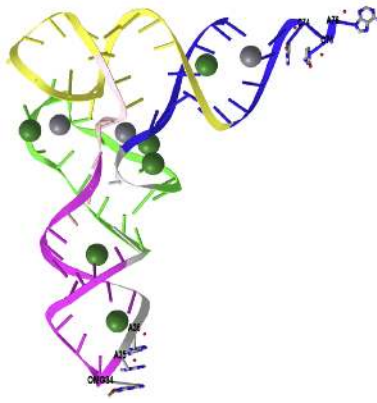


Figure [Math Processing Error]: Yeast phenylalanine tRNA (1EHZ). (Copyright; author via source). Click the image for a popup or use this external link: <https://structure.ncbi.nlm.nih.gov/1...yq7wdue7MNuT56>

The coloring, which matches that in Figure 6 above, is shown below:

- Acceptor Stem - blue
- D Arm - green
- Anticodon= magenta
- Variable Loop - pink
- TΨC - yellow

Post-transcriptional enzyme-catalyzed modification of tRNA occurs at many base and sugar positions and influences specific anticodon–codon interactions, regulating translation, its efficiency, and fidelity. This phenomenon of nucleoside modification is particularly remarkable, resulting in a rich structural diversity of tRNA, of which over 93 modifications have been characterized.

The variety of post-transcriptional modifications can be classified into two groups according to their complexity. The first group comprises the majority of modified bases, which have simple methylations at the ribose or base moiety, usually introduced by a single enzymatic reaction. Simple modifications can be found at almost every position of the tRNA molecule with a high density in the tRNA core region, where tertiary interactions between D- and T-arm stabilize the three-dimensional fold, as shown in Figure [Math Processing Error]. The second group includes complex modifications, whose synthesis requires the sequential activity of several enzymes. Most often, these hypermodified nucleosides are found in the anticodon of tRNAs, where they play a direct role in codon recognition and create what is known as the *wobble base* or *wobble position*.

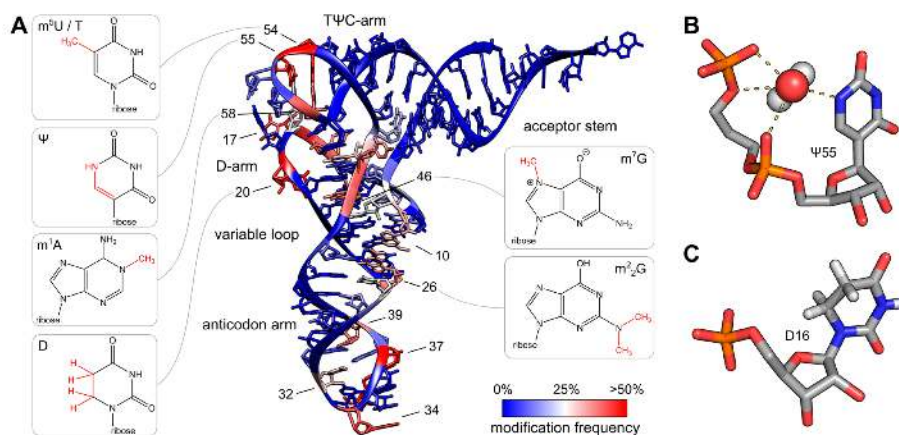


Figure [Math Processing Error]: Post-transcriptional modifications in tRNA. Lorenz, C., et. al. (2017) Biomolecules 7(2):35

Panel (A) shows the colored tRNA structure, which indicates the modification frequency of each base. The modification data were taken from the tRNAmoviz database and plotted on the crystal structure of tRNA^{Phe} from *S. cerevisiae*. Blue-colored bases are rarely modified, while red-colored bases are hotspots for modification. tRNAs possess two regions with high modification levels—the anticodon loop (especially positions 34 and 37) and the core or elbow region, where D- and T-loop bases interact with each other and stabilize the tertiary fold. For some important positions, the chemical structure of the most frequent modification at this position is shown.

Panel (B) shows the three-dimensional structure of pseudouridine at position 55 of tRNA^{Phe} from *S. cerevisiae*. The additional H-bond donor at N1 interacts with the 5'-adjacent phosphates via a coordinated water molecule. The hydrogen bond to N1 was not resolved in the crystal structure. The ribose shows a stabilizing C3'-endo conformation. PDB: 1EHZ-

Panel (C) shows the three-dimensional structure of D16 in the D-arm of tRNA_i^{Met} from *Schizosaccharomyces pombe*. The C5-C6 bond of dihydrouridine is reduced, which leads to a non-planar structure of the base. The ribose takes the less stable C2'-endo conformation. PDB: 2MN0.

A *wobble base pair* is a pairing between two nucleotides in RNA molecules that does not follow Watson-Crick base pair rules. The four main *wobble base pairs* are guanine-uracil (G-U), hypoxanthine-uracil (I-U), hypoxanthine-adenine (I-A), and hypoxanthine-cytosine (I-C), as shown in Figure [Math Processing Error]. To maintain the consistency of nucleic acid nomenclature, “I” is used for hypoxanthine because hypoxanthine is the nucleobase of the inosine nucleotide; nomenclature otherwise follows the names of nucleobases and their corresponding nucleosides (e.g., “G” for both guanine and guanosine, as well as for deoxyguanosine). The thermodynamic stability of a wobble base pair is comparable to that of a Watson-Crick base pair. Wobble base pairs are fundamental to RNA secondary structure and are crucial for the accurate translation of the genetic code.

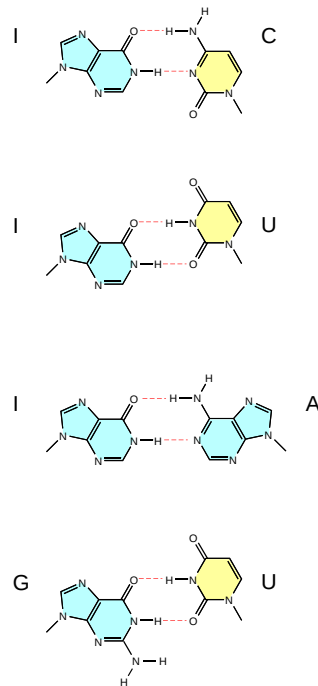


Figure [Math Processing Error]: The Four Main Wobble Base Pairs. The top three diagrams illustrate the flexibility of the inosine nucleotide in forming hydrogen bond interactions with cytosine, uracil, and adenine. tRNA and mRNA wobble base pairing also includes guanine hydrogen binding with uracil. Figure from [Fdardel](#)

The *wobble base* position is usually the first position of the anticodon (read in the 5' – 3' direction), which aligns with the 3rd position of the mRNA codon. This helps to explain the *degeneracy* found within the genetic code as shown in Figure 3 above and Figure [Math Processing Error]. *Degeneracy* refers to the ability of a single tRNA to recognize multiple different codons within mRNA.

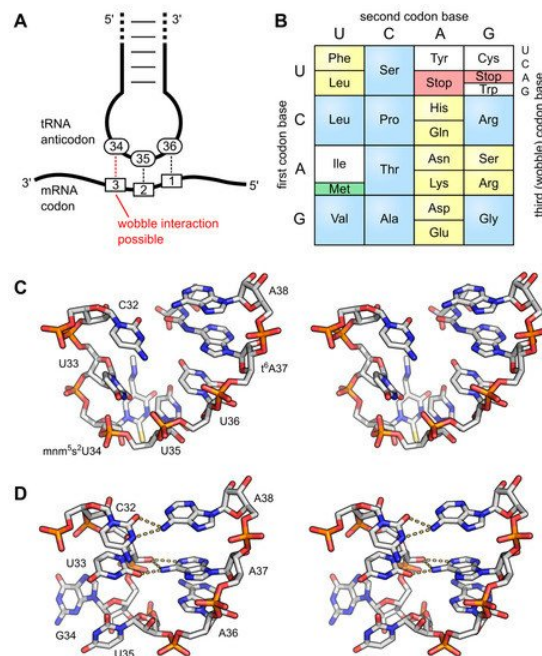


Figure [Math Processing Error]: Anticodon Loop Structure and Codon Degeneracy. [Lorenz, C., et. al. \(2017\) Biomolecules 7\(2\):35](#)

Panel (A) shows the interaction of the anticodon bases (34–36) of a tRNA with the corresponding bases of the mRNA codons (3, 2, 1). A wobble interaction is possible between codon base 3 and anticodon base 34. The latter is frequently modified and directs the

wobble interactions with the third codon base.

Panel (B) illustrates the standard genetic code as a simple decoding table, where 2-fold degenerate codon boxes are colored yellow and 4-fold degenerate boxes are colored blue. Start and stop codons are colored green and red, respectively.

Panel (C) shows a stereo image of the well-structured anticodon loop of tRNA^{Lys} from *E. coli*. Modifications mnm⁵s²U34 and t⁶A37 prevent wrong base pairing inside the 7-nucleotide loop and promote the formation of the conserved U-turn motif. The stacked anticodon bases are located on the same side of the loop. PDB: 1FL8;

Panel (D) shows a stereo image of a collapsed and unmodified anticodon loop of tRNA^{Tyr} from *Bacillus subtilis*. Here, bases 32 and 38, as well as 33 and 37, interact with each other, and the U-turn motif is missing. The anticodon bases are not ordered and are on opposite sides of the loop. PDB: 2LAC.

A prominent example is tRNA^{Ile}, which carries the anticodon UAU. In principle, this anticodon can read codons AUA (for isoleucine) and AUG (for methionine). Yet, it has been shown in some instances that tRNA^{Ile} with an unmodified UAU anticodon exists, but it has a strong preference for its cognate AUA codon, while it rarely misreads AUG. In most organisms, however, tRNA^{Ile} carries the anticodon CUA. To avoid misreading of the methionine codon by this tRNA, C34 (position 1 of the anticodon) is modified to lysidine (k²C34, with the chemical structure shown in Figure 27.1.10, which restricts codon recognition to only AUA and thereby changes the amino acid identity of the tRNA from methionine to isoleucine. In the archaeal species *Haloarcula marismortui*, *Methanococcus maripaludis*, and *Sulfolobus solfataricus*, this tRNA^{Ile} carries a different modification at C34, serving the same purpose of restricting interaction to AUA codons. Here, the original cytosine is modified at the C2-oxo position, which is replaced by agmatine (decarboxy-arginine), resulting in agmatidine (C⁺ or agm²C), as shown in Figure [Math Processing Error]. A complementary modification is that of N⁴-acetylcytosine (ac⁴C34, whose chemical structure is shown in Figure 27.1.8) in the elongator-tRNA^{Met} of *E. coli*, which prevents the recognition of the AUA isoleucine codon. In non-plant mitochondria, however, both AUG and AUA codons are read as methionine. Hence, mitochondrial tRNA^{Met} (carrying the anticodon CAU) has to recognize both codon forms. This is achieved by the introduction of 5-formylcytidine (f⁵C, Figure 11, at position 34, a modification that pairs with both A and U residues at the corresponding codon position 3.

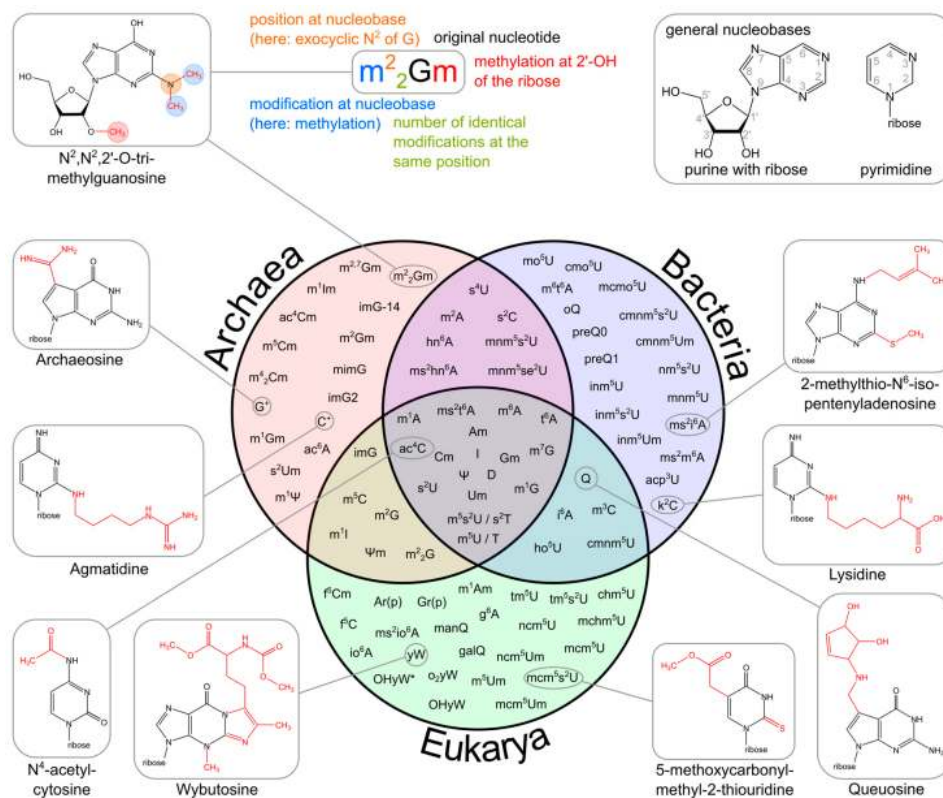


Figure [Math Processing Error]: Variability of tRNA Modifications. Lorenz, C., et. al. (2017) *Biomolecules* 7(2):35

The upper part of the image illustrates the systematic abbreviation of RNA modifications with N²,N²,2'-O-trimethylguanosine (m²₂Gm) as an example. It also shows the atom numbering in the purine and pyrimidine rings as well as in the ribose. An abbreviation preceding the base letter indicates a base modification, whereas letters following the base denote ribose alterations.

Superscripted numbers specify the position at the base, and subscripted numbers indicate the frequency of identical modification at the same position. Abbreviations are as follows: ac—acetyl, acp—aminocarboxypropyl, chm—carboxyhydroxymethyl, cmo—oxyacetic acid, cmnm—carboxymethylaminomethyl, f—formyl, g—glycyl, gal—galactosyl, hn—hydroxynorvalylcarbamoyl, ho—hydroxy, i—isopentenyl, inm—isopentenylaminomethyl, io—cis-hydroxyisopentenyl, m—methyl, man—mannosyl, mchm—carboxyhydroxymethyl methyl ester, mcm—methoxycarbonylmethyl, mcmo—oxyacetic acid methyl ester, mnm—methylaminomethyl, mo—methoxy, ncm—carbamoylmethyl, nm—aminomethyl, r(p)—5-O-phosphono-b-d-ribofuranosyl, s—thio, se—seleno, t—threonylcarbamoyl, tm—taurinomethyl. The Venn diagram summarizes data collected from the RNA modification database and contains the 93 post-transcriptional modifications that are found in tRNAs. Several examples mentioned throughout the text are accompanied by their corresponding chemical structures.

Organisms vary in the number of tRNA genes in their genome. For example, the nematode worm *C. elegans*, a commonly used model organism in genetics studies, has 29,647 genes in its nuclear genome, of which 620 code for transfer RNA (tRNA). The budding yeast *Saccharomyces cerevisiae* has 275 tRNA genes in its genome.

The human genome contains approximately 20,848 protein-coding genes, of which 497 are nuclear genes that encode cytoplasmic tRNA molecules, and 324 tRNA-derived pseudogenes (tRNA genes considered to be non-functional). Regions in nuclear chromosomes, very similar in sequence to mitochondrial tRNA genes, have also been identified (tRNA-lookalikes). These tRNA-look-alikes are also considered part of the nuclear mitochondrial DNA, which comprises genes transferred from the mitochondria to the nucleus. As with all eukaryotes, there are 22 mitochondrial tRNA genes in humans. Mutations in some of these genes have been associated with severe diseases, such as MELAS syndrome.

Cytoplasmic tRNA genes can be grouped into 49 families according to their anticodon features. These genes are found on all chromosomes, except the 22nd and the Y chromosomes. High clustering on 6p is observed (140 tRNA genes), as well as on chromosome 1. Currently, it is unclear why there is so much redundancy within the genome to decode 61 of the 64 possible codons (the other three are stop codons used to terminate translation).

26.1.3: Aminoacyl tRNA Synthetases

Aminoacyl-tRNA synthetases (aaRSs) are universally distributed enzymes that catalyze the esterification of a tRNA to its cognate amino acid (i.e., the amino acid corresponding to the anticodon triplet of the tRNA according to the genetic code). The product of this reaction, an aminoacyl-tRNA (aa-tRNA), is delivered by elongation factors to the ribosome to take part in protein synthesis.

Aminoacyl-tRNA synthetases are named after the aminoacyl-tRNA product generated; as such, methionyl-tRNA synthetase (abbreviated as MetRS) charges tRNA^{Met} with methionine. In eukaryotes, an alternative nomenclature is often employed using the one-letter code of the amino acid (MARS), and a number is added to refer to the cytosolic (MARS1) or the mitochondrial (MARS2) variants. A total of 23 aaRSs have been described so far, one for each of the 20 proteinogenic amino acids (except for lysine, for which there are two) plus pyrrolysyl-tRNA synthetase (PylRS) and phosphoseryl-tRNA synthetase (SepRS), enzymes with a more restricted distribution that are only found in some bacterial and archaeal genomes. It is also worth noting that in eukaryotes, the protein synthesis machinery of mitochondria and chloroplasts generally utilizes their own, bacterial-like sets of synthetases and tRNAs that are distinct from their cytosolic counterparts.

The aminoacyl-tRNA synthetases catalyze a two-step reaction that leads to the esterification of an amino acid to the 3' end of a tRNA, accompanied by the hydrolysis of one molecule of ATP, yielding aminoacyl-tRNA, AMP, and PP_i. In the first step, termed amino acid activation, both the amino acid and ATP bind to the catalytic site of the enzyme, triggering a nucleophilic attack of the α -carboxylate oxygen of the amino acid to the α -phosphate group of the ATP, condensing into aminoacyl-adenylate (aa-AMP), which remains bound to the enzyme, and PP_i, which is expelled from the active site, as shown in Figure [\[Math Processing Error\]](#).

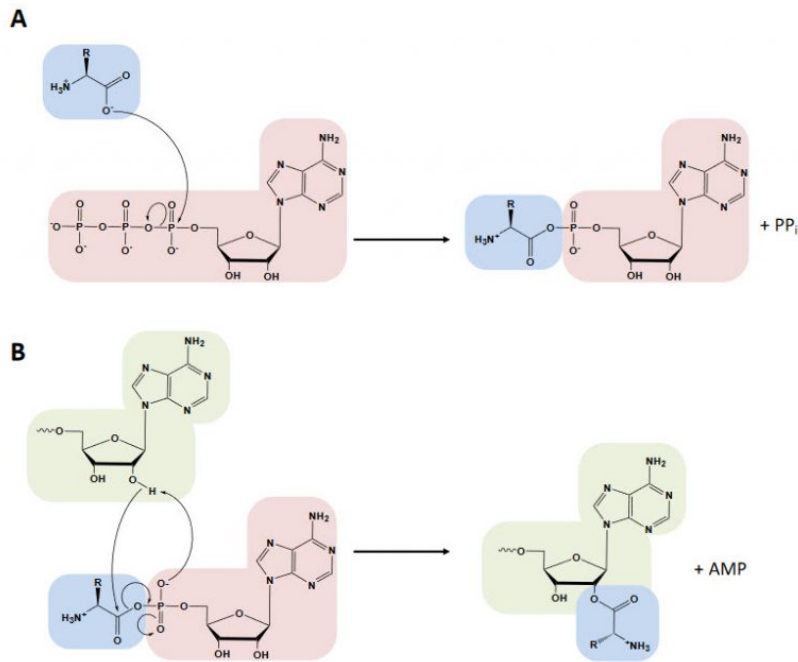


Figure *[Math Processing Error]*: The Aminoacylation Reaction. In the first step, the amino acid (blue) is activated with ATP (red) in the synthetase active site (not depicted), forming the aminoacyl-AMP and releasing PP_i. The amino acid is transferred to the tRNA (green) and AMP is released (depicted in the image transfer to the 3'-OH characteristic of class I aaRS, while in class II transfer happens at the 2'-OH with a 3'-OH attack in the second step). Gomez, M.A.R., and Ibba M. (2020) RNA, doi: 10.1261/ma.071720.119

Although tRNA is usually not required for this first step, certain synthetases do require the tRNA species for productive amino acid activation. In the second part of the reaction, either the 2'- or 3'-hydroxyl group of the terminal adenine nucleotide attacks the carbonyl carbon of the adenylate, forming aminoacyl-tRNA and AMP (Figure 27.1.12 B). While the two-step aminoacylation reaction is universally conserved, the aaRSs that catalyze it show extensive structural and, in some instances, functional diversity.

The 23 known aaRSs can be divided into two major classes based on the architecture of their active sites: Class I and Class II. In class I synthetases, the catalytic domain bears a dinucleotide or Rossmann fold (RF) featuring a five-stranded parallel β -sheet connected by α -helices. It is usually located at or near the N-terminus of the protein. This RF contains the highly conserved motifs HIGH and KMSKS, separated by a connecting domain termed connective peptide 1 (CP1), as shown in Figure *[Math Processing Error]* (panel A). Class II active site architecture is organized as seven-stranded β -sheets flanked by α -helices and features three motifs that show a lesser degree of conservation than those in class I (panel B). Both classes also exhibit pronounced differences in their modes of substrate binding. Class I aaRSs bind the minor groove of the tRNA acceptor stem (with the exceptions of TrpRS and TyrRS) and aminoacylate the 2'-OH of the ribose of the terminal adenosine, while class II aaRSs approach tRNA from the major groove and transfer amino acid to the 3'-OH (except PheRS). The mode of ATP binding is also different between both classes, being bound in an extended configuration in class I, while class II binds a bent configuration with the γ -phosphate folding back over the adenine ring. The kinetics of the aminoacylation reaction can also be used as a distinctive mechanistic signature, as aminoacyl-tRNA release is the rate-limiting step for class I enzymes (except for IleRS and some GluRS), while for class II, it is the amino acid activation rate instead.

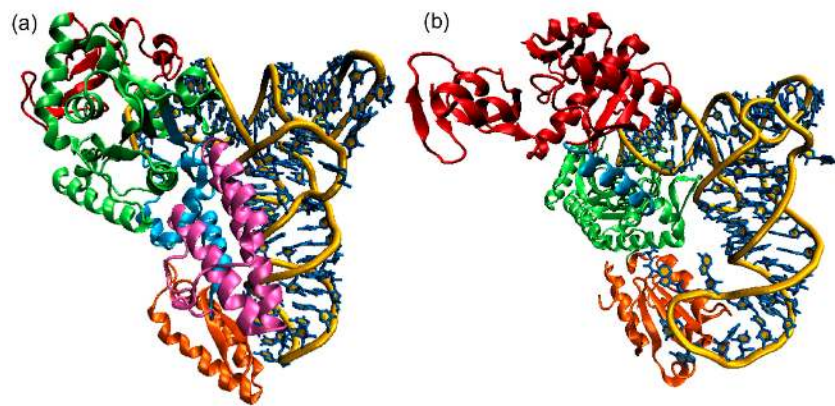
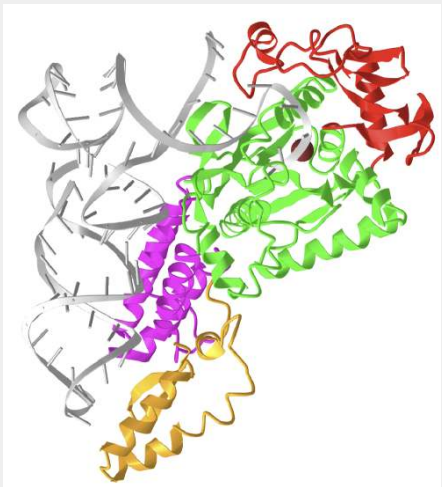
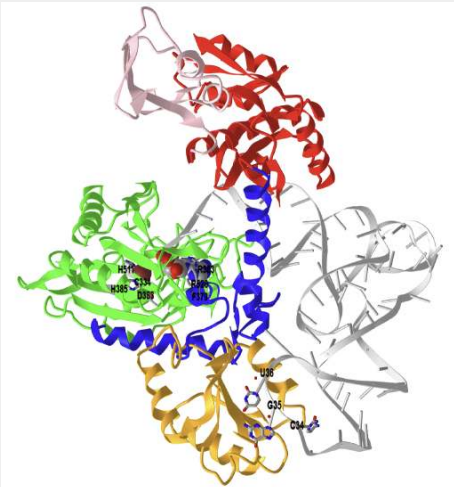


Figure [Math Processing Error]: Structures of Class I and Class II AARSs. Li, R., et. al. (2015) Int J. Mol. Sci. 16(7):15872-15902.

Panel (a) shows the *E. coli* CysRS:tRNA^{Cys} complex. The CP domain (red) and Rossmann fold catalytic domain (green), stem contact fold (cyan), helical bundle domain (magenta), and anticodon binding domain (orange) of CysRS are shown in a ribbon diagram;

Panel (b) A single monomer of the homodimeric *E. coli* ThrRS:tRNA^{Thr} complex. The two N-terminal domains (red), catalytic domain (green), linker (cyan), and anticodon binding domain (orange) of ThrRS are shown in a ribbon diagram. For both structures, the tRNAs are shown in a stick diagram (blue) with a trace of their backbone (yellow).

Figure [Math Processing Error] shows an [interactive iCn3D model](#) of the Class I and II aminoacyl-tRNA synthetases

Class I E. Coli Cysteinyl-tRNA synthetase -tRNA(Cys) (1U0B)	class II E. coli threonyl-tRNA synthase - tRNA(Thr) (1QF6)
	
(Copyright; author via source). Click the image for a popup or use this external link: https://structure.ncbi.nlm.nih.gov/i...P4oRWxKvJHFDL6	(Copyright; author via source). Click the image for a popup or use this external link: https://structure.ncbi.nlm.nih.gov/i...6aXwfqXGCpD7c6

The color coding is analogous to Figure [Math Processing Error] with the tRNA shown in gray.

For the Class II *E. coli* threonyl-tRNA synthase:tRNA(Thr), key conserved amino acids in the active site are shown in CPK-colored sticks and labeled. The anticodon in the tRNA is also shown in colored sticks. Note the large conformational change in the anticodon loop.

The *E. coli* threonyl-tRNA synthetase is interesting in that it represses the translation of its mRNA. A Zn²⁺ ion is involved in binding specificity for the amino acid. It is coordinated by H385, H511, and a water molecule (not shown). R363 interacts with the alpha phosphate, while F379 and R520 align on both sides of the adenine ring. D383 interacts with the amine group of the substrate threonine.

To ensure the faithful translation of the genetic message, synthetases must identify and pair particular tRNAs with their cognate amino acid, which relies on the proper recognition of both substrates. This can prove highly challenging for the synthetases, as not only do they have to discriminate the correct tRNA isoacceptor among a set of other tRNAs that are very similar in structure and chemical composition, but also be able to select the cognate amino acid amidst an extremely large pool of similar amino acids, both proteinogenic and non-proteinogenic. The evolutionary pressure to maintain fidelity has driven aaRSs to develop an elevated specificity for their substrates, both the tRNA and the amino acid.

In addition, some synthetases have evolved editing activities that specifically target and hydrolyze misactivated amino acids and/or misacylated tRNAs. To date, editing activity has been described in 10 out of the 23 aaRSs. In class I synthetases, this activity is located in the highly conserved CP1 domain, although in some enzymes such as MetRS and LysRS, the editing activity resides in the catalytic site. In class II synthetases, however, the editing domains are more idiosyncratic. Editing mechanisms can be divided into two categories, *pre-transfer* or *post-transfer editing*, depending on whether the editing occurs before or after the transfer of the amino acid to the tRNA, as shown in Figure [Math Processing Error].

Pre-transfer editing has been described in both class I and class II aaRSs and takes place after aa-AMP synthesis but before the aminoacyl moiety is transferred to the tRNA. Although the tRNA does not participate in the reaction itself, it has been reported that tRNA binding promotes editing activity in some aaRSs and is a requirement in IleRS and LeuRS. Pre-transfer editing can follow two main pathways. The first one is the selective release of the aa-AMP to the cytosol, where the labile phosphoester bond is spontaneously hydrolyzed. The second route involves the enzymatic breakdown of the product and may happen either in the active site or in an independent editing site.

Post-transfer editing occurs after the transfer of the amino acid to the tRNA and involves the hydrolysis of the ester bond in a domain separate from the active site. The specific mechanism of editing is idiosyncratic to each synthetase, but in general, once formed, the aa-tRNA triggers a conformational change, and the 3' terminus containing the aa is translocated from the active site to the editing site, sometimes traversing distances as large as 40 Å. As the core of the tRNA remains bound to the enzyme, this translocation often involves a rearrangement of the 3' terminus to relocate to the editing site.

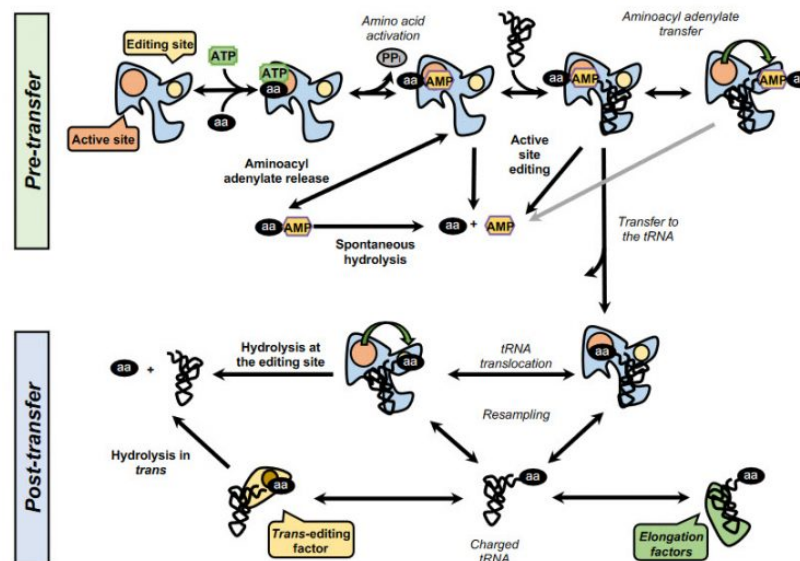


Figure [Math Processing Error]: Aminoacyl tRNA Synthetase Editing Pathways. Schematic overview of the editing pathways employed by the synthetases. In the figure above, the events are in italics, while the editing paths are in bold. The pathways are divided between pre-transfer and post-transfer pathways. In the *pre-transfer editing*, the activated non-cognate aminoacyl-adenylate may be released from the enzyme and hydrolyze spontaneously or be edited within the active site or a specialized active site. During *post-transfer editing*, the aminoacyl-tRNA can be translocated to the editing site or released and cleared by a dedicated trans-editing factor. When a correct cognate has formed, the resulting aminoacyl-tRNA binds the elongation factors and proceeds to translation in the ribosome. Figure from: Gomez, M.A.R., and Ibba M. (2020) RNA, doi: 10.1261/rna.071720.119

26.1.4: Ribosome Structure

The ribosome is a highly conserved molecular machine. In all organisms, it is composed of two unequal subunits, which consist of a distinct set of ribosomal RNA (rRNA) and ribosomal proteins (RPs) that combine to form a large nucleoprotein complex. The ribosome structures in all living organisms harbor three different tRNA binding sites: The A-site, where decoding occurs and the

correct aminoacyl-tRNA (aa-tRNA) is selected based on the mRNA codon displayed, the P-site, which carries the peptidyl-tRNA, and the E-site, which binds exclusively deacylated tRNAs that are exiting the ribosome. Thus, during translation, the tRNA moves from the A-site through the P- and E-sites, where it leaves the ribosome, as shown in Figure [Math Processing Error].

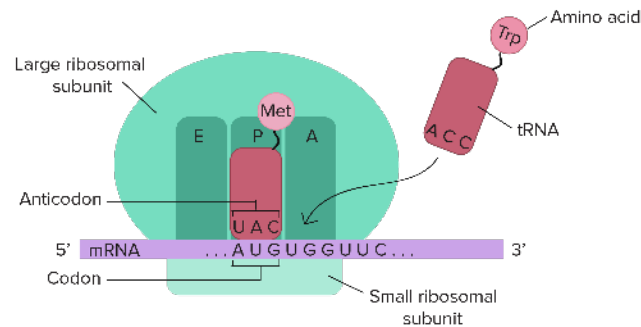


Figure [Math Processing Error]: **Schematic Structure of an Active Ribosome.** The Kahn Academy where it was modified from Openstax College Biology

The mRNA (shown in purple) is assembled between the small subunit and the large subunit of the ribosome (shown in green). tRNA molecules (shown in red) that are loaded with their cognate amino acid (shown in pink) are transitioned through the A-P-E sites of the ribosome during the elongation phase of translation. Movement of the tRNA molecules also shifts the position of the mRNA, causing the next three codon bases to line up in the A-site of the ribosome.

The catalytic peptidyl transferase activity occurs when the tRNA molecules are bound in the A- and P-sites, transferring the nascent peptide to the incoming tRNA molecule (Fig. 27.1.15). Ribosomes are ribozymes because the rRNA performs the catalytic peptidyl transferase activity that links amino acids together. The complexity of the ribosome structure is reflected in the process of protein synthesis, which can be divided into three major steps: initiation, elongation, termination, and recycling.

Ribosomes are either free-floating in the cytoplasm or they can be associated with the intracellular membranes that make up the rough endoplasmic reticulum (rER). Proteins translated into the rER will often be transported out of the cell or embedded into the plasma membrane. These processes are illustrated in Figure [Math Processing Error].

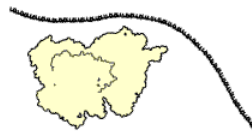


Figure [Math Processing Error]: **A ribosome translating a protein that is secreted into the endoplasmic reticulum.** Figure from: Bensaccount

Ribosomes from bacteria, archaea, and eukaryotes in the three-domain system resemble each other to a remarkable degree. They differ in their size, the majority of their rRNA sequence, and the ratio of protein to RNA. Figure [Math Processing Error] shows the eukaryotic rRNA from the large subunit of the ribosome with highly conserved nucleotide elements (>90% sequence identity) within all of the domains of life, termed universal CNEs or uCNEs, indicated. The differences in sequence and structure between eukaryotes and prokaryotes allow some antibiotics to kill bacteria by inhibiting their ribosomes while leaving human ribosomes unaffected.

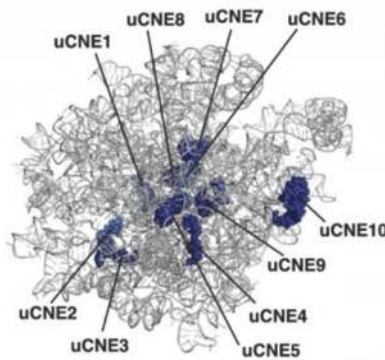


Figure [Math Processing Error]: **Universally Conserved Nucleotide Elements (uCNEs) in the Large Subunit of the Eukaryotic Ribosome.** rRNA from the yeast large subunit of the ribosome is shown from the crown view (from the subunit interface) with the L1 stalk at the upper left. uCNEs that are $\geq 90\%$ conserved in sequence in all domains of life are indicated in dark blue. Figure from: Doris, S.M., et. al. (2015) RNA 21:1719-1730

In all species, more than one ribosome may move along a single mRNA chain at one time (as a *polysome*), each “reading” its sequence and producing a corresponding protein molecule. In this way, many proteins can be translated from a single mRNA molecule. Within bacteria, translation is also coupled with transcription, as the two processes are not physically separated from one another. This is illustrated in Figure [Math Processing Error]. In eukaryotic organisms, *polysomes* form during the process of translation. However, transcription and translation are not coupled, as the processes are separated into the nucleus and cytoplasm, respectively.

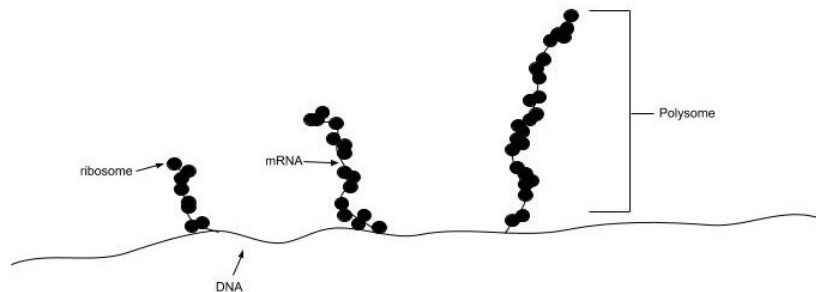


Figure [Math Processing Error]: **Bacterial Transcription and Translation are Coupled.** Translation of nascent mRNA in bacteria occurs before the release of the nascent mRNA from the RNA Polymerase and DNA template. Multiple ribosomes assemble on the nascent mRNA and begin the translation of multiple protein products from the newly formed mRNA template. The structure of the ribosomes assembled on the mRNA is referred to as a *polysome*. Figure from: Gonzalezmg1

The mitochondrial ribosomes of eukaryotic cells share many functional features with those in bacteria, reflecting the likely evolutionary origin of mitochondria.

26.1.4.1: Prokaryotic Ribosome Structure

Prokaryotic ribosomes have a mass of about 2500 kDa and a size of 70S (or Svedberg units: A Svedberg unit is a measure of the sedimentation rate in a centrifuge and thus is representative of size). A complete ribosome (70S) can be dissociated into a large subunit (50S) and a small subunit (30S), as shown in Figure [Math Processing Error]. The small subunit is formed by the interactions of 21 different proteins and a 16S RNA molecule, whereas the large subunit contains 34 different proteins and two RNA molecules, a 23S, and a 5S species.

The rate-limiting step in protein synthesis is the formation of the 70S initiation complex, which will be discussed in detail in the next section.

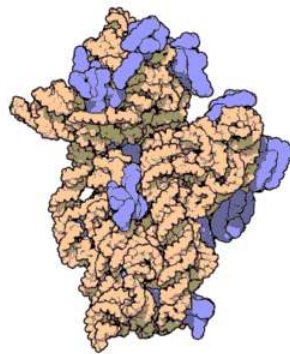


Figure [Math Processing Error]: The Small and Large Prokaryotic Ribosomal Subunits (Left) Atomic structure of the 30S subunit from *Thermus thermophilus*. Proteins are shown in blue, and the single RNA chain is in orange. (Right) Atomic structure of the 50S subunit from *Haloarcula marismortui*. Proteins are shown in blue, and the two RNA chains are in orange and yellow. The small patch of green in the center of the subunit is the active site where the peptidyl transfer reaction occurs. Figure from: Goodsell, D. (2000) Molecule of the Month

Figure [Math Processing Error] shows an [interactive iCn3D model](#) of the E. Coli ribosome (7K00).

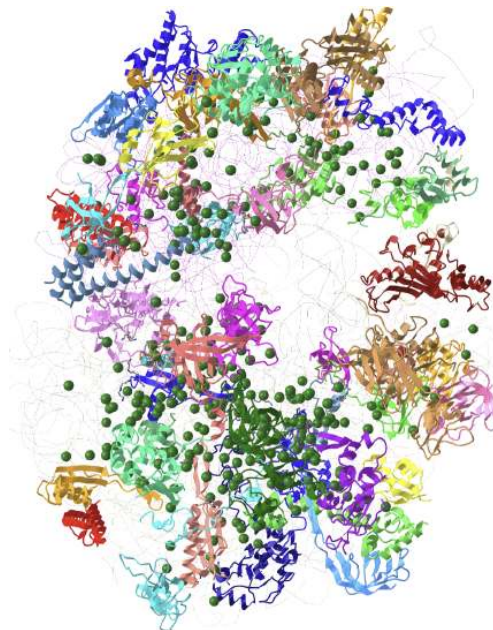


Figure [Math Processing Error]: E. Coli ribosome (7K00). (Copyright; author via source). Click the image for a popup or use this external link: <https://structure.ncbi.nlm.nih.gov/icn3d/share.html?51UCkQmuV3BBqeEJ6>

The RNA is shown in a faint trace backbone. The proteins are shown as cartoons with different colors. Note the large number of Mg^{2+} ions.

The structure was determined using cryo-EM at high resolution. Some ribosomal proteins have isopeptide bonds (formed using side-chain amine and carboxyl groups) as well as some thioamide backbone replacements for the usual amide links.

26.1.4.2: Eukaryotic Ribosome Structure

Eukaryotic ribosomes are larger than their prokaryotic counterparts at approximately 80S (although there is some modest variation between eukaryotic species). Human cytosolic ribosomes are composed of a large subunit (60S) that contains the 28S, 5.8S, and 5S rRNAs and 47 ribosomal proteins (RPs), and a small subunit (40S) that contains the 18S rRNA and 33 RPs.

The assembly of eukaryotic ribosomal subunits starts in the nucleolus, where RNA polymerase I transcribes the major rRNA precursor (a 45S pre-rRNA), from which, after processing and removal of the external and internal transcribed spacers (ETS and ITS), the mature rRNAs are generated, as shown in Figure [Math Processing Error]. The pre-rRNA is modified during transcription by small nucleolar ribonucleoproteins (snoRNPs), processed by RNA nucleases, and assembled with numerous RPs. After

processing the rRNA precursor, the pre-40S and pre-60S subunits follow separate biogenesis routes. Here we will describe the assembly of the 60S subunit in more detail.

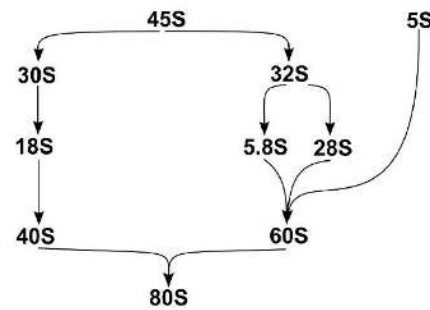


Figure [Math Processing Error]: Schematic Representation of Ribosomal RNA Processing in Eukaryotes. The 45S rRNA transcript is processed by endonucleotide cleavage and splicing to form the 5.8S and 28S components of the large ribosomal subunit and the 18S component of the small ribosomal subunit. The 5S subunit is transcribed independently from the other rRNA components and is incorporated into the large ribosomal subunit. Figure from FeatherPluma

Although the exact assembly of the 60S subunit is not currently known, a model has been postulated that suggests that in the nucleolus, after circularization of rRNA domains, early 60S assembly is carried out sequentially, as shown in Figure [Math Processing Error]. As the transcription of the pre-rRNA proceeds, the rRNA quickly develops a core secondary structure that facilitates the interaction of key Assembly Factors (AFs) and RPs prior to transcriptional termination. Specifically, during this time, the 5.8S portion, ITS2, domains I and II, and partially domain VI are folded in the earliest observed intermediate (state A in Fig. 23). Thus, it appears that the solvent-exposed back side of the large subunit forms like an exoskeleton and construction proceeds by formation of the exit tunnel. This model agrees with a previously suggested model of hierarchical folding based on RP depletion phenotypes. The peptidyl transferase center (PET) is predicted to be one of the later folding steps in the process (state F in Fig. Fig 23) Although the very late-folding peptidyl transferase center is the evolutionarily oldest part of the ribosome, it is likely that folding the solvent side first brings the advantage of providing a stable scaffold for the developing 60S subunit. The folding and assembly of the 40S subunit follow a similar pattern. The two subunits remain unattached until activated in the cytoplasm through the binding of a mRNA transcript with the small subunit. This marks the initiation of the formation of the initiation complex, which will signal the start of the transcriptional process.

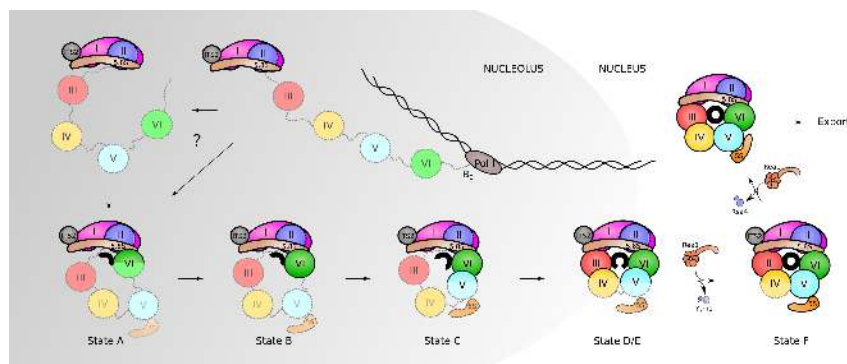


Figure [Math Processing Error]: Assembly Sequence of the Pre-rRNA Domains in the Large Subunit of a Eukaryotic Ribosome. Kater, L., et. al. (2017) Cell 171(7):1599-1610

Assembly of RPs and AFs to the nascent 35S rRNA precursor starts co-transcriptionally. Very early on, the pre-rRNA is circularized as domain VI binds to domains I and II, and the 5.8S portion of the precursor rRNA. The formation of the PET (displayed here as a black circle) starts with this circularization. Its maturation progresses as rRNA domains fold in the following order: VI, V, III, and IV. Full assembly of the PET is only achieved when domain V is completely folded, as observed in state F. After that, only a few additional steps need to occur before the particles are exported to the cytoplasm, where they undergo final maturation.

26.1.5: Summary

This chapter provides an integrated overview of translation—the process by which the information encoded in mRNA is converted into functional proteins—by examining its chemical, structural, and mechanistic foundations across both prokaryotic and eukaryotic systems.

- **Translation Fundamentals and Compartmentalization:**

The chapter begins by contrasting the spatial and temporal dynamics of translation in prokaryotes and eukaryotes. In prokaryotes, the absence of compartmentalization permits simultaneous transcription and translation (coupled processes), whereas in eukaryotes, transcription occurs in the nucleus and translation in the cytoplasm, adding additional layers of regulation.

- **Peptide Bond Formation and Directionality:**

The synthesis of proteins is driven by the formation of peptide bonds through a dehydration reaction, which joins the carboxylic acid of one amino acid with the amine group of another. The ribosome—a ribozyme composed of rRNA and proteins—catalyzes this process directionally from the amino (N) terminus to the carboxyl (C) terminus.

- **mRNA Decoding and the Genetic Code:**

The ribosome reads mRNA in non-overlapping triplets known as codons, with each codon specifying one amino acid. The genetic code is highly degenerate, meaning that multiple codons can code for the same amino acid; this degeneracy, particularly at the wobble base position, helps minimize the impact of mutations. The universality of the genetic code, with few exceptions (e.g., in mitochondria), underscores a common evolutionary heritage.

- **tRNA Structure, Maturation, and Modifications:**

Transfer RNAs serve as the crucial adaptors that match codons with their corresponding amino acids. The chapter reviews the canonical cloverleaf secondary structure and the L-shaped tertiary structure of tRNAs, highlighting their essential features—the acceptor stem, D-arm, anticodon arm, variable loop, and TΨC-arm. It further details the maturation process, which includes removal of leader and trailer sequences, intron splicing, and the post-transcriptional addition of the CCA sequence. Extensive post-transcriptional modifications, especially within the anticodon loop and the tRNA core, fine-tune tRNA structure and function, ensuring efficient and accurate translation.

- **Aminoacyl-tRNA Synthetases and Fidelity Mechanisms:**

Aminoacyl-tRNA synthetases (aaRSs) are responsible for charging tRNAs with their cognate amino acids via a two-step reaction. The chapter distinguishes between the two major classes of these enzymes (Class I and Class II), which differ in active site architecture, substrate binding, and the hydroxyl group targeted for aminoacylation. Additionally, it explores the editing mechanisms—both pre-transfer and post-transfer—that maintain the high fidelity of protein synthesis by preventing misacylation.

- **Ribosome Structure and Assembly:**

A central focus is the structure of the ribosome, the molecular machine that orchestrates translation. In both prokaryotes (70S ribosomes) and eukaryotes (80S ribosomes), the ribosome is organized into two subunits that house three functional sites (A, P, and E sites) for tRNA binding and mRNA decoding. The chapter discusses the structural differences and similarities between ribosomes across domains of life, including details on rRNA components and ribosomal proteins. It also covers ribosome assembly—from co-transcriptional rRNA processing in the nucleolus (for eukaryotes) to the formation of polysomes, where multiple ribosomes simultaneously translate a single mRNA.

Overall, the chapter synthesizes key concepts in translation—from the decoding of mRNA and precise peptide bond formation to the intricate interplay between tRNA modifications, aminoacyl-tRNA synthetase specificity, and ribosomal architecture. Understanding these processes is essential for appreciating how cells accurately synthesize proteins, which in turn underpins all aspects of cellular function and regulation in both prokaryotic and eukaryotic organisms. This comprehensive overview equips students with a solid foundation for further exploration of the molecular mechanisms that drive gene expression and protein synthesis.

26.1.6: References

This chapter was remixed and adapted from the following resources under [creative commons licensing](#):

1. Wikipedia contributors. (2020, June 27). Wobble base pair. In *Wikipedia, The Free Encyclopedia*. Retrieved 16:47, August 12, 2020, from https://en.Wikipedia.org/w/index.php?title=Wobble_base_pair&oldid=964760055
2. Lorenz, C., Lünse, C., and Mörl, M. (2017) tRNA modifications: Impact on the structure and thermal adaptation. *Biomolecules* 7(2)35. Available at: <https://www.mdpi.com/2218-273X/7/2/35/html>
3. Pan, T. (2018) Modifications and functional genomics of human transfer RNA. *Cell Research* 28:395-404. Available at: <https://www.nature.com/articles/s41422-018-0013-y#Sec3>

4. Bednářová, A., Hanna, M., Durham, I., Van Cleave, T., England, A., Chaudhuri, A., and Krishnan, N. (2017) Lost in translation: Defects in transfer RNA modifications and neurological disorders. *Front. Mol. Neurosci.* 10:135. Available at: https://www.researchgate.net/publication/316440980_Lost_in_Translation_Defects_in_Transfer_RNA_Modifications_and_Neurological_Disorders
5. Wikipedia contributors. (2020, May 17). Transfer RNA. In *Wikipedia, The Free Encyclopedia*. Retrieved 23:03, August 15, 2020, from https://en.Wikipedia.org/w/index.php?title=Transfer_RNA&oldid=957227343
6. Gomez, M.A.R., and Ibba M. (2020) Aminoacyl-tRNA Synthetases. *RNA*, doi: 10.1261/rna.071720.119 Available at: <https://rnajournal.cshlp.org/content/early/2020/04/17/rna.071720.119.abstract>
7. Li, R., Macnamara, L.M., Leuchter, J.D., Alexander, R.W., and Cho, S.S. (2015) MD Simulations of tRNA and Aminoacyl-tRNA Synthetases: Dynamics, Folding, Binding, and Allostery. *Int. J. Mol. Sci.* 16(7):15872-15902. Available at: <https://www.mdpi.com/1422-0067/16/7/15872>
8. Wikipedia contributors. (2020, August 15). Ribosome. In *Wikipedia, The Free Encyclopedia*. Retrieved 05:22, August 16, 2020, from <https://en.Wikipedia.org/w/index.php?title=Ribosome&oldid=973144885>
9. Kater, L., Thoms, M., Barrio-Garcia, C., Cheng, J., Ismail, S., Ahmed, Y.L., Bange, G., Kressler, D., Berninghausen, O., Sinning, I., Hurt, E., and Beckmann, R. (2017) Visualizing the assembly pathway of nucleolar Pre-60S Ribosomes. *Cell* 171(7):1599-1610. Available at: <https://www.sciencedirect.com/science/article/pii/S0092867417314290>
10. Bock, L.V., Kolár, M.H., Grubmüller, H. (2018) Molecular simulations of the ribosome and associated translation factors. *Cur. Op. Struc. Bio.* 49:27-35. Available at: <https://www.sciencedirect.com/science/article/pii/S0959440X1730132X>
11. Doris, S.M., Smith, D.R., Beamesderfer, J.N., Raphael, B.J., Nathanson, J.A., and Gerbi, S.A. (2015) Universal and domain-specific sequences in 23S-28S ribosomal RNA identified by computational phylogenetics. *RNA* 21:1719-1730. Available at: https://www.researchgate.net/publication/281141702_Universal_and_domain-specific_sequences_in_23S-28S_ribosomal_RNA_identified_by_computational_phylogenetics
12. Aleksashin, M.A., Leppik, M., Hochenberry, A.J., Klepacki, D., Vázquez-Laslop, N., Jewett, M.C., Remme, J., and Mankin A.S. (2019) Assembly and functionality of the ribosome with tethered subunits. *Nature Communications* 10:930. Available at: <https://www.nature.com/articles/s41467-019-08892-w#rightslink>
13. Wikipedia contributors. (2020, July 3). Formylation. In *Wikipedia, The Free Encyclopedia*. Retrieved 22:03, August 16, 2020, from <https://en.Wikipedia.org/w/index.php?title=Formylation&oldid=965827476>
14. Gualerzi, C.O., and Pon C.L. (2015) Initiation of mRNA translation in bacteria: structural and dynamic aspects. *Cell Mol Life Sci.* 72:4341-4367. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4611024/>
15. Wikipedia contributors. (2020, June 14). Kozak consensus sequence. In *Wikipedia, The Free Encyclopedia*. Retrieved 06:03, August 18, 2020, from https://en.Wikipedia.org/w/index.php?title=Kozak_consensus_sequence&oldid=962444929
16. Knight, J.R.P., Garland, G., Pöyry, T., Mead, E., Vlahov, N., Sfakianos, A., Frosso, S., De-Lima-Hedayioglu, F., Mallucci, G.R., von der Haar, T., Smales, C.M., Sansom, O.J., and Willis, A.E. (2020) Control of translation elongation in health and disease. *Dis. Mod. and Mech.* 13: dmm043208. Available at: <https://dmm.biologists.org/content/13/3/dmm043208>
17. Adio, S., Sharma, H., Senyushkina, T., Karki, P., Maracci, C., Wohlgemuth, I., Holtkamp, W., Peske, R., and Rodina, M.V. (2018) Dynamics of ribosomes and release factors during translation termination in *E. coli*. *eLife* 7:e34253. Available at: <https://elifesciences.org/articles/34252>
18. Ge, X., Oliveira, A., Hjort, K., Bergfors, T., Guitierrez-de-Terán, H., Andersson, D.I., Sanyal, S., and Åqvist, J. (2019) Inhibition of translation termination by small molecules targeting ribosomal release factors. *Scientific Reports* 9: 15424. Available at: <https://www.nature.com/articles/s41598-019-51977-1#rightslink>
19. Svidritskiy, E., Demo G., Loveland A.B., Xu, C., and Korostelev, A.A. (2019) Extensive ribosome and RF2 rearrangements during translation termination. *eLife* 8:e46850. Available at: <https://elifesciences.org/articles/46850>
20. Sauert, M., Temmel, H., and Moll, I. (2015) Heterogeneity of the translational machinery: Variations on a common theme. *Biochimie* 114:39-47. Available at: <https://www.sciencedirect.com/science/article/pii/S0300908414003952>

This page titled **26.1: The Genetic Code** is shared under a **not declared** license and was authored, remixed, and/or curated by **Henry Jakubowski and Patricia Flatt**.

- **11.2: Diffusion Across a Membrane - Passive and Facilitated Diffusion** by **Henry Jakubowski and Patricia Flatt** has no license indicated.
- **Current page** by **Henry Jakubowski and Patricia Flatt** has no license indicated.

26.2: Protein Synthesis

Learning Goals (ChatGPT o3-mini)

- **Understand Ribosomal Subunit Functions:**
 - Describe the roles of the 30S (decoding) and 50S (peptidyl transferase) subunits in translation.
 - Explain how these subunits assemble to form the active 70S ribosome at the start codon and dissociate at the stop codon.
- **Explore the Role of mRNA Features in Initiation:**
 - Identify the Translation Initiation Region (TIR) and its components, including the start codon and Shine–Dalgarno (SD) sequence.
 - Discuss how the SD sequence, located 4–9 nucleotides upstream of the start codon, facilitates proper mRNA docking with the 30S subunit.
 - Analyze how variations in TIR sequences affect translation efficiency.
- **Examine the Function of the Initiator tRNA:**
 - Explain the unique role of the initiator tRNA^{fMet}, including its enzymatic aminoacylation with methionine and subsequent formylation.
 - Understand why formylated methionine (fMet) is used exclusively for initiation and the subsequent removal of the formyl group (and sometimes the methionine residue) after protein synthesis.
- **Detail the Roles of Prokaryotic Initiation Factors:**
 - Outline the functions of the three key initiation factors (IF1, IF2, and IF3) in assembling the 30S initiation complex.
 - Describe how IF1 binds in the A site to stabilize the ribosome, while IF2 (a GTPase) facilitates the recruitment of tRNA^{fMet}, and IF3 ensures the fidelity of start codon selection.
 - Explain the sequential steps of initiation—from the formation of the 30S-IF complex, mRNA and tRNA^{fMet} recruitment, to the joining of the 50S subunit to form the 70S initiation complex.
- **Understand the Role of GTP Hydrolysis in Initiation:**
 - Describe how the binding of GTP to IF2 and its subsequent hydrolysis upon 50S docking induce conformational changes that lead to the release of initiation factors.
 - Analyze how these GTP-dependent steps serve as kinetic checkpoints to ensure initiation fidelity.
- **Contrast Prokaryotic and Eukaryotic Initiation:**
 - Compare the simpler prokaryotic system with the more complex eukaryotic translation initiation, noting the role of additional eukaryotic initiation factors (eIFs) and the scanning mechanism.
 - Explain the significance of mRNA cap recognition, the Kozak sequence, and mRNA circularization in eukaryotic initiation.
- **Integrate Initiation into the Overall Translation Cycle:**
 - Relate how the accurate assembly of the initiation complex sets the stage for efficient elongation and overall protein synthesis.
 - Discuss the importance of initiation factor dissociation (e.g., IF2 leaving the ribosome) in allowing elongation factors to bind and peptide bond formation to commence.

These learning goals aim to ensure that students not only understand the molecular details and steps involved in prokaryotic translation initiation but also appreciate the broader context and regulatory mechanisms that ensure high fidelity during protein synthesis.

26.2.1: Prokaryotic Initiation

The small subunit of the ribosome (the 30S) interprets genetic information by selecting aminoacyl-tRNAs that are cognate to the mRNA codons in the decoding center. The large subunit (the 50S) contains the catalytic peptidyl transferase center, where amino

acids are polymerized into proteins. Small and large subunits unite together at the start codon of a gene to form the 70S ribosome and dissociate again at the stop codon upon completing the synthesis of the encoded protein. This process consists of three phases: initiation, elongation, and termination. In this section, we will focus on the initiation of translation.

In bacteria, the initiation phase of protein synthesis involves a limited number of key actors. Aside from the two ribosomal subunits, key roles are played by the initiator tRNA^{fMet}, the translation initiation region (TIR) of the mRNA, and three protein factors – the initiation factors (IFs) IF1, IF2, and IF3 – that ensure speed and accuracy in the overall process. The initiator tRNA^{fMet} contains a methionine residue that has been enzymatically modified to contain an N-terminal formyl group, as shown in Figure [Math Processing Error]. fMet is only used for the initiation of protein synthesis and is thus found only at the N-terminus of the protein. Unmodified methionine is utilized throughout the remainder of the translation process. Once protein synthesis is completed, the formyl group on methionine may be removed by peptide deformylase. On occasion, the entire methionine residue can be further removed by the enzyme methionine aminopeptidase.

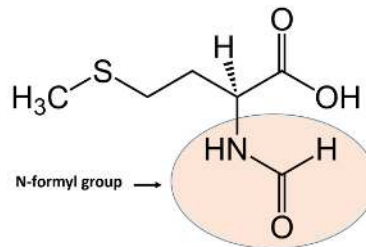


Figure [Math Processing Error]: Structure of N-formylmethionine.

The TIR sequence within the mRNA contains the start codon and usually an upstream untranslated region that interacts with the small subunit of the ribosome. The bacterial cell produces and expresses a plethora of different mRNAs with varying TIR sequences and structures; the efficiency with which these individual transcripts are translated depends not only on their abundance and stability but also on the nature of the TIR. Thus, unlike the other aforementioned actors that represent constants, the mRNA TIRs represent essentially the only variable in the process of mRNA initiation site selection and can affect translation efficiency.

Although the triplet AUG is by far the most frequent initiation codon found in TIRs, other initiation triplets (i.e., GUG, UUG, AUU, AUC, and AUA) are found in bacteria, and the central U is the only universally conserved base of the start codon. Among the triplets mentioned above, those having a 3'-G (i.e., AUG, GUG, and UUG) are recognized equivalently and most efficiently by IF3 during the initiation complex formation.

Another important characteristic of a large number of bacterial mRNA TIRs is the presence of a Shine–Dalgarno (or SD) sequence that is complementary to the 3' end sequence of 16S rRNA (the anti-SD sequence or aSD). The SD sequence, when present, is usually at an optimal distance of 4–9 nucleotides upstream of the initiation codon, as shown in Figure [Math Processing Error]. While the SD sequence plays an important role in the efficient translation of many mRNA transcripts, it is not essential. Many other mRNA sequences fully lack an SD sequence but are still efficiently transcribed. Thus, the SD sequence is only one example of TIR mechanisms that can play an important role in mRNA binding with the small subunit of the ribosome.

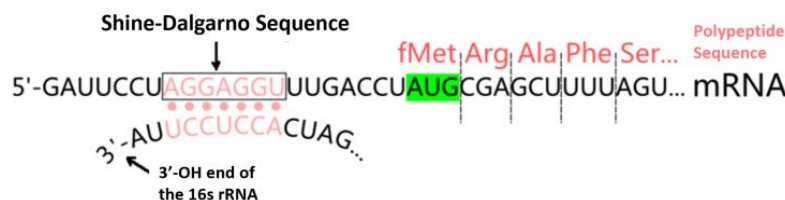


Figure [Math Processing Error]: Shine-Dalgarno Translation Initiation Sequence. Figure modified from [Shakeistone](#)

Prokaryotic mRNA sequences often share a highly conserved sequence upstream of the start codon known as the Shine-Dalgarno sequence. This consensus sequence is complementary to the 3' end of the 16S rRNA sequence in the small subunit of the ribosome. It is a crucial feature for the binding and docking of many mRNA molecules to the small ribosomal subunit during transcription initiation.

The three protein initiation factors, IF1, IF2, and IF3, determine the kinetics and fidelity of the overall initiation process. The three IFs are bound, one copy each, to specific sites of the 30S subunit where they assist with the formation of the initiation complex and assembly of the 70S ribosome.

As noted above, the initiator tRNA is first aminoacylated with methionine whose α -NH₂ group is eventually blocked by a specific formyl transferase (TMF) to produce a tRNA^{fmet} molecule. This modification prevents interaction with the elongation factor EF-Tu (which we will see plays a crucial role in the elongation phase of translation, but not in the initiation phase). Blocking EF-Tu binding ensures instead the recognition and binding of tRNA^{fmet} by initiation factor IF2, effectively docking it with the 30S subunit. Furthermore, tRNA^{fmet} binds with high affinity to the ribosomal P-site, unlike all other aminoacyl-tRNAs that bind to the A-site in a ternary complex with EF-Tu and GTP (details will be presented in the next section). In the P-site, the initiator tRNA must be recognized as correct by the other initiation factors IF3 and IF1.

To form the 30S initiation complex, IF3 and IF2 are the first factors to bind to the 30S subunit, creating an unstable 30S-IF3-IF2 complex, as shown in Figure [Math Processing Error] (panel A). The binding of IF1 causes a conformational change in the 30S subunit, stabilizing the complex and facilitating the recruitment of tRNA^{fmet} by IF2. Notably, IF1 binds in the A site of the 30S subunit, where it contacts ribosomal protein S12. Recruitment of the tRNA^{fmet} can also stabilize the mRNA interactions with the 30S subunit through the formation of hydrogen bonds between the codon of the mRNA and the anticodon of the tRNA^{fmet}. Note that the binding of mRNA to the 30S subunit is IF-independent and can take place at any time during the 30S assembly process. Two potential routes of mRNA association are shown Figure [Math Processing Error], panel A, where the mRNA is assembled either before or after tRNA^{fmet} recruitment.

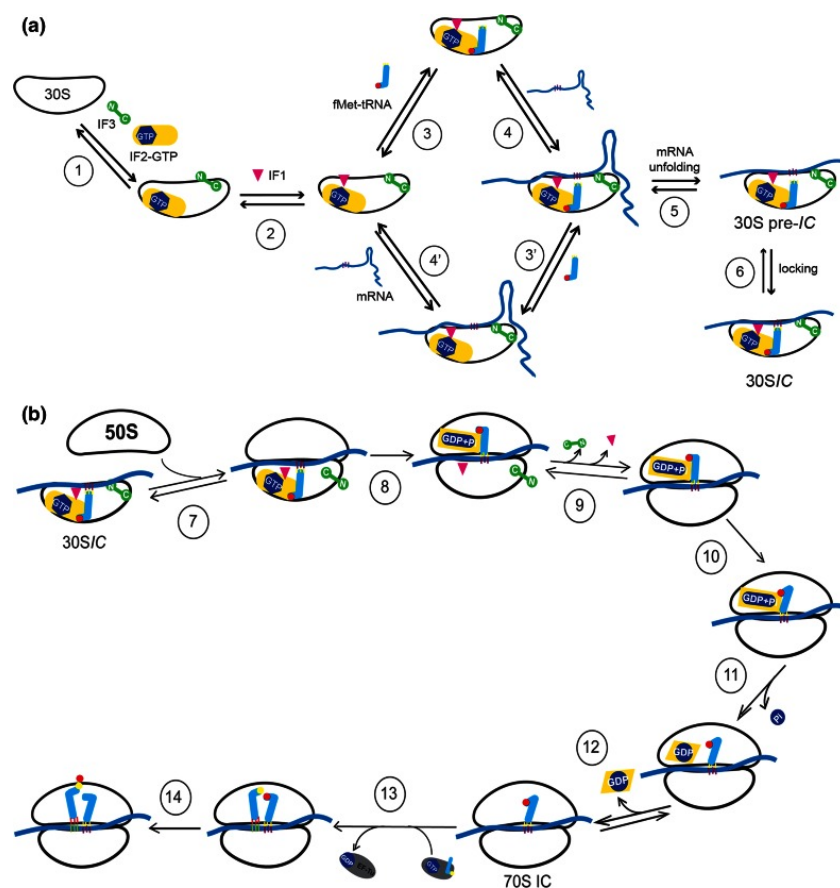


Figure [Math Processing Error]: Initiation of Translation. Figure from: [Gualerzi, C.O., and Pon, C.L. \(2015\) Cell Mol Life Sci. 72:4341-4367.](#)

- Step 1: A vacant 30S ribosomal subunit binds IF3 and IF2.
- Step 2: IF1 binds to the 30S subunit in the presence of both IF3 and IF2.
- Steps 3 and 3': In the presence of all three factors, tRNA^{fmet} is recruited.
- Steps 4 and 4': The mRNA is bound with different on and off rates depending on its TIR structure; mRNAs with strong secondary structures are bound more slowly than those having little or no secondary structure.
- Step 5: mRNAs containing secondary structures must be unfolded in a process that is facilitated by IF2 bound to GTP and antagonized by IF3.

Step 6: The isomerization of the 30S pre-IC allows the P-site codon–anticodon interaction to yield a more stable 30SIC from which mRNA and fMet-tRNA are more stably bound.

Step 7: A 30SIC, containing IF1, IF2-GTP, IF3 and mRNA whose initiation triplet is P-site decoded by fMet-tRNA, is docked by a 50S subunit.

Step 8: Upon contact with the 50S subunit, the GTPase function of IF2 is activated, and GTP is rapidly hydrolyzed, leaving GDP+Pi bound to IF2.

Step 9: This reversible conformational transition represents the last kinetic checkpoint of translation initiation fidelity by IF3 and IF1, as IF3 and IF1 dissociate from the complex.

Step 10: The first-order isomerization of the IF2-GDP structure causes a shift in the ribosome structure, representing the rate-limiting step in 70SIC formation.

Step 11: Pi is dissociated from IF2-GDP.

Step 12: IF2 leaves the ribosome (or moves away from the A-site), clearing the way for EF-Tu binding.

Step 13: the EF-Tu-GTP-aminoacyl-tRNA complex binds to the 70SIC and through several steps (not represented here) delivers to the ribosomal A-site the aminoacyl-tRNA encoded by the second mRNA codon.

Step 14: the tRNA^{fMet} bound in the P-site of the peptidyl transferase center donates its formyl-methionine to the A-site-bound aminoacyl-tRNA to yield the initiation dipeptide fMet-aa. Initiation is then complete, and the elongation phase can begin.

Following the recruitment of the mRNA and the tRNA^{fmet} to the 30S initiation complex loaded with the IF2, IF3, and IF1 initiation factors, the 50S subunit is docked very rapidly to yield an initially unstable 70S initiation complex (Fig. 27.2.3 b). It should be noted that the IF2 protein is a GTP hydrolase enzyme and, as such, binds with the cofactor GTP before the recruitment of the 50S subunit. Contact between the IF2 GTPase-activating center and the 50S subunit causes the rapid hydrolysis of GTP to GDP + Pi.

The formation of the 70S complex causes the dissociation of the initiation factors. IF2 is the last factor to be dissociated, leaving the ribosome after having positioned tRNA^{fMet} in the P-site of the 70S initiation complex. It must be placed in the correct orientation to facilitate peptide bond formation. GDP and Pi also dissociate from the complex with the removal of IF2. The elongation factor, EF-G, is then free to chaperone the first tRNA into the A-site, and the first peptide bond is formed (Step 13 of Fig. 27.2.3 b). This marks the beginning of the elongation phase of protein synthesis.

26.2.2: Eukaryotic Initiation

Eukaryotic translation initiation is more complex than prokaryotic systems and requires the actions of at least 11 eukaryotic initiation factors (eIFs), plus additional auxiliary factors (Table [\[Math Processing Error\]](#)). We will not cover the action of all these eIFs in detail here, but rather focus on a few key steps as outlined in Figure [\[Math Processing Error\]](#).

Table [\[Math Processing Error\]](#): **Comparison of Prokaryotic and Eukaryotic Translation Initiation Factors**

Prokaryotic Factors	Function	GTP Hydrolase	Eukaryotic Factors	Function'	GTP Hydrolase'
IF1	Associates with 30S subunit at A-site. Prevents premature assembly of large subunit	No	eIF1, eIF1A	induce open conformation of mRNA binding channel of 40S subunit	No
IF2	Recruits fMet-tRNA to P-site	Yes	eIF2, eIF5B	delivers tRNAi to 40S subunit	Yes
IF3	Required for correct association of mRNA with small subunit rRNA	No	eIF3	binds to 40S subunit and helps recruit mRNA through interaction with other eIFs	No
			eIF4E	Bind to 5' CAP of mRNA	No
			eIF4G	Binds PABP	No
			eIF4A	RNA helicase important for resolving secondary structure	No
			eIF5	GTPase activating protein that helps eIF2 in recruiting the large subunit	No
			eIF6	Binds large subunit of ribosome and helps with eIF3 to link with small subunit	No
			PABP	Binds to polyA tail of mRNA and links it to 5' end	No

First, the initiator tRNA_i is recruited to the small ribosomal subunit (40S) to form a ternary complex with the GTP-bound eukaryotic initiation factor 2 (eIF2). Additional factors, such as eIF3, strongly enhance the formation of this 43S pre-initiation complex. eIF3 also interacts with the eIF4F complex, which consists of three other initiation factors: eIF4A, eIF4E, and eIF4G. eIF4G is a scaffolding protein that directly associates with both eIF3 and the other two components. eIF4E is the 5'-cap-binding protein. The binding of the mRNA cap by eIF4E is often considered the rate-limiting step in *cap-dependent initiation*, and the concentration of eIF4E serves as a regulatory nexus in translational control. Certain viruses cleave a portion of eIF4G that binds to eIF4E, thus preventing cap-dependent translation and hijacking the host machinery in favor of viral (cap-independent) messages. eIF4A is an ATP-dependent RNA helicase that aids the ribosome by resolving certain secondary structures formed along the mRNA transcript. The poly(A)-binding protein (PABP) also associates with the eIF4F complex via eIF4G and binds the poly-A tail of most eukaryotic mRNA molecules. This protein has been implicated in playing a role in the circularization of the mRNA during translation. The 43S preinitiation complex, accompanied by the protein factors, moves along the mRNA chain toward its 3'-end, in a process known as 'scanning', to reach the start codon (typically AUG). After recognition of the start codon, the large ribosomal subunit (60S) assembles to form the 80S initiation complex, which is ready for elongation. Alternatively, under distinct conditions or on certain transcripts, internal initiation can occur in a cap-independent manner at so-called *internal ribosome entry sites (IRES)*. Eukaryotic translation initiation is shown in Figure [Math Processing Error].

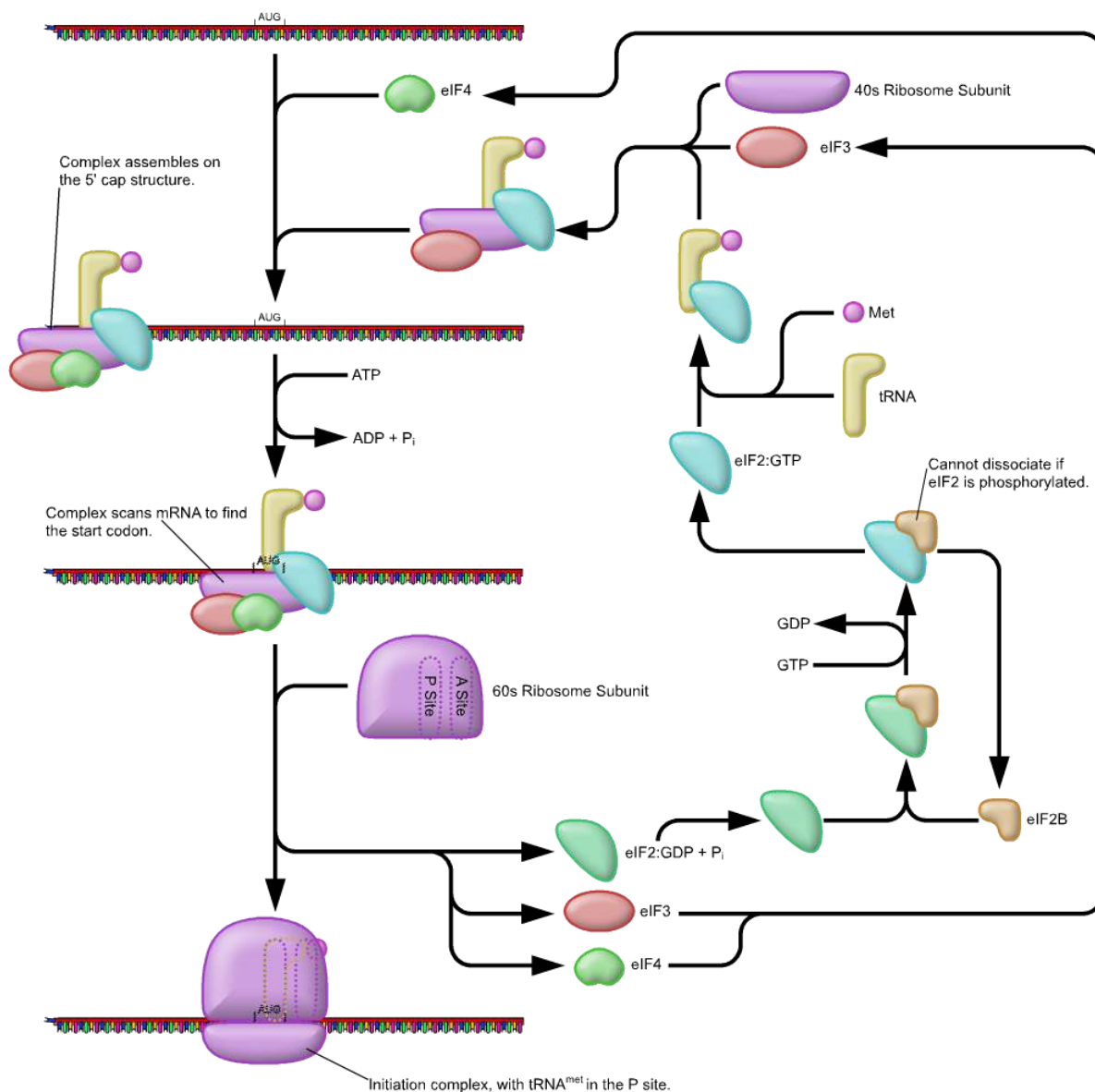


Figure [Math Processing Error]: Eukaryotic Translation Initiation. Figure from: [Eukaryotic Translation, Wikiwand](https://www.wikidwelling.com/wiki/Eukaryotic_Translation)

This is a simplified diagram of eukaryotic translation initiation detailing some of the eIFs involved in the process. eIF2 is critical for recruiting the initiation tRNA_i to the 40S subunit. eIF3 enhances the activity of eIF2 and also promotes the binding of the 43S pre-initiation complex with the mRNA. eIF3 binds to the mRNA through the interaction of the eIF4 factors, causing the scanning of the pre-initiation complex down the mRNA to locate the start codon (usually AUG). Poly A Binding Proteins (PABPs) bind with the polyA tail sequence of the mRNA and also interact with the eIF4 factors, causing the circularization of the mRNA.

As seen in prokaryotic systems with the favored Shine-Dalgarno sequence upstream of the start codon within the mRNA sequence, there are also preferred nucleotide sequences within the local vicinity of the start codon in eukaryotic mRNAs, as well. In eukaryotic mRNA, this is known as the **Kozak sequence** (Figure [Math Processing Error]). The sequence was initially defined as 5'- (gcc)gccRccAUGG-3' where:

1. The underlined nucleotides indicate the translation start codon, coding for Methionine.
2. Upper-case letters indicate highly conserved bases, i.e., the 'AUGG' sequence is constant or rarely, if ever, changes.
3. 'R' indicates that a purine (adenine or guanine) is always observed at this position (with adenine being more frequent according to Kozak rules)
4. A lower-case letter denotes the most common base at a position where the base can nevertheless vary
5. the sequence in parentheses (gcc) is of uncertain significance.

The AUG is the initiation codon encoding a methionine amino acid at the N-terminus of the protein. (Rarely, GUG is used as an initiation codon, but methionine is still the first amino acid as it is the met-tRNA in the initiation complex that binds to the mRNA). Variation within the Kozak sequence alters the "strength" of the translational start site. Kozak sequence strength refers to the favorable ability of initiation, affecting how much protein is synthesized from a given mRNA. This is shown in Figure [Math Processing Error].

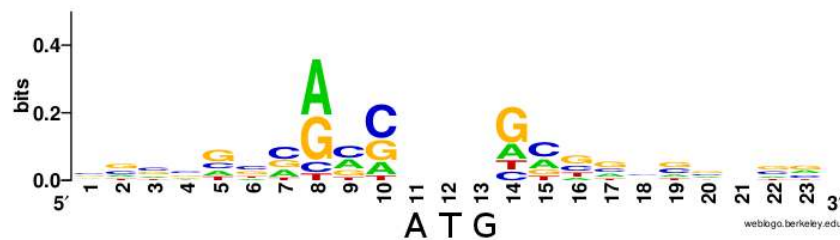


Figure [Math Processing Error]: The Kozak Sequence. The sequence schematic above shows the most conserved bases around the initiation codon from over 10,000 human mRNAs. Larger letters indicate a higher frequency of incorporation. Note the larger size of A and G at the 8 position (-3, Kozak position) and at the G at position 14, which corresponds to (+4) position in the Kozak sequence. Figure from: [Audrey Michel](#)

26.2.3: The Elongation Phase of Translation

Both prokaryotic and eukaryotic elongation phases of transcription utilize similar elongation factors during the process. Table [Math Processing Error] provides a summary of their functions.

Prokaryotic Factors	Function	GTP Hydrolase	Eukaryotic Factors	Function'	GTP Hydrolase'
EF-Tu	chaperones aminoacyl-tRNA into ribosome	Yes	eEF-1 α	chaperones aminoacyl-tRNA into ribosome	Yes
EF-Ts	Recharges EF-Tu with GDP-GTP exchange	No	eEF-1 $\beta\gamma$	Recharges eEF-1 α with GDP-GTP exchange	No
EF-G	Required for translocation of mRNA and tRNAs after peptide bond formation	Yes	eEF-2	Required for translocation of mRNA and tRNAs after peptide bond formation	Yes
EF-P	Helps facilitate protein exit through tunnel and prevent stalling	No	eIF5A	Helps facilitate protein exit through tunnel and prevent stalling	No

Table [Math Processing Error]: Comparison of Prokaryotic and Eukaryotic Translation Elongation Factors

26.2.4: Prokaryotic Elongation

The prokaryotic elongation phase of transcription requires the activity of three primary elongation factors (EFs), EF-Tu, EF-Ts, and EF-G. During elongation, aminoacyl-tRNAs are delivered to the ribosome in the form of a ternary complex: the tRNA, a translational GTPase (in bacteria: EF-Tu or SelB), and a GTP molecule, as shown in Figure [Math Processing Error]. The tRNA

decodes the information on the mRNA by forming hydrogen bonds (H-bonds) between the codon and the corresponding nucleobases of the anticodon. Remarkably, the free-energy difference between correct (cognate) and incorrect (near-cognate, non-cognate) base pairing alone does not explain the very high fidelity of decoding. Rather, high fidelity is achieved by a two-step decoding process: initial selection leading to GTPase activation and proofreading. In addition to the free-energy difference, kinetic effects contribute to the discrimination. The GTP hydrolysis rate is increased, and the recognition of the correct codon decreases the tRNA rejection rate.

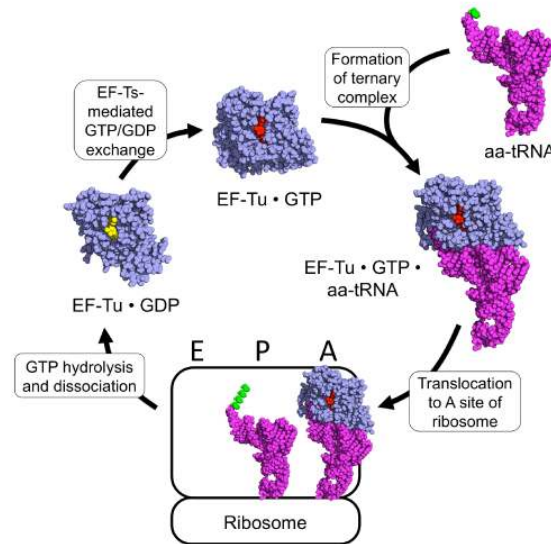


Figure [\[Math Processing Error\]](#): The EF-Tu tRNA Chaperone Protein. Ef-Tu (shown in blue) is a GTPase enzyme chaperone that, when bound to GTP (shown in red), can form a ternary complex with an amino acid-containing tRNA molecule (tRNA shown in purple; amino acid shown in green). The ternary complex translocates to the A-site of the ribosome, where the anticodon of the incoming tRNA hybridizes with the codon of the mRNA. An exact match of the codon-anticodon hybridization leads to the hydrolysis of GTP by EF-Tu and its subsequent dissociation from the ribosome (the hydrolysis product GDP is shown in yellow). EF-Tu is unable to release the GDP on its own. It requires a protein-mediated exchange of the GDP for a molecule of GTP using the elongation factor, EF-Ts. Figure from [Awchen](#)

Small-subunit nucleotides A1492 and A1493 adopt a flipped-out conformation in the presence of a tRNA and, in this conformation, the tRNA anticodon hydrogen bonds with the codon of the mRNA, forming a mini-helix structure as shown in Figure [\[Math Processing Error\]](#). The flipped-out nucleotides A1492, A1493, along with G530, were found to shield the codon-anticodon base pairs from solvent. This shielding of near-cognate base pairs from the solvent is incomplete, increasing the free-energy difference between near-cognate and cognate base pairs and increasing flexibility within the docking region. This reduces the strength of hydrogen bonding between a non-cognate tRNA, causing the inappropriate tRNA to leave the A-site before peptide bond formation can occur. This increases the fidelity and discrimination of tRNA selection, such that only the correct cognate tRNA is incorporated into the A-site.

Figure [Math Processing Error]: Cognate Interaction of tRNA and mRNA in *E. coli*. [Fislage, M., et. al. \(2018\) Nuc. Acids Res. 46\(11\):5861-5874.](#)

Panel (A) shows the interaction of cognate tRNA anticodon (shown in red) with the mRNA codon (shown in blue) when wild type EF-Tu binds and chaperones the tRNA complex to the A-site of the ribosome. Panel (B) is the same figure as shown in (A), with the ribosomal nucleotides A1492 and A1493 depicted in the flipped out conformation.

Interestingly, aminoglycosides, a class of antibiotics, bind to the decoding center and lock nucleotides A1492/A1493 in the flipped-out conformation as shown in Figure [Math Processing Error]. In this way, aminoglycosides promote the accommodation of near-cognate, thus wrong, tRNAs into proteins during synthesis, causing widespread mutagenesis. This is toxic to the bacteria and leads to the death of bacterial cells.

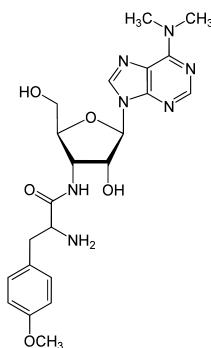


Figure [Math Processing Error]: Structure of the Aminoglycoside Antibiotic, Puromycin. Figure from: [Yikrazuul](#)

After GTP hydrolysis, the GTPase EF-Tu dissociates from the tRNA. At this point, EF-Tu is tightly bound to a molecule of GDP and cannot release it on its own to be recycled for a second round of tRNA chaperoning. The recharging of EF-Tu is executed by the Elongation Factor Thermo stable (EF-Ts), as shown in Figure [Math Processing Error]. The binding of EF-Ts to EF-Tu-GDP causes a conformational change in EF-Tu, allowing the release of GDP. The binding of a new molecule of GTP with the EF-Tu protein causes the dissociation of EF-Ts and fully recharges EF-Tu.

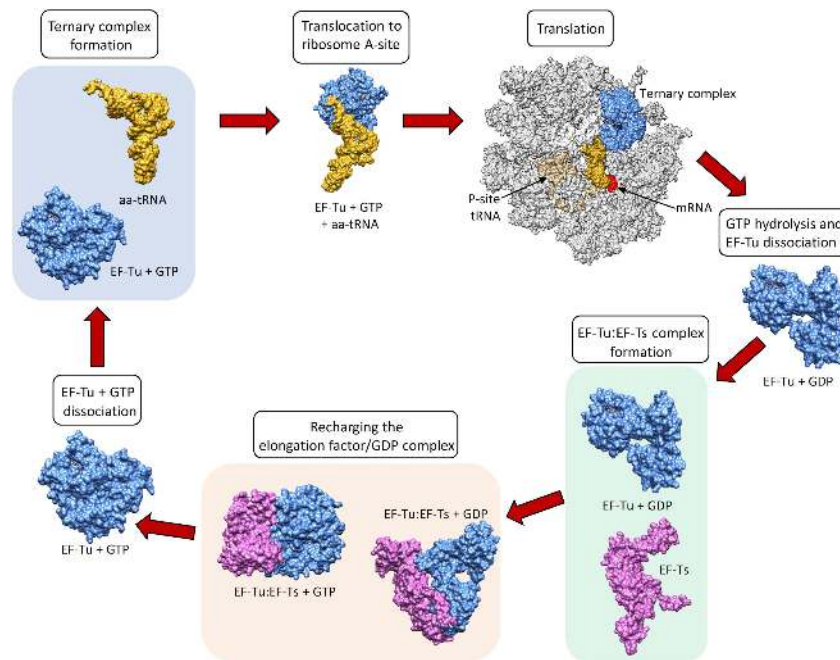


Figure [Math Processing Error]: Recharging of EF-Tu is Mediated by EF-Ts. When EF-Tu is bound with GTP (shown in blue), it can bind an aa-tRNA complex (shown in yellow), causing the translocation of the aa-tRNA to the A-site of the ribosome (shown in grey). The anticodon of the tRNA is positioned in the A-site, such that it can form H-bonds with the codon of the mRNA (shown in red). Binding of the cognate aa-tRNA in the A-site causes GTP hydrolysis by EF-Tu and subsequent dissociation of EF-Tu (GDP) from the ribosome. Following the ribosomal dissociation, EF-Tu(GDP) binds with a molecule of EF-Ts (shown in purple). Binding of EF-Ts causes a conformational change in EF-Tu, resulting in the release of GDP. A molecule of GTP is then incorporated into EF-Tu while it is complexed with EF-Ts. This causes the dissociation of EF-Ts from the complex, and EF-Tu is recharged. [Harvey, K.L., et. al. \(2019\) Front. Microbiol. 10:2351](#)

Dissociation of EF-Tu from the ribosome allows the tRNA to move into the peptidyl transferase center (A-site) on the large subunit. At the core of ribosomal translation is the catalysis of peptide bond formation, as shown in Figure [Math Processing Error]. Current reaction models suggest a substrate-assisted mechanism. Simulations indicate that the transition state forms due to extensive hydrogen bonding with water molecules and the surrounding rRNA bases, and that C-O bond cleavage occurs after C-N bond formation. Peptide bond formation results in the transition of the amino acid docked on the P-site tRNA to the nascent growing peptide that is now held on the tRNA in the A-site. Note that this mechanism causes the nascent growing peptide to always grow in the N-to-C direction.

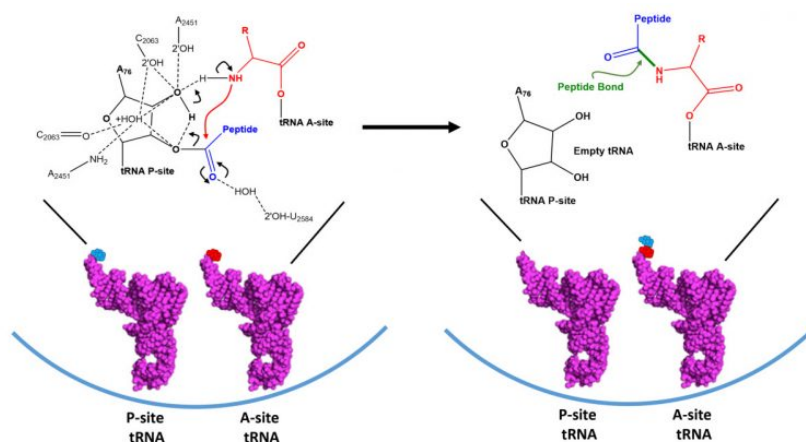


Figure [Math Processing Error]: Proposed Mechanism of Peptide Bond Formation in the Ribosome. (Left) tRNA molecules docked in the P- and A-sites are oriented such that the two amino acids move near. Hydrogen bond formation with the ribosomal rRNA and surrounding water molecules stabilizes the formation of the transition state, enabling the amine nitrogen from the A-site amino acid to attack the carbonyl carbon of the P-site amino acid. The formation of the peptide bond occurs before the C-O bond breakage, resulting in a transient oxyanion intermediate. (Right) The products of the peptidyl transferase reaction include the nascent growing peptide attached to the A-site tRNA and an empty tRNA in the P-site. Figure modified from [Awchen](#)

Once the peptide bond is formed, the ribosome needs to translocate down the mRNA to make the next mRNA codon available within the A-site. This also requires the shifting of the tRNA molecules, such that the tRNA in the A-site, which is now tethered to the nascent peptide, shifts to the P-site. The P-site tRNA (which is now empty) shifts to the E-site, and if there was an empty tRNA in the E-site, it will shift to exit the ribosome. Shifting the tRNAs and mRNAs within the ribosome core requires the action of the EF-G elongation factor, as shown in Figure [Math Processing Error].

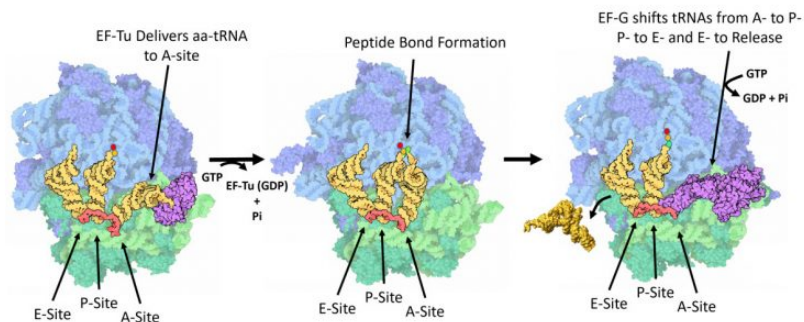


Figure [Math Processing Error]: One Cycle of Elongation. Figure modified from: [Goodsell, D. \(2010\) Molecule of the Month](#) (Left): During one round of amino acid elongation on a nascent peptide, the EF-Tu protein binds with the cognate aa-tRNA molecule and shuttles it to the A-site of the ribosome. GTP hydrolysis by EF-Tu leads to the hybridization of the anticodon of the tRNA with the codon of the mRNA and causes the dissociation of the EF-Tu (GDP-bound) from the ribosome. (Center): Following the dissociation of EF-Tu, the peptide bond is formed, leading to the transfer of the nascent peptide from the tRNA in the P-site to the tRNA in the A-site. (Right): Peptide bond formation leads to a conformational change in the ribosome that allows the binding of EF-G (GTP-bound) near the A-site of the ribosome. Rapid hydrolysis of GTP by EF-G causes a large conformational shift in the protein that twists the large subunit of the ribosome and shifts the bound tRNAs from the A- to the P-site; from the P- to the E-site; or from the E-site to exiting the ribosome.

Figure [Math Processing Error] shows an [interactive iCn3D model](#) of the eukaryotic 80S ribosome, bound to mRNA and tRNAs (6GX3). (Very long load time)

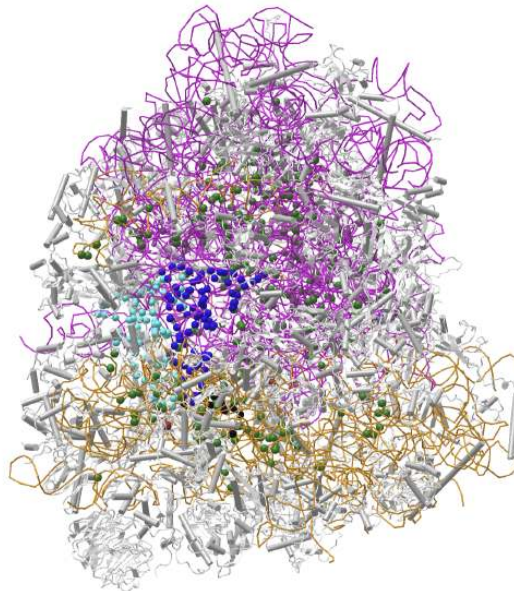


Figure [Math Processing Error]: eukaryotic 80S ribosome with bound mRNA and tRNAs (6GX3) . (Copyright; author via source). Click the image for a popup or use this external link: <https://structure.ncbi.nlm.nih.gov/1...Ak7ZvtRxaoAQp9> (Very long load time)

color coding as follows:

- gray: protein tube
- coiled coils: RNA trace
- black spheres: mRNA

- dark blue spheres: ap/P-site tRNA
- cyan spheres: pe/E-site-tRNA
- green spheres: Mg^{2+}

EF-G is a GTP hydrolase protein that binds to the A-site of the ribosome. The EF-G protein exhibits high flexibility, enabling it to act as a hinge. Folding of EF-G is dependent on GTP hydrolysis. Thus, when binding to the ribosome, the rapid hydrolysis of GTP acts as a power stroke, causing the EF-G protein to fold and inducing a conformational shift in the ribosome that enables the translocation of tRNA residues and mRNA. Translocation of tRNAs is accompanied by large-scale collective motions of the ribosome: relative rotation of ribosomal subunits and L1-stalk motion, as shown in Figure [Math Processing Error]. The L1 stalk, which is a flexible part of the large subunit, is in contact and moves along with the tRNA from the P to the E site. Once in the EF-G-GDP form, the factor quickly dissociates from the ribosome, opening up the A-site for the recruitment of the next aa-tRNA molecule. The elongation cycle will continue to be repeated until a termination codon is reached.

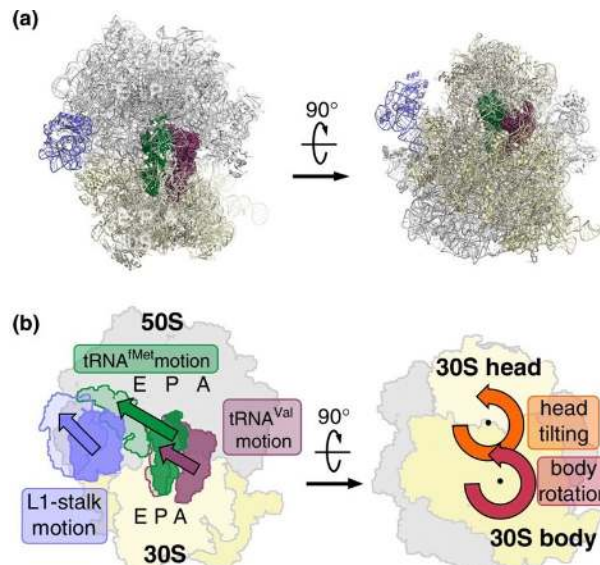


Figure [Math Processing Error]: Large-scale Motion of the Large Subunit of the Ribosome During Translocation. (a) Pre-translocation structure of the ribosome with tRNAs in A and P sites (green, brown). The L1 stalk of the large subunit is shown in purple. (b) Motions accompanying tRNA translocation. [Bock, L.V., Kolár, M.H., Grubmüller, H. \(2018\) Cur. Op. Struc. Bio. 49:27-35.](#)

26.2.5: Eukaryotic Elongation

The elongation phase in eukaryotic translation is similar to that in prokaryotic translation. Essentially, the mRNA is decoded by the ribosome in a process that requires the selection of each aminoacyl-transfer RNA (aa-tRNA), which is dictated by the mRNA codon in the ribosome acceptor (A) site, peptide bond formation and movement of both tRNAs and the mRNA through the ribosome, as shown in Figure [Math Processing Error]. A new amino acid is incorporated into a nascent peptide at a rate of approximately one every sixth of a second. The first step of this process requires guanosine triphosphate (GTP)-bound eukaryotic elongation factor 1A (eEF1 α) to recruit an aa-tRNA to the aminoacyl (A) site, which has an anticodon loop cognate to the codon sequence of the mRNA. The anticodon of this sampling tRNA does not initially base-pair with the A-site codon. Instead, the tRNA dynamically remodels to generate a codon-anticodon helix, which stabilizes the binding of the tRNA-eEF1 α -GTP complex to the ribosome A site. This helical structure is energetically favorable for cognate or correct pairing, and so discriminates between the non-cognate or unpaired and single mismatched or near-cognate species. This is important for the accuracy of decoding since it provides a mechanism to reject a non-cognate tRNA that carries an inappropriate amino acid. The pairing of the tRNA and codon induces GTP hydrolysis by eEF1 α , which is then evicted from the A site. In parallel with this process, the ribosome undergoes a conformational change that stimulates contact between the 3' end of the aa-tRNA in the A site and the tRNA carrying the polypeptide chain in the peptidyl (P) site. The shift in position of the two tRNAs [A to the P site and P to the exit (E) site] results in ribosome-catalyzed peptide bond formation and the transfer of the polypeptide to the aa-tRNA, thus extending the polypeptide by one amino acid. The second stage of the elongation cycle requires a GTPase, eukaryotic elongation factor 2 (eEF2), which enters the A-site and, through the hydrolysis of GTP, induces a conformational change in the ribosome. This stimulates ribosome translocation to allow the next aa-tRNA to enter the A-site, thus starting a new cycle of elongation.

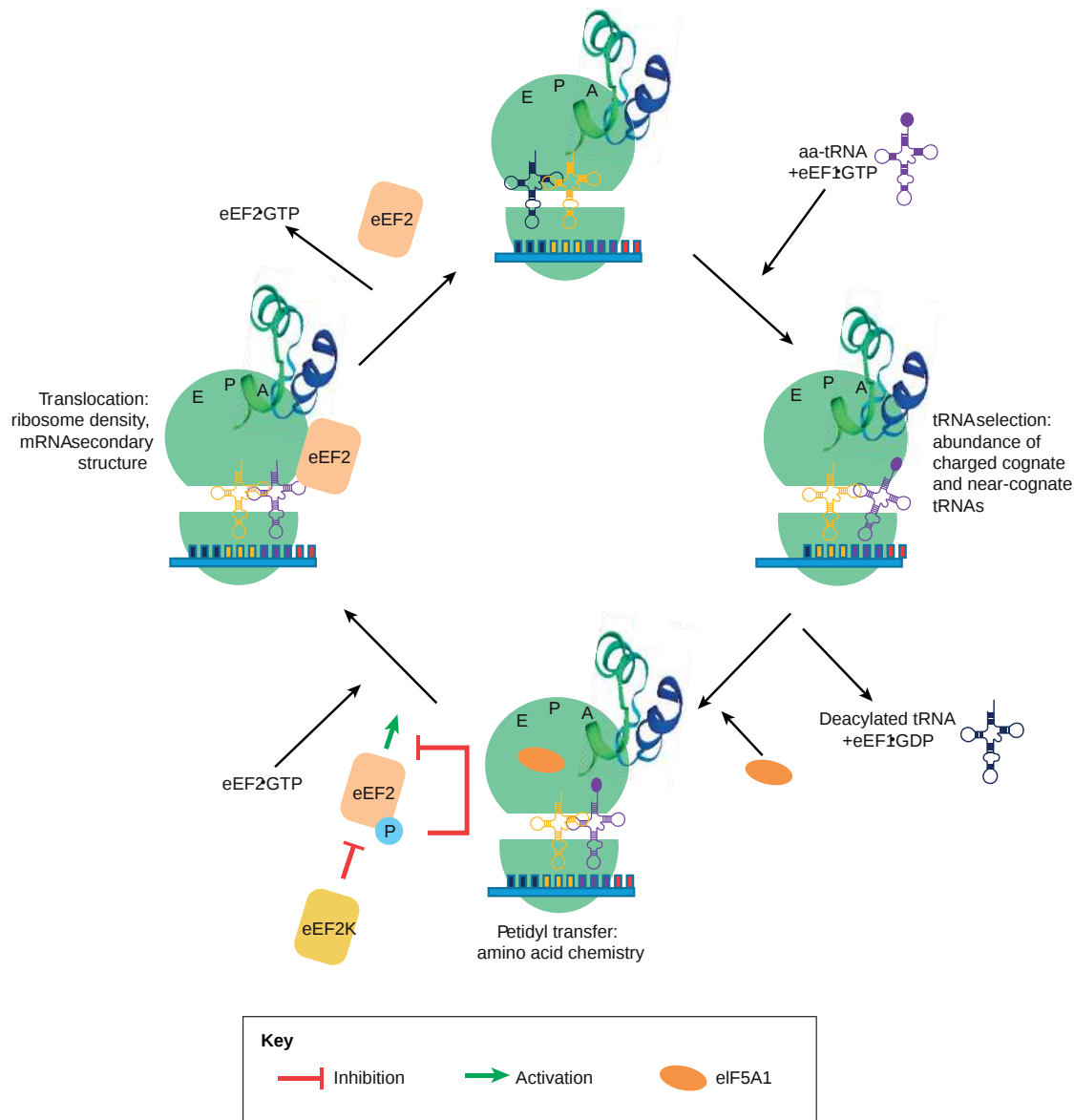


Figure [Math Processing Error]: Eukaryotic Translation Elongation Phase. [Knight, J.R.P., et. al. \(2020\) Disease Models & Mechanisms 13, dmm043208.](#)

This schematic represents the four basic steps of eukaryotic translation elongation. The ribosome contains three tRNA-binding sites: the aminoacyl (A), peptidyl (P), and exit (E) sites. In the first step of peptide elongation, the tRNA, which is in a complex with eIF1 and GTP and contains the cognate anticodon to the mRNA coding sequence, enters the A site. Recognition of the tRNA leads to the hydrolysis of GTP and eviction of eIF1 from the A site. In parallel, the deacylated tRNA in the E site is ejected. The A site and the P site tRNAs interact, allowing ribosome-catalyzed peptide bond formation to occur. This involves the transfer of the polypeptide to the aa-tRNA, thus extending the nascent polypeptide by one amino acid. eIF5A allosterically assists in the formation of certain peptide bonds, e.g., proline-proline. eEF2 then enters the A site and, through the hydrolysis of GTP, induces a change in the ribosome conformation and stimulates translocation. The ribosome is then in a correct conformation to accept the next aa-tRNA and commence another cycle of elongation.

26.2.5.1: The Ribosome as a Ribozyme

Protein synthesis from a mRNA template occurs on a ribosome, a nanomachine composed of proteins and ribosomal RNAs (rRNA). Peptide bond formation occurs when another tRNA-amino acid molecule binds to an adjacent codon on mRNA. The

tRNA has a cloverleaf tertiary structure with some intra-strand H-bonded secondary structure. The last three nucleotides at the 3' end of the tRNA are CpCpA. The amino acid is esterified to the terminal 3'OH of the terminal A by a protein enzyme, aminoacyl-tRNA synthetase.

Covalent amide bond formation between the second amino acid and the first, forming a dipeptide, occurs at the peptidyl transferase center, located on the larger ribosomal subunit (50S in bacteria and 60S in eukaryotes). The ribosome ratchets down the mRNA so the dipeptide-tRNA is now at the P or Peptide site, awaiting a new tRNA-amino acid at the A or Amino site. Figure [Math Processing Error]: The figure below shows a schematic of the ribosome with bound mRNA on the 30S subunit and tRNAs covalently attached to an amino acid (or the growing peptide) at the A and P sites, respectively.

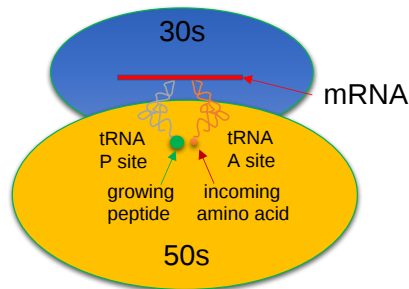


Figure [Math Processing Error]: Ribosome with bound mRNA on the 30S subunit and tRNAs covalently attached to amino acid (or the growing peptide) at the A and P site

A likely mechanism (derived from crystal structures with bound substrates and transition state analogs) for the formation of the amide bond between a growing peptide on the P-site tRNA and the amino acid on the A-site tRNA is shown in Figure [Math Processing Error]. Catalysis does not involve any of the ribosomal proteins (not shown) since none is close enough to the peptidyl transferase center to provide amino acids that could participate in general acid/base catalysis, for example. Hence, the rRNA must act as the enzyme (i.e., it is a ribozyme). Initially, it was thought that a proximal adenosine with a perturbed pKa could, at physiological pH, be protonated/deprotonated and hence act as a general acid/base in the reaction. However, none was found. The most likely mechanism to stabilize the oxyanion transition state at the electrophilic carbon attack site is precisely located water, which is positioned at the oxyanion hole by H-bonds to uracil 2584 on the rRNA. The cleavage mechanism involves the concerted proton shuffle shown below. In this mechanism, the substrate (peptide-tRNA) facilitates its own cleavage, as the 2'OH is positioned to initiate the proton shuttle mechanism. (A similar mechanism might occur to facilitate hydrolysis of the fully elongated protein from the P-site tRNA.) Of course, all of this requires perfect positioning of the substrates and isn't that what enzymes do best? The primary mechanisms for catalyzing peptide bond formation by the ribosome, acting as a ribozyme, are intramolecular catalysis and transition state stabilization facilitated by an appropriately positioned water molecule.

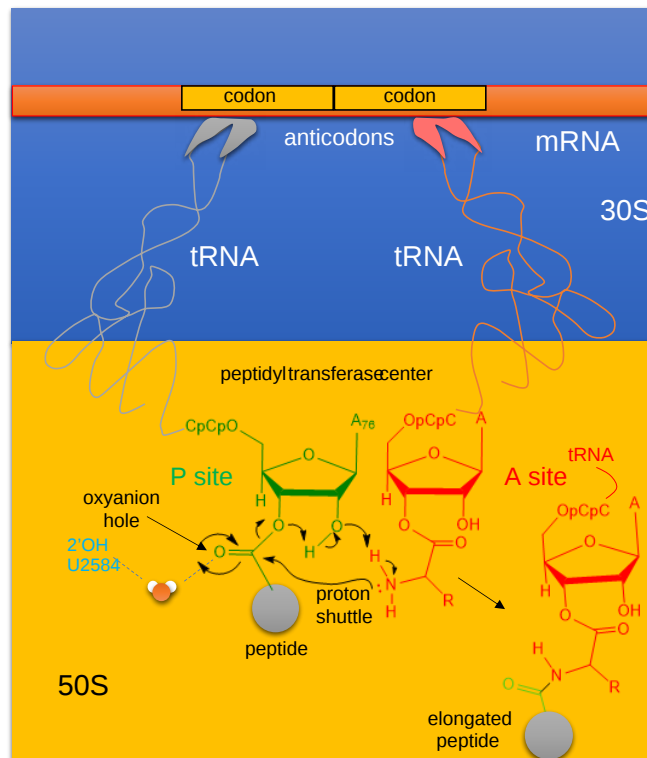


Figure [Math Processing Error]: formation of the amide bond between a growing peptide on the P-site tRNA and the amino acid on the A-site tRNA.

26.2.6: Translation Termination

26.2.6.1: Prokaryotic Termination

Termination of bacterial protein synthesis occurs when a stop codon is presented in the ribosomal A-site and is recognized by a class I release factor, RF1 or RF2. These release factors (RFs) have different but overlapping specificities, where RF1 recognizes UAA and UAG, and RF2 recognizes UAA and UGA, with strong discrimination against sense codons. The RFs are multi-domain proteins, where binding and stop codon recognition by domain 2 at the decoding site causes the universally conserved GGQ motif of domain 3 to insert into the A-site of the PTC, some 80 Å away from the decoding site. This event triggers hydrolysis of the peptidyl-tRNA bond in the P-site of the PTC, and the nascent peptide chain can then be released via the ribosomal exit tunnel, as shown in Figure [Math Processing Error]. After peptide release, RF1 and RF2 dissociate from the post-termination complex. The dissociation is accelerated by a class II release factor called RF3, which functions as a translational GTPase that binds and hydrolyzes GTP during termination.

While RF3 increases the efficiency of peptide hydrolysis, it is not an essential protein for the process. In gene knockout studies, RF3 is dispensable for the growth of *Escherichia coli*, and its expression is not conserved in all bacterial lineages. For example, RF3 is not present in the thermophilic model organisms of the *Thermus* and *Thermatoga* genera, as well as in infectious *Chlamydiales* and *Spirochaetae*. This means that both RF1 and RF2 are capable of performing a complete round of termination independently of RF3 or that other GTPases from the elongation or initiation phases of translation can compensate for the action of RF3.

The release factors RF1 and RF2 acquire an open conformation (Figure [Math Processing Error]) on the 70S ribosome, which is distinctly different from the closed conformation observed in crystal structures of free RFs. The conformational equilibrium of the free RFs in solution indicates that the open conformation dominates at approximately 80%.

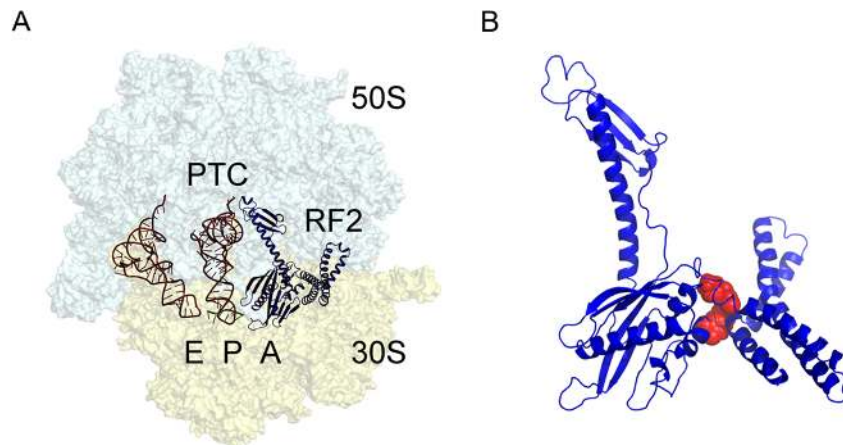


Figure [Math Processing Error]: The bacterial 70S ribosome termination complex with RF2. (A) View of the ribosome termination complex with E- and P-site tRNAs (brown), mRNA (green), and RF2 (dark blue). (B) Close-up view of the hinge region of RF2 between domains 1 and 4 used for virtual screening, where the putative binding region is indicated by a docked ligand (red). [Ge, X., et. al. \(2019\) Scientific Reports 9:15424.](#)

During peptide hydrolysis, the RF factors cause rotational and conformational changes within the ribosome that allow the binding of a ribosome recycling factor (RRF) and the EF-G GTPase, which leads to the dissociation of the large subunit from the small subunit and the release of the mRNA, as shown in Figure [Math Processing Error].

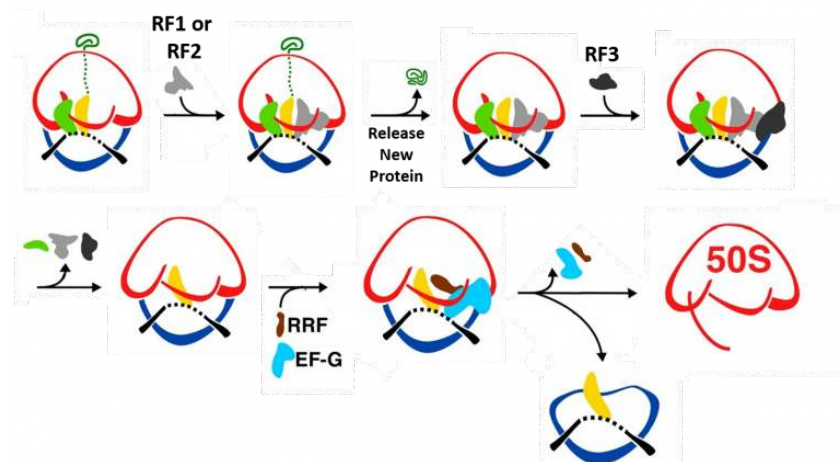


Figure [Math Processing Error]: Termination of Translation. Subsequent binding of the ribosome recycling factor, RRF, and EF-G causes the dissociation of the large and small ribosomal subunits and the release of the mRNA. Figure modified from: [Bock, L.V., Kolár, M.H., Grubmüller, H. \(2018\) Cur. Op. Struc. Bio. 49:27-35.](#)

When a stop codon enters the A-site of the ribosome RF1 or RF2 enter the A-site and bind with the mRNA. This leads to the hydrolysis of the protein and release through the exit tunnel. Binding of RF3 and GTP hydrolysis causes the dissociation of the RF factors and a conformational change in the ribosome structure.

26.2.7: Eukaryotic Termination

In eukaryotes and archaea, on the other hand, a single, omnipotent RF reads all three stop codons. Although the mechanism of translation termination is the same, there is neither sequence nor structural homology between the bacterial RFs and the eukaryotic eRF1, apart from the universally conserved GGQ motif, which is required for peptide hydrolysis from the tRNA. The eRF3 GTPase coordinates the release of eRF1 following the hydrolysis of its GTP. In Archaea, there is no eRF3 homolog, instead the aEF1A protein mediates this function. The process of eukaryotic ribosomal disassembly and recycling is currently not well understood, but it appears to involve an ABC-type ATPase called ABCE1. Mitochondria have independent release factors (RFs) that can recognize standard and non-standard stop codons, and are more homologous with bacterial systems of ribosomal recycling and disassembly.

26.2.8: Summary of Translation

An overall summary of prokaryotic translation is given in Figure [\[Math Processing Error\]](#).



Figure [\[Math Processing Error\]](#): Summary of Prokaryotic Translation. [Bock, L.V., Kolár, M.H., Grubmüller, H. \(2018\) Cur. Op. Struc. Bio. 49:27-35.](#)

Left panel: Structure of the bacterial ribosome in complex with EF-Tu (PDB 5AFI).

Right Panel: Scheme of the bacterial translation cycle. 30S: small subunit; 50S: large subunit; IF1, IF2, IF3: initiation factors; fMet-tRNA: N-formylmethionine tRNA; aa-tRNA: aminoacyl tRNA; EF-Tu, EF-G: elongation factors; RF1, RF2, RF3: release factors; RRF: ribosome recycling factor; green trace: nascent protein. The question mark stands for a stop codon recognition.

26.2.9: Summary

This chapter provides a detailed exploration of translation initiation in prokaryotes, outlining the molecular components and stepwise events that set the stage for accurate protein synthesis. Key highlights include:

Overview of Ribosomal Subunits and Initiation Components

- **Ribosomal Architecture:**

The small (30S) and large (50S) ribosomal subunits perform distinct roles—30S decodes the mRNA by matching codons with tRNA anticodons, while 50S houses the peptidyl transferase center for peptide bond formation. These subunits join at the start codon to form a functional 70S ribosome.

- **mRNA Translation Initiation Region (TIR):**

The TIR contains the start codon (most frequently AUG) and, in many bacterial mRNAs, an upstream Shine–Dalgarno (SD) sequence. The SD sequence, which is complementary to the 3' end of the 16S rRNA, helps dock the mRNA to the 30S subunit and positions the start codon in the P-site for accurate initiation.

- **Initiator tRNA^{fMet}:**

The initiator tRNA is uniquely charged with methionine that is enzymatically modified by formylation (fMet). This modification is critical because it directs the tRNA to the P-site and prevents its interaction with elongation factors during initiation.

Role of Initiation Factors and Stepwise Assembly

- **Initiation Factors (IF1, IF2, IF3):**

These proteins bind to specific sites on the 30S subunit to enhance both the speed and accuracy of complex formation:

- IF3 helps maintain fidelity by preventing premature association with the 50S subunit.
- IF2, a GTPase, facilitates the recruitment of fMet-tRNA^{fMet} and its proper positioning in the P-site.
- IF1 binds in the A-site, inducing conformational changes that stabilize the complex.

- **Sequential Assembly Process:**

The initiation process proceeds through a series of steps:

1. The vacant 30S subunit binds IF3 and IF2.

2. IF1 subsequently associates, triggering a conformational change.
3. fMet-tRNA^{fMet} is recruited—its high-affinity binding to the P-site is critical.
4. mRNA binds to the 30S subunit; its TIR and SD sequence help align the start codon.
5. After formation of a stable 30S pre-initiation complex, the 50S subunit docks.
6. IF2 hydrolyzes GTP upon 50S joining, prompting the dissociation of all initiation factors.
7. The resulting 70S initiation complex is now primed for elongation.

Comparison with Eukaryotic Initiation

- While the prokaryotic system is streamlined with only three initiation factors, eukaryotic initiation is more complex, involving at least 11 eukaryotic initiation factors (eIFs) and additional auxiliary proteins.
- Eukaryotic mRNAs utilize a 5' cap structure and a Kozak consensus sequence to guide ribosome binding, and the small ribosomal subunit (40S) forms a 43S pre-initiation complex that scans the mRNA for the start codon.
- Despite these differences, both systems share the fundamental goal of accurately positioning the initiator tRNA at the start codon.

This chapter integrates the molecular choreography of prokaryotic translation initiation—from the critical interactions of ribosomal subunits, initiator tRNA^{fMet}, and initiation factors, to the precise alignment of the mRNA via the TIR and SD sequence. Understanding these steps provides essential insight into how bacteria ensure rapid and accurate protein synthesis, setting the stage for subsequent elongation and termination. This detailed mechanistic perspective is foundational for appreciating the evolutionary and functional differences between prokaryotic and eukaryotic translation systems.

26.2.10: References

This chapter was remixed and adapted from the following resources under [creative commons licensing](#):

1. Wikipedia contributors. (2020, June 27). Wobble base pair. In *Wikipedia, The Free Encyclopedia*. Retrieved 16:47, August 12, 2020, from https://en.Wikipedia.org/w/index.php?title=Wobble_base_pair&oldid=964760055
2. Lorenz, C., Lünse, C., and Mörl, M. (2017) tRNA modifications: Impact on structure and thermal adaptation. *Biomolecules* 7(2)35. Available at: <https://www.mdpi.com/2218-273X/7/2/35/html>
3. Pan, T. (2018) Modifications and functional genomics of human transfer RNA. *Cell Research* 28:395-404. Available at: <https://www.nature.com/articles/s41422-018-0013-y#Sec3>
4. Bednářová, A., Hanna, M., Durham, I., Van Cleave, T., England, A., Chaudhuri, A., and Krishnan, N. (2017) Lost in translation: Defects in transfer RNA modifications and neurological disorders. *Front. Mol. Neurosci.* 10:135. Available at: https://www.researchgate.net/publication/316440980_Lost_in_Translation_Defects_in_Transfer_RNA_Modifications_and_Neurological_Disorders
5. Wikipedia contributors. (2020, May 17). Transfer RNA. In *Wikipedia, The Free Encyclopedia*. Retrieved 23:03, August 15, 2020, from https://en.Wikipedia.org/w/index.php?title=Transfer_RNA&oldid=957227343
6. Gomez, M.A.R., and Ibba M. (2020) Aminoacyl-tRNA Synthetases. *RNA*, doi: 10.1261/rna.071720.119 Available at: <https://rnajournal.cshlp.org/content/early/2020/04/17/rna.071720.119.abstract>
7. Li, R., Macnamara, L.M., Leuchter, J.D., Alexander, R.W., and Cho, S.S. (2015) MD Simulations of tRNA and Aminoacyl-tRNA Synthetases: Dynamics, Folding, Binding, and Allostery. *Int. J. Mol. Sci.* 16(7):15872-15902. Available at: <https://www.mdpi.com/1422-0067/16/7/15872>
8. Wikipedia contributors. (2020, August 15). Ribosome. In *Wikipedia, The Free Encyclopedia*. Retrieved 05:22, August 16, 2020, from <https://en.Wikipedia.org/w/index.php?title=Ribosome&oldid=973144885>
9. Kater, L., Thoms, M., Barrio-Garcia, C., Cheng, J., Ismail, S., Ahmed, Y.L., Bange, G., Kressler, D., Berninghausen, O., Sinning, I., Hurt, E., and Beckmann, R. (2017) Visualizing the assembly pathway of nucleolar Pre-60S Ribosomes. *Cell* 171(7):1599-1610. Available at: <https://www.sciencedirect.com/science/article/pii/S0092867417314290>
10. Bock, L.V., Kolár, M.H., Grubmüller, H. (2018) Molecular simulations of the ribosome and associated translation factors. *Cur. Op. Struc. Bio.* 49:27-35. Available at: <https://www.sciencedirect.com/science/article/pii/S0959440X1730132X>
11. Doris, S.M., Smith, D.R., Beamesderfer, J.N., Raphael, B.J., Nathanson, J.A., and Gerbi, S.A. (2015) Universal and domain-specific sequences in 23S-28S ribosomal RNA identified by computational phylogenetics. *RNA* 21:1719-1730. Available at: https://www.researchgate.net/publication/281141702_Universal_and_domain-specific_sequences_in_23S-28S_ribosomal_RNA_identified_by_computational_phylogenetics
12. Aleksashin, M.A., Leppik, M., Hochenberry, A.J., Klepacki, D., Vázquez-Laslop, N., Jewett, M.C., Remme, J., and Mankin A.S. (2019) Assembly and functionality of the ribosome with tethered subunits. *Nature Communications* 10:930. Available at:

<https://www.nature.com/articles/s41467-019-08892-w#rightslink>

13. Wikipedia contributors. (2020, July 3). Formylation. In *Wikipedia, The Free Encyclopedia*. Retrieved 22:03, August 16, 2020, from <https://en.Wikipedia.org/w/index.php?title=Formylation&oldid=965827476>
14. Gualerzi, C.O., and Pon C.L. (2015) Initiation of mRNA translation in bacteria: structural and dynamic aspects. *Cell Mol Life Sci.* 72:4341-4367. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4611024/>
15. Wikipedia contributors. (2020, June 14). Kozak consensus sequence. In *Wikipedia, The Free Encyclopedia*. Retrieved 06:03, August 18, 2020, from https://en.Wikipedia.org/w/index.php?title=Kozak_consensus_sequence&oldid=962444929
16. Knight, J.R.P., Garland, G., Pöyry, T., Mead, E. Vlahov, N., Sfakianos, A., Frosso, S., De-Lima-Hedayioglu, F., Mallucci, G.R., von der Haar, T., Smales, C.M., Sansom, O.J., and Willis, A.E. (2020) Control of translation elongation in health and disease. *Dis. Mod. and Mech.* 13: dmm043208. Available at: <https://dmm.biologists.org/content/13/3/dmm043208>
17. Adio, S., Sharma, H., Senyushkina, T., Karki, P., Maracci, C., Wohlgemuth, I., Holtkamp, W., Peske, R., and Rodina, M.V. (2018) Dynamics of ribosomes and release factors during translation termination in *E. coli*. *eLife* 7:e34253. Available at: <https://elifesciences.org/articles/34252>
18. Ge, X., Oliveira, A., Hjort, K., Bergfors, T., Guitierrez-de-Teran, H., Andersson, D.I., Sanyal, S., and Åqvist, J. (2019) Inhibition of translation termination by small molecules targeting ribosomal release factors. *Scientific Reports* 9: 15424. Available at: <https://www.nature.com/articles/s41598-019-51977-1#rightslink>
19. Svidritskiy, E., Demo G., Loveland A.B., Xu, C., and Korostelev, A.A. (2019) Extensive ribosome and RF2 rearrangements during translation termination. *eLife* 8:e46850. Available at: <https://elifesciences.org/articles/46850>
20. Sauert, M., Temmel, H., and Moll, I. (2015) Heterogeneity of the translational machinery: Variations on a common theme. *Biochimie* 114:39-47. Available at: <https://www.sciencedirect.com/science/article/pii/S0300908414003952>

This page titled [26.2: Protein Synthesis](#) is shared under a [not declared](#) license and was authored, remixed, and/or curated by [Henry Jakubowski](#) and [Patricia Flatt](#).

- [11.2: Diffusion Across a Membrane - Passive and Facilitated Diffusion](#) by [Henry Jakubowski](#) and [Patricia Flatt](#) has no license indicated.
- [Current page](#) by [Henry Jakubowski](#) and [Patricia Flatt](#) has no license indicated.

26.3: Translational Regulation and Protein Degradation

[Search Fundamentals of Biochemistry](#)



Learning Goals (ChatGPT o3-mini)

- **Understand Ribosome Heterogeneity and Its Functional Implications:**
 - Describe how variations in ribosomal protein (RP) composition and rRNA sequences (including differential post-translational modifications) contribute to ribosome heterogeneity.
 - Explain the “ribosome filter hypothesis” and how specialized ribosomes can selectively translate subsets of mRNAs.
 - Analyze examples of rRNA gene copy variation and chemical modifications (e.g., 2'-OH methylation and pseudouridylation) as sources of functional diversity in ribosomes.
- **Evaluate the Impact of mRNA Sequence and Secondary Structure on Translation Efficiency:**
 - Identify key features in the mRNA (GC content, codon usage, codon pairs, and 5'UTR secondary structure) that influence the translational output.
 - Explain how interactions with small RNAs (such as miRNAs) and protein factors (e.g., Argonautes) can lead to mRNA degradation or repression of translation.
 - Discuss how the primary sequence and structural elements of mRNA serve as regulatory signals in translation.
- **Examine the Influence of the Nascent Peptide on Ribosome Function:**
 - Describe how the nascent peptide exits through the ribosomal tunnel and the role of the tunnel in modulating translation.
 - Analyze how specific peptide sequences and secondary structures can induce ribosome stalling and the role of specialized factors (e.g., EF-P) in relieving stalling.
 - Assess how interactions between the nascent peptide and the exit tunnel can be targeted by antibiotics (e.g., macrolides) to modulate translation.
- **Understand the Structure and Function of the Proteasome in Protein Degradation:**
 - Describe the architecture of the 26S proteasome, including the 20S core particle (α - and β -subunits) and the regulatory 19S cap.
 - Explain the mechanism by which proteasome activators open the gate of the 20S core to allow substrate entry and subsequent proteolysis.
 - Discuss the roles of the catalytic β -subunits ($\beta 1$, $\beta 2$, and $\beta 5$) in substrate cleavage and how their specificities (caspase-, trypsin-, and chymotrypsin-like activities) determine proteolytic outcomes.
- **Analyze the Process and Regulation of Ubiquitination:**
 - Outline the enzymatic cascade of ubiquitin activation, conjugation, and ligation involving E1, E2, and E3 enzymes.
 - Differentiate between mono-ubiquitination, multi-mono-ubiquitination, and polyubiquitination, and describe how these modifications target proteins for degradation or alter their cellular functions.
 - Discuss how deubiquitinating enzymes (DUBs) recycle ubiquitin and regulate proteasome-mediated degradation.
- **Integrate Translational Regulation with Protein Homeostasis:**
 - Connect how modulation of translation efficiency (via ribosome heterogeneity, mRNA structure, and nascent peptide interactions) interfaces with the degradation machinery (proteasome and ubiquitination system) to maintain cellular protein homeostasis.
 - Evaluate the cellular significance of having specialized regulatory mechanisms at both the translational and post-translational levels in response to environmental cues and developmental signals.

These learning goals are designed to provide a comprehensive framework for understanding how translation is finely tuned by a range of molecular mechanisms—from the diversity of ribosomal components and mRNA signals to the regulation of protein turnover by the ubiquitin–proteasome system.

26.3.1: Regulation of Translation

26.3.1.1: Heterogeneity of Ribosome Structure

Over the years, numerous studies conducted in eukaryotes have presented evidence that ribosomes can vary in their protein and rRNA complements between different cell types and developmental states. These observations culminated in the postulation of the "*ribosome filter hypothesis*" by Mauro and Edelman in 2002. The authors propose that the ribosome composition functions as a determinant of translation. Depending on the RPs and rRNA sequences represented in the respective ribosome, the complex acts like a filter that selects for specific mRNAs and hence modulates translation, as shown in Figure [Math Processing Error]. RP heterogeneity can arise from the differential expression of paralogs or homologs of RP proteins within different cell types or occur due to differential post-translational modifications of RPs, such as phosphorylation. The protein-to-rRNA ratio may also slightly vary within ribosomal composition, affecting translation efficiency and selectivity.

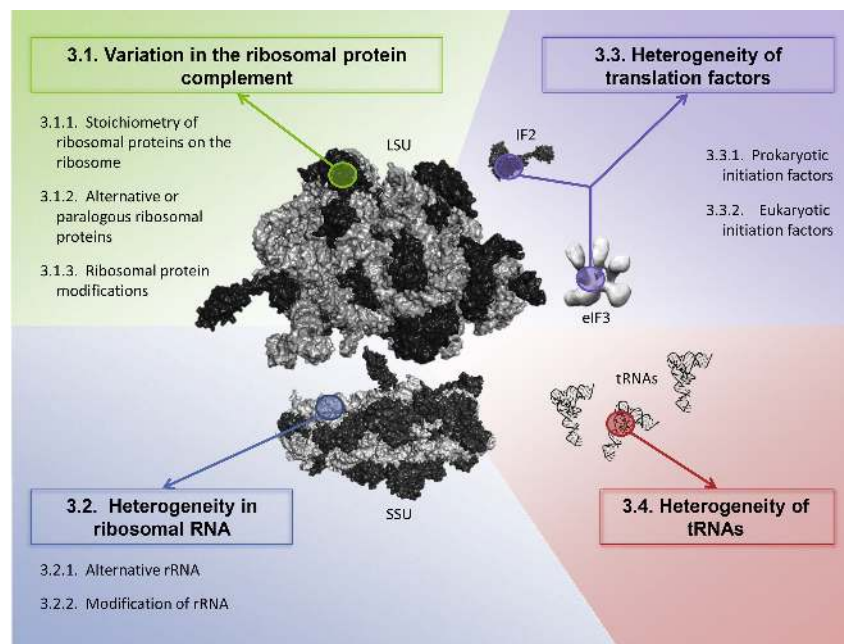


Figure [Math Processing Error]: Components of the translation machinery that have the potential to contribute to functional heterogeneity. Structures were visualized with Polyview 3D software using the following maps: *E. coli* large ribosomal subunit (LSU; pdb accession code: 3d5a, rRNA in light gray, proteins in dark gray); small ribosomal subunit (SSU; pdb accession code: 3d5b); *E. coli* IF2 (pdb accession code: 1g7r); human eIF3 (EMD-2166); yeast tRNA^{Phe} (pdb accession code: 6tna). [Sauert, M., Temmel, H., and Moll, I. \(2015\) Biochimie 114: 39-47.](#)

RNA genes are also present in multiple copies throughout the genomes of organisms from all domains of life. For example, the bacteria *Streptomyces coelicolor* harbor six copies of divergent large subunit (LSU) rRNA genes that constitute at least five different LSU rRNA species in a cell. These genes were shown to be differentially transcribed during the morphological development of the organism. Similarly, *B. subtilis* harbors ten *rRNA* operons, and their reduction to one copy increased the doubling time as well as the sporulation frequency and the motility of the resulting mutant.

Modification of the rRNA also provides another avenue of ribosomal heterogeneity. Similar to tRNA, rRNA residues can be chemically modified and commonly have 2-OH methylation. The conversion of uridine to pseudouridine is also a common occurrence. In eukaryotes, the modifications are facilitated by snoRNAs, and their tissue-specific expression might be a source for ribosome specialization. In light of the increasing evidence, ribosome heterogeneity, although still far from being entirely understood, proves to be an integral mechanism for modulating and fine-tuning protein synthesis in response to environmental signals in all organisms.

26.3.1.2: Effects of Sequence and Secondary Structure in mRNA

The amount of protein produced from any given mRNA (i.e., the translational output) is influenced by multiple factors specified by the primary nucleotide sequence. These factors include GC content, codon usage, codon pairs, and secondary structure. For example, 5' UTR sequences in the mRNA may interact with small miRNAs, leading to RNA interference. miRNA interactions may also target mRNA for degradation (Figure [Math Processing Error]). This process is facilitated by protein chaperones known

as argonautes. This antisense-based process involves steps that first process the miRNA so that it can base-pair with a region of its target mRNAs. Once the base pairing occurs, other proteins direct the mRNA to be destroyed by nucleases. Fire and Mello were awarded the 2006 Nobel Prize in Physiology or Medicine for this discovery.

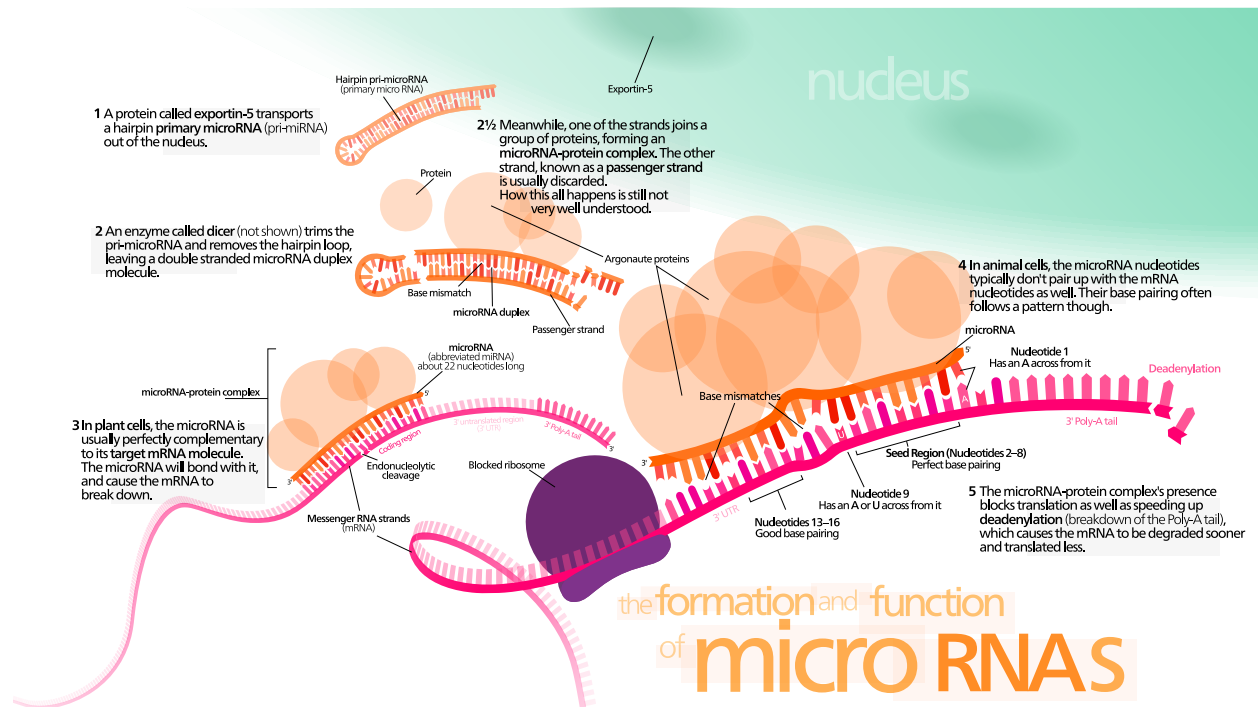


Figure [Math Processing Error]: Role of Micro RNA (miRNA) in the Inhibition of Eukaryotic mRNA Translation. [Wikimedia Commons](#)

- Step (1) shows how Exportin-5 transports a hairpin primary micro RNA (pri-miRNA) out of the nucleus and into the cytoplasm.
- Step (2) shows how Dicer (not shown) trims the pri-miRNA and removes the hairpin loop. A group of proteins, known as Argonautes, form a miRNA/protein complex.
- Steps (3,4) show how miRNA/protein complex hydrogen bonds with mRNA based on complementary sequence homology, and blocks translation.
- Step (5) shows the miRNA/protein complex binding speeds up the breakdown of the polyA tail of the mRNA, causing the mRNA to be degraded sooner.

26.3.1.3: Effects of the Nascent Peptide on Ribosome Efficiency

Since the Peptidyl Transferase Center (PTC) is buried within the large subunit, during translation, the nascent peptide chain (NC) exits through a 100 Å-long tunnel (Figure [Math Processing Error]). The exit tunnel plays an active role in protein synthesis. Certain peptide sequences specifically interact with tunnel walls and induce ribosome stalling. Furthermore, the exit tunnel is a binding site for a clinically important class of antibiotics known as the [macrolides](#).

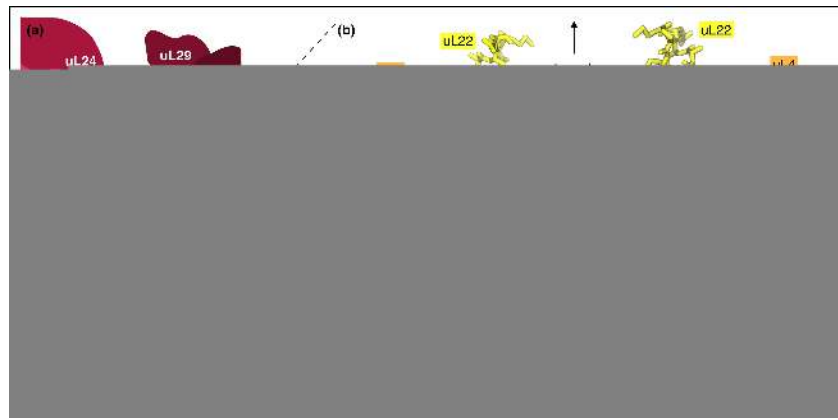


Figure *[Math Processing Error]*: The Ribosomal Exit Tunnel (a) Scheme of the ribosome exit tunnel with several proteins highlighted. NC, nascent peptide chain; ERY, erythromycin; PTC, peptidyl transferase center. (b) Context of the erythromycin (ERY, in green) binding; figure based on PDB: 5JTE. Several large subunit nucleotides are highlighted in bold red. Two proteins uL4 and uL22 form a constriction site. The nascent peptide is shown as a transparent surface structure. [Bock, L.V., Kolár, M.H., Grubmüller, H. \(2018\) Cur. Op. Struc. Bio. 49:27-35](#)

When synthesizing proteins containing proline stretches (i.e., several prolines in a row), ribosomes become stalled. Stalling is alleviated by a specialized elongation factor, EF-P, in bacteria. Recently, cryo-EM structures of a ribosome stalled by a proline stretch with and without EF-P were resolved. In simulations of the PTC region, elongation factor P (EF-P) was observed to stabilize the P-site tRNA in a conformation compatible with peptide bond formation. In contrast, in the absence of EF-P, the P-site tRNA moved away from the A-site tRNA.

The exit tunnel can accommodate 30–60 AAs, depending on the level of NC compaction. The rate of translation of about 4–22 AA per second in bacteria provides the NC with sufficient time to explore its conformational space and to start folding when still bound to the ribosome-tRNA complex.

26.3.2: Proteasome

The 26S proteasome is the central element of proteostasis regulation in eukaryotic cells. It is required for the degradation of protein factors in multiple cellular pathways, and it plays a fundamental role in cell stability. The 26S proteasome has a structural configuration that confines the proteolytic active sites in a location unreachable for native and functional proteins, thus preventing uncontrolled degradation. The proteolytically active sites are found in the interior of a barrel-shaped core particle (CP or 20S). The entrances of the tunnel, located at the distal ends of the barrel, are commonly occupied by the regulatory particle (RP or 19S), a complex protein assembly that functions as a substrate processing machine. The regulatory particle has the important role of receiving, deubiquitinating, unfolding, and translocating substrates to the CP, and it adopts different configurations depending on the activity states they exhibit. This process typically requires ATP hydrolysis. Moreover, conformationally distinct proteasomes may exhibit different subcellular distributions, depending on the functional requirements of each cell type and environmental situation. Proteasomes are distributed throughout the cell, detected in both the cytoplasm and the nucleus, and can localize to specific sites with high protein metabolism or with distinct protein degradation requirements.

The core of the proteasome consists of a symmetrical cylinder-shaped structure composed of four stacked rings, each containing seven different subunits, and is called the 20S proteasome, as shown in Figure *[Math Processing Error]*. The two outer rings are each composed of seven α -subunits (α 1- α 7 or PSMA1-7). During proteasome assembly, the α -rings serve as the backbone for the incorporation of β -subunits, followed by the dimerization of two half-proteasomes. In mature proteasomes, the α -rings regulate substrate entrance, as the α -subunits have hydrophobic loops that close the 20S barrel, preventing the random entry of substrates.

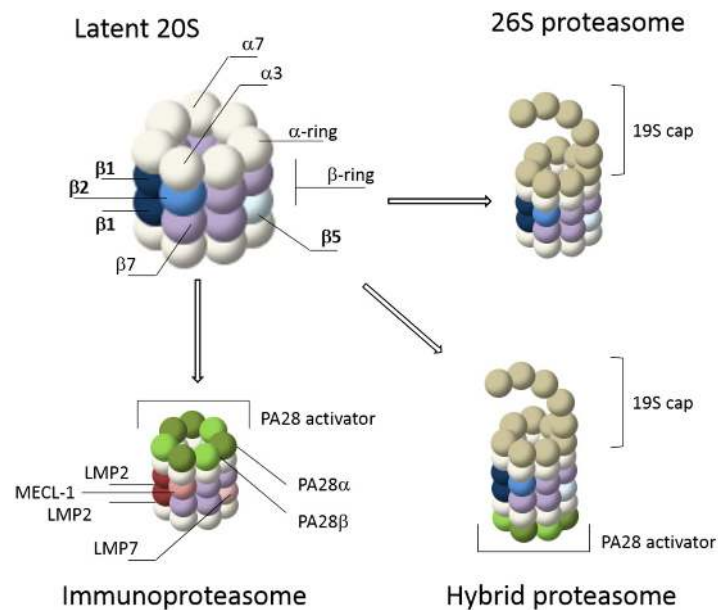


Figure [Math Processing Error]: Sabine et al., Frontiers in Molecular Biosciences, 6 (2019), <https://www.frontiersin.org/articles...olb.2019.00056>. DOI=10.3389/fmolb.2019.00056. Creative Commons Attribution License (CC BY).

The 20S core of the proteasome consists of 4 stacked rings. The outer rings contain seven α -subunits (white) while the inner rings contain seven β -subunits (purple). The catalytic subunits, $\beta 1$, $\beta 2$, and $\beta 5$, are depicted in shades of blue. Gate opening of the 20S core occurs via capping by proteasome activators such as the 19S cap or PA28. The 19S cap is the most abundant activator, and it forms the 26S proteasome together with the 20S core. Different cells have different caps. For example, interferoIFN- γ stimulation induces *de novo* formation of immunoproteasomes, which contain the immune subunits $\beta 1i$ (LMP2), $\beta 5i$ (LMP7), and $\beta 2i$ (MECL-1) (shades of red), as well as proteasome activation by PA28 $\alpha\beta$ (shades of green). Proteasomes in neural tissue are discussed below.

In general, protein entry can only be established after gate opening by proteasome activators (PAs) such as the 19S cap, after which substrates can enter the interior of the 20S core for degradation. The inner two rings of the 20S barrel consist of the subunits $\beta 1$ - $\beta 7$ (PSMB1-7). Each β -ring contains three catalytic subunits, termed $\beta 1$, $\beta 2$, and $\beta 5$. In mature 20S complexes, the pro-peptides of these catalytic β -subunits are auto-catalytically removed. Upon autocatalytic processing, the N-terminal threonine residues become exposed as the catalytically reactive residues, harboring both the nucleophile (the hydroxyl group) and the catalytic base (the N-terminal amine) involved in peptide bond cleavage. Each catalytic subunit has selectivity toward specific residues. $\beta 1$ has caspase- or peptidyl-glutamyl peptidase-like activity, preferring cleavage at the C-terminus of acidic residues. $\beta 2$ has trypsin-like activity and cleaves after basic residues, while $\beta 5$ has chymotrypsin-like activity and prefers cleavage after hydrophobic residues. Figure [Math Processing Error] shows an [interactive iCn3D model](#) of the Human 20S Proteasome (6RGQ).

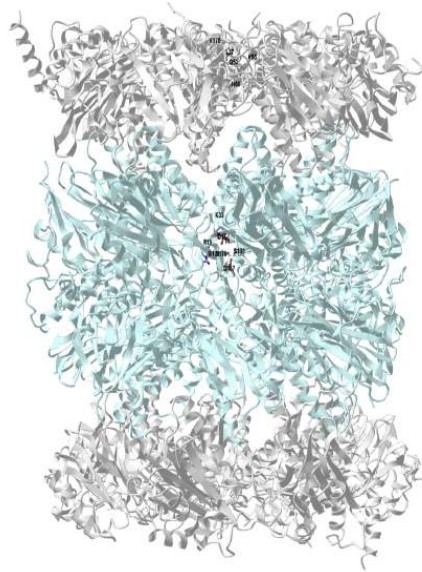


Figure *[Math Processing Error]*: Human 20S Proteasome (6RGQ). (Copyright; author via source). Click the image for a popup or use this external link: <https://structure.ncbi.nlm.nih.gov/1...9gDp6nj7Co2Ud9>

The alpha subunits are shown in light gray, and the beta subunits are in light cyan. Active site residues, as defined by databases, are shown for one of the alpha chains (L37, Q53, K55, H68, V170) and one of the beta chains (T1, D17, R19, K33, S130, D167, S170, G171).

A possible generic mechanism for the cleavage of peptide bonds by the beta chain is shown in Figure *[Math Processing Error]*. The mechanism illustrates how the newly exposed N-terminal threonine functions as both the nucleophile (via its hydroxyl group) and the catalytic base.

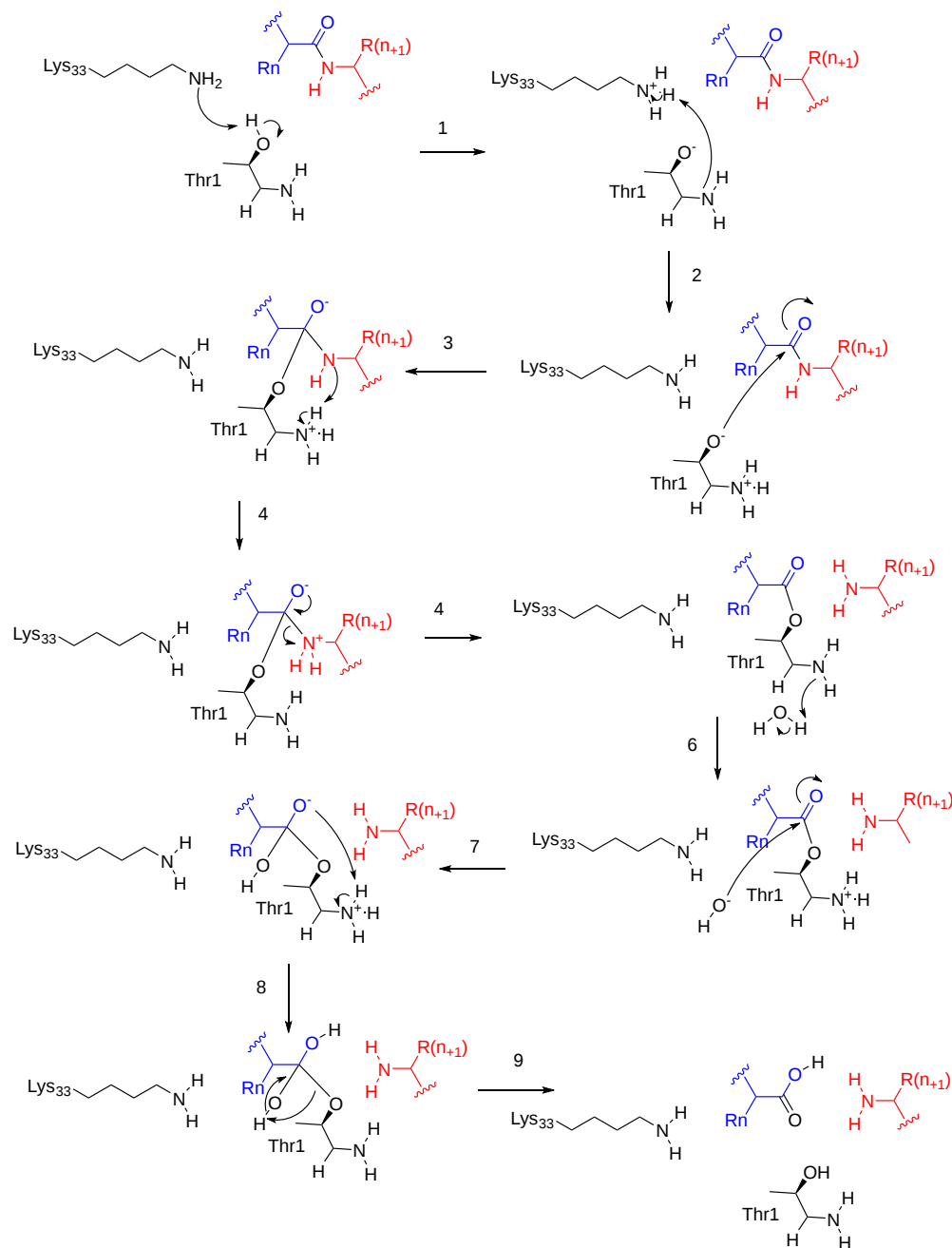


Figure [Math Processing Error]: Possible mechanism for the beta chain proteasome cleavage of proteins. after Arjun Saha, Gabriel Oanca, Dibyendu Mondal, and Arie Warshel. *The Journal of Physical Chemistry B* **2020**124 (27), 5626-5635. DOI: 10.1021/acs.jpcb.0c04435

Proteins are normally targeted to the proteasome using ubiquitin labels attached covalently to a lysine residue, usually in a chained form that produces a polyubiquitinated protein. The process of protein polyubiquitination is carried out by a highly specialized and diverse enzymatic system, which includes families of ubiquitin-activating enzymes (E1), ubiquitin-conjugating enzymes (E2), and ubiquitin ligases (E3).

The covalent attachment of ubiquitin to specific target proteins is primarily accomplished through a stepwise enzymatic cascade reaction, where ubiquitin is attached to substrates via the concerted action of ubiquitin-activating enzyme (E1), ubiquitin-conjugating enzyme (E2), and ubiquitin ligase (E3). The attachment of ubiquitin or ubiquitin chains to the substrate is a successive process, as shown in Figure [Math Processing Error]. First, the C-terminal carboxylic acid is activated by adenylation using a molecule of ATP, forming an **adenylate (AMP-) intermediate**. The **adenylate** acts as a good leaving group during the next reaction, where an E1-ubiquitin thioester bond is formed between the C-terminal Gly carboxyl group of ubiquitin and the active site Cys of the E1 enzyme. AMP leaves the active site at this point. Note that ATP is used in many reactions to activate carboxylic acid

functional groups through the formation of an adenylate intermediate, and that this will be seen as a theme in many different types of reactions throughout this textbook. Once the ubiquitin is docked as a thioester on E1, it can be transferred to a Cys residue of the E2 enzyme to form an E2-ubiquitin thioester-linked intermediate. This enzymatic reaction is known as a **transesterification**.

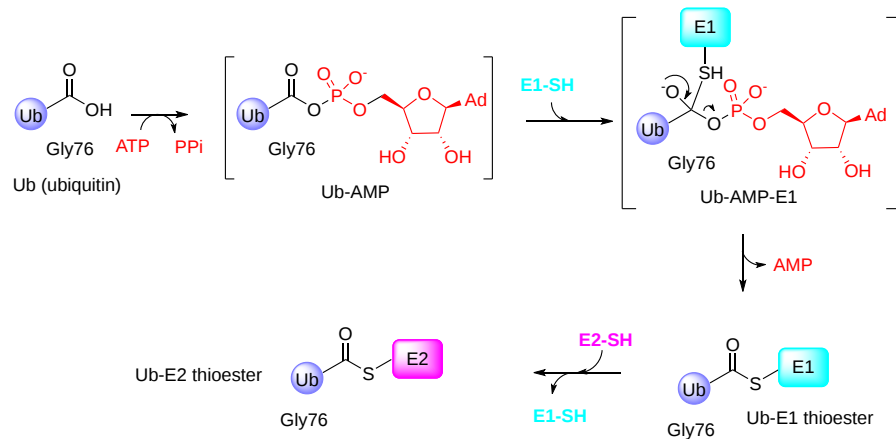


Figure [Math Processing Error]: Enzymatic Activity of the E1 Ubiquitin Activating Enzyme. Conversion of Gly 76 on ubiquitin to E1 and E2 thioesters. After Hann, Z.S., et. al. (2019) *PNAS* 116(31)15475-15484.

Eventually, the E2 transfers the ubiquitin to the substrate protein by E3. Ubiquitin is conjugated to the target protein through an isopeptide bond between its C-terminal glycine (Gly76) and the ϵ -amino group of a lysine residue. There are three typical ways of linking ubiquitin with the substrate, as shown in Figure [Math Processing Error]. The first is called mono-ubiquitination, which refers to the modification of a single site on a substrate by a single ubiquitin molecule. The second is multi-mono-ubiquitination, which involves adding several ubiquitin molecules repetitively to distinct sites. The third is called polyubiquitination, which includes both linear and branched polyubiquitination. Ubiquitin molecules are added to the same site on a substrate. In the second and third ways of linking, the previously attached ubiquitin serves as the “acceptor” of subsequently added ubiquitin. Of course, polyubiquitin chains linked by the same Lys are homogeneous, while those linked at different Lys are heterogeneous or mixed ones.

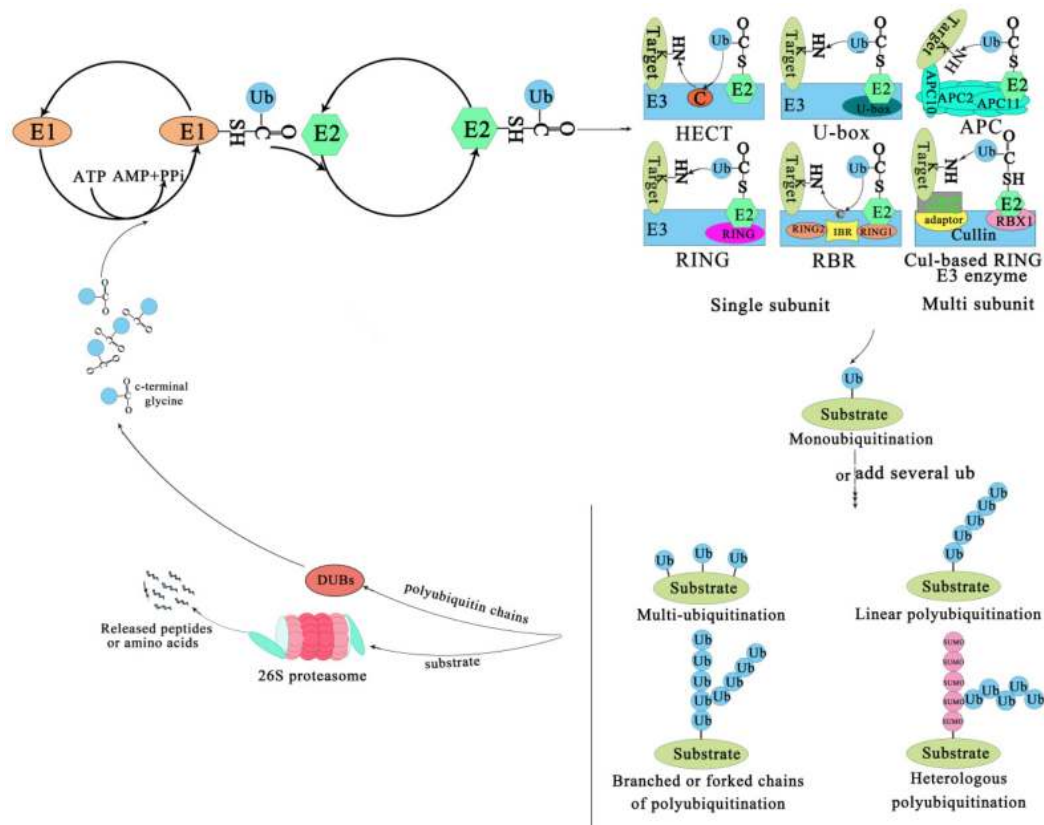


Figure [Math Processing Error]: The Ubiquitin-Proteasome System. Liu, W., et.al. (2020) Int J Mol Sci 21(8):2894

Three major enzymes catalyze the process of ubiquitination from activation to attachment to the substrate. The substrates labeled by ubiquitin are degraded by the 26S proteasome or play a non-degradative role in other processes. Abbreviations: APC, Anaphase-

promoting complex; DUBs, Deubiquitinating enzymes; E1, Ubiquitin-activating enzyme; E2, Ubiquitin-conjugating enzyme; E3, Ubiquitin-ligase enzyme; Cul-based, Cullin-RING box1-Ligase; HECT, Homology to E6-AP C Terminus; Ub, Ubiquitin; SUMO, Small ubiquitin-related modifier; RBX1, RING-Box 1; RING, Really interesting new gene; RBR, RING1-IBR(cysteine/histidine-rich region)-RING2.

Subsequently, the substrate complex tagged by ubiquitin is either degraded by the 26S proteasome or executes non-proteolytic functions, such as regulating gene expression, cellular trafficking, or other biological processes. In most cases, polyubiquitinated proteins are recognized and degraded by the 26S proteasome, and the ubiquitin or ubiquitin chain is hydrolyzed and freed by deubiquitinating enzymes (DUBs) for reuse in further conjugation cycles after being removed from the substrate protein (Figure 27.3.6).

The family of E3 enzymes is large and diverse. It is estimated that there are 600-700 E3 enzymes in humans, representing approximately 5% of the human genome. Thus, E3 enzymes can be very substrate specific, leading to the specialized degradation of a small subset of proteins within the cell. E2 enzymes are the intersection between E1 and E3 enzymes, helping to determine the ubiquitination of specific target proteins by interacting with various types of E3 enzymes.

Figure [\[Math Processing Error\]](#) shows an [interactive iCn3D model](#) of the Yeast proteasome in resting state (C1-a) (6J2X). (long load)

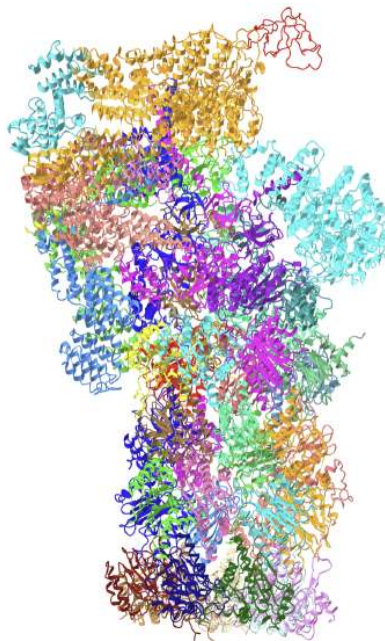


Figure [\[Math Processing Error\]](#): Yeast proteasome in resting state (C1-a) (6J2X). (Copyright; author via source). Click the image for a popup or use this external link: <https://structure.ncbi.nlm.nih.gov/icn3d/share.html?1AmGmkghbZD2zute7>

It's too complex to discuss in much detail. The chains are presented in different colors. You should be able to find the alpha and beta chains of the core 20S particle. The regulatory cap proteins, comprising the lid, are positioned on top of the alpha ring.

26.3.3: Summary

This chapter delves into the intricate regulatory mechanisms that govern protein synthesis and turnover, integrating translational control with proteostasis. It covers how variations at multiple levels—from the composition of ribosomes and mRNA structure to the influence of nascent peptides and the protein degradation machinery—collectively fine-tune gene expression.

Ribosome Heterogeneity and Specialized Translation

- **Diversity in Ribosome Composition:**

Studies in eukaryotes have revealed that ribosomes can differ in their ribosomal protein (RP) and rRNA content between cell types and developmental stages. This heterogeneity, arising from differential expression of RP paralogs or post-translational modifications (such as phosphorylation), supports the "ribosome filter hypothesis." According to this model, specialized ribosomes selectively translate subsets of mRNAs, thereby acting as determinants of translation efficiency and specificity.

- **Variability in rRNA Genes and Modifications:**

The presence of multiple, sometimes divergent, rRNA gene copies and chemically modified rRNA residues (e.g., 2'-OH methylation and pseudouridylation) further contributes to the functional diversity of ribosomes, allowing cells to adapt protein synthesis in response to environmental signals.

mRNA Sequence, Structure, and Translational Output

- **Influence of mRNA Features:**

The primary nucleotide sequence of an mRNA—including its GC content, codon usage, and codon pair patterns—as well as the secondary structure of its 5' untranslated region (UTR), play pivotal roles in determining translational output. These features affect the binding of ribosomes and the efficiency of initiation.

- **Role of Small RNAs:**

Interactions with microRNAs (miRNAs) and the involvement of Argonaute proteins can target specific mRNAs for degradation or repress their translation, adding another layer of regulation that fine-tunes protein levels.

Nascent Peptide Effects on Translation

- **Exit Tunnel Interactions:**

As the nascent peptide emerges from the peptidyl transferase center through a 100 Å-long exit tunnel, specific peptide sequences may interact with the tunnel walls. Such interactions can induce ribosome stalling, which is alleviated by specialized elongation factors like EF-P in bacteria.

- **Antibiotic Targeting:**

The ribosomal exit tunnel is also a binding site for macrolide antibiotics (e.g., erythromycin), which exploit these interactions to interfere with translation and inhibit bacterial growth.

Proteasome-Mediated Protein Degradation and Ubiquitination

- **Structure and Function of the Proteasome:**

The 26S proteasome, central to eukaryotic protein homeostasis, comprises a 20S core particle and regulatory 19S caps. The 20S core is a barrel-shaped structure formed by four stacked rings, with the outer rings (α -subunits) regulating substrate entry and the inner rings (β -subunits) housing the proteolytic active sites.

- **Ubiquitin Tagging System:**

Protein degradation is selectively targeted by polyubiquitination. This multistep process involves the sequential action of ubiquitin-activating enzymes (E1), conjugating enzymes (E2), and ligases (E3). Ubiquitin is attached to substrates via isopeptide bonds, marking them for recognition and degradation by the proteasome. Deubiquitinating enzymes (DUBs) subsequently recycle ubiquitin for reuse.

Integration of Translational Control and Proteostasis

- **Coordinated Regulation:**

The chapter emphasizes how regulation of translation—via ribosome heterogeneity, mRNA structural features, and nascent peptide dynamics—is intricately linked with the mechanisms controlling protein degradation. Together, these processes ensure that cellular protein levels are finely tuned in response to developmental cues and environmental stress.

In summary, this chapter provides a comprehensive view of the multifaceted regulation of translation and protein homeostasis. By understanding the variations in ribosome structure, the regulatory roles of mRNA features and nascent peptides, and the targeted degradation of proteins via the ubiquitin–proteasome system, students gain insight into the dynamic control of gene expression at both the synthesis and degradation levels.

26.3.4: References

This chapter was remixed and adapted from the following resources under [creative commons licensing](#):

1. Wikipedia contributors. (2020, June 27). Wobble base pair. In *Wikipedia, The Free Encyclopedia*. Retrieved 16:47, August 12, 2020, from https://en.Wikipedia.org/w/index.php?title=Wobble_base_pair&oldid=964760055
2. Lorenz, C., Lünse, C., and Mörl, M. (2017) tRNA modifications: Impact on structure and thermal adaptation. *Biomolecules* 7(2)35. Available at: <https://www.mdpi.com/2218-273X/7/2/35/html>

3. Pan, T. (2018) Modifications and functional genomics of human transfer RNA. *Cell Research* 28:395-404. Available at: <https://www.nature.com/articles/s41422-018-0013-y#Sec3>
4. Bednářová, A., Hanna, M., Durham, I., Van Cleave, T., England, A., Chaudhuri, A., and Krishnan, N. (2017) Lost in translation: Defects in transfer RNA modifications and neurological disorders. *Front. Mol. Neurosci.* 10:135. Available at: https://www.researchgate.net/publication/316440980_Lost_in_Translation_Defects_in_Transfer_RNA_Modifications_and_Neurological_Disorders
5. Wikipedia contributors. (2020, May 17). Transfer RNA. In *Wikipedia, The Free Encyclopedia*. Retrieved 23:03, August 15, 2020, from https://en.Wikipedia.org/w/index.php?title=Transfer_RNA&oldid=957227343
6. Gomez, M.A.R., and Ibba M. (2020) Aminoacyl-tRNA Synthetases. *RNA*, doi: 10.1261/rna.071720.119 Available at: <https://rnajournal.cshlp.org/content/early/2020/04/17/rna.071720.119.abstract>
7. Li, R., Macnamara, L.M., Leuchter, J.D., Alexander, R.W., and Cho, S.S. (2015) MD Simulations of tRNA and Aminoacyl-tRNA Synthetases: Dynamics, Folding, Binding, and Allostery. *Int. J. Mol. Sci.* 16(7):15872-15902. Available at: <https://www.mdpi.com/1422-0067/16/7/15872>
8. Wikipedia contributors. (2020, August 15). Ribosome. In *Wikipedia, The Free Encyclopedia*. Retrieved 05:22, August 16, 2020, from <https://en.Wikipedia.org/w/index.php?title=Ribosome&oldid=973144885>
9. Kater, L., Thoms, M., Barrio-Garcia, C., Cheng, J., Ismail, S., Ahmed, Y.L., Bange, G., Kressler, D., Berninghausen, O., Sinning, I., Hurt, E., and Beckmann, R. (2017) Visualizing the assembly pathway of nucleolar Pre-60S Ribosomes. *Cell* 171(7):1599-1610. Available at: <https://www.sciencedirect.com/science/article/pii/S0092867417314290>
10. Bock, L.V., Kolár, M.H., Grubmüller, H. (2018) Molecular simulations of the ribosome and associated translation factors. *Cur. Op. Struc. Bio.* 49:27-35. Available at: <https://www.sciencedirect.com/science/article/pii/S0959440X1730132X>
11. Doris, S.M., Smith, D.R., Beamesderfer, J.N., Raphael, B.J., Nathanson, J.A., and Gerbi, S.A. (2015) Universal and domain-specific sequences in 23S-28S ribosomal RNA identified by computational phylogenetics. *RNA* 21:1719-1730. Available at: https://www.researchgate.net/publication/281141702_Universal_and_domain-specific_sequences_in_23S-28S_ribosomal_RNA_identified_by_computational_phylogenetics
12. Aleksashin, M.A., Leppik, M., Hochenberry, A.J., Klepacki, D., Vázquez-Laslop, N., Jewett, M.C., Remme, J., and Mankin A.S. (2019) Assembly and functionality of the ribosome with tethered subunits. *Nature Communications* 10:930. Available at: <https://www.nature.com/articles/s41467-019-08892-w#rightslink>
13. Wikipedia contributors. (2020, July 3). Formylation. In *Wikipedia, The Free Encyclopedia*. Retrieved 22:03, August 16, 2020, from <https://en.Wikipedia.org/w/index.php?title=Formylation&oldid=965827476>
14. Gualerzi, C.O., and Pon C.L. (2015) Initiation of mRNA translation in bacteria: structural and dynamic aspects. *Cell Mol Life Sci.* 72:4341-4367. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4611024/>
15. Wikipedia contributors. (2020, June 14). Kozak consensus sequence. In *Wikipedia, The Free Encyclopedia*. Retrieved 06:03, August 18, 2020, from https://en.Wikipedia.org/w/index.php?title=Kozak_consensus_sequence&oldid=962444929
16. Knight, J.R.P., Garland, G., Pöyry, T., Mead, E., Vlahov, N., Sfakianos, A., Frosso, S., De-Lima-Hedayioglu, F., Mallucci, G.R., von der Haar, T., Smales, C.M., Sansom, O.J., and Willis, A.E. (2020) Control of translation elongation in health and disease. *Dis. Mod. and Mech.* 13: dmm043208. Available at: <https://dmm.biologists.org/content/13/3/dmm043208>
17. Adio, S., Sharma, H., Senyushkina, T., Karki, P., Maracci, C., Wohlgemuth, I., Holtkamp, W., Peske, R., and Rodina, M.V. (2018) Dynamics of ribosomes and release factors during translation termination in *E. coli*. *eLife* 7:e34253. Available at: <https://elifesciences.org/articles/34252>
18. Ge, X., Oliveira, A., Hjort, K., Bergfors, T., Guitierrez-de-Terán, H., Andersson, D.I., Sanyal, S., and Åqvist, J. (2019) Inhibition of translation termination by small molecules targeting ribosomal release factors. *Scientific Reports* 9: 15424. Available at: <https://www.nature.com/articles/s41598-019-51977-1#rightslink>
19. Svidritskiy, E., Demo G., Loveland A.B., Xu, C., and Korostelev, A.A. (2019) Extensive ribosome and RF2 rearrangements during translation termination. *eLife* 8:e46850. Available at: <https://elifesciences.org/articles/46850>
20. Sauert, M., Temmel, H., and Moll, I. (2015) Heterogeneity of the translational machinery: Variations on a common theme. *Biochimie* 114:39-47. Available at: <https://www.sciencedirect.com/science/article/pii/S0300908414003952>
21. Shen, Q., Wang, G., Li, S., Liu, X., Lu, S., Chen, Z., Song, K., Yan, J., Geng, L., Huang, Z., Huang, W., Chen, G., and Zhang, J. (2016) ASDv3.0: Unraveling allosteric regulation with structural mechanisms and biological networks. *Nucleic Acids Research* 44(D1):D527-D535. Available at: <https://academic.oup.com/nar/article/44/D1/D527/2503129>
22. Wikipedia contributors. (2020, January 19). Isozyme. In *Wikipedia, The Free Encyclopedia*. Retrieved 03:42, May 19, 2020, from <https://en.Wikipedia.org/w/index.php?title=Isozyme&oldid=936548836>

23. Wikipedia contributors. (2020, April 30). COX-2 inhibitor. In *Wikipedia, The Free Encyclopedia*. Retrieved 07:00, May 22, 2020, from [en.Wikipedia.org/w/index.php?title=COX-2_inhibitor&oldid=954080651](https://en.wikipedia.org/w/index.php?title=COX-2_inhibitor&oldid=954080651)
24. Clarkson, C.W. (2018) Major Side Effects of NSAIDs and COX-2 Selective Inhibitors. *TUSOM Pharmwiki*. Available at: http://tmedweb.tulane.edu/pharmwiki/doku.php/nsaid_side_effects?do=
25. Santos, A.L. and Lindner, A.B. (2017) Protein Posttranslational Modifications: Roles in Aging and Age-Related Disease. *Oxidative Medicine and Cellular Longevity*, Article ID: 5716409. Available at: <https://www.hindawi.com/journals/omcl/2017/5716409/#copyright>
26. Wikipedia contributors. (2020, May 12). Protein phosphorylation. In *Wikipedia, The Free Encyclopedia*. Retrieved 17:38, May 25, 2020, from [en.Wikipedia.org/w/index.php?title=Protein_phosphorylation&oldid=956228409](https://en.wikipedia.org/w/index.php?title=Protein_phosphorylation&oldid=956228409)
27. Szylveszter, K.P., Németh, T., and Mócsai, A. (2019) Tyrosine Kinases in Autoimmune and Inflammatory Skin Diseases. *Front. Immunol.* 10.3389(2019.01862). Available at: <https://www.frontiersin.org/articles/10.3389/fimmu.2019.01862/full>
28. Wikipedia contributors. (2020, May 7). Histone. In *Wikipedia, The Free Encyclopedia*. Retrieved 21:12, May 25, 2020, from [en.Wikipedia.org/w/index.php?title=Histone&oldid=955458038](https://en.wikipedia.org/w/index.php?title=Histone&oldid=955458038)
29. Wikipedia contributors. (2020, May 13). Mucin. In *Wikipedia, The Free Encyclopedia*. Retrieved 01:10, June 7, 2020, from [en.Wikipedia.org/w/index.php?title=Mucin&oldid=956387296](https://en.wikipedia.org/w/index.php?title=Mucin&oldid=956387296)
30. Wikipedia contributors. (2020, March 15). Allosteric regulation. In *Wikipedia, The Free Encyclopedia*. Retrieved 04:07, June 7, 2020, from [en.Wikipedia.org/w/index.php?title=Allosteric_regulation&oldid=945637073](https://en.wikipedia.org/w/index.php?title=Allosteric_regulation&oldid=945637073)
31. Dixit, Ajay. Dawra, Rajinder K. Dudeja, Vikas. Saluja, Ashok K. (2016). Role of trypsinogen activation in genesis of pancreatitis. *Pancreapedia: Exocrine Pancreas Knowledge Base*, DOI: [10.3998/panc.2016.25](https://doi.org/10.3998/panc.2016.25)
32. Coll-Martinez, B., and Crosas, B. (2019) How the 26S Proteasome Degrades Ubiquitinated Proteins in the Cell. *Biomolecules* 9(9):395. Retrieved from: <https://www.mdpi.com/2218-273X/9/9/395>
33. Liu, W., Tang, X., Qi, X., Ghimire, S., Ma, R., Li, S., Zhang, N., and Si H. (2020) The Ubiquitin Conjugating Enzyme: An Important Ubiquitin Transfer Platform in Ubiquitin-Proteasome System. *Int J Mol Sci* 21(8):2894. Retrieved from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7215765/>

This page titled [26.3: Translational Regulation and Protein Degradation](#) is shared under a [not declared](#) license and was authored, remixed, and/or curated by [Henry Jakubowski and Patricia Flatt](#).

- [Current page](#) by Henry Jakubowski and Patricia Flatt has no license indicated.
- [11.2: Diffusion Across a Membrane - Passive and Facilitated Diffusion](#) by Henry Jakubowski and Patricia Flatt has no license indicated.
- [23.1: Gene Mapping and Chromosomal Karyotypes](#) by Henry Jakubowski and Patricia Flatt is licensed [CC BY-SA 4.0](#).

CHAPTER OVERVIEW

27: Regulation of Gene Expression

[Return to Fundamentals of Biochemistry](#)

[Search Fundamentals of Biochemistry](#)

[27.1: Regulation of Gene Expression in Bacteria](#)

[27.2: Regulation of Gene Expression in Eukaryotes](#)

This page titled [27: Regulation of Gene Expression](#) is shared under a [not declared](#) license and was authored, remixed, and/or curated by [Henry Jakubowski and Patricia Flatt](#).

27.1: Regulation of Gene Expression in Bacteria

Learning Goals (ChatGPT o3-mini)

- **Explain the Importance of Gene Regulation:**
 - Describe how identical genomes can produce diverse phenotypes in multicellular organisms and enable adaptive responses in single-celled organisms.
 - Discuss why precise control of gene expression is essential for cellular differentiation, metabolism, and overall homeostasis.
- **Differentiate Between Structural and Regulatory Genes:**
 - Identify the roles of structural genes (encoding proteins or enzymes) and regulatory genes (encoding transcription factors and other regulators) in controlling gene expression.
- **Understand Operon Organization in Prokaryotes:**
 - Explain the concept of the operon, including the organization of structural genes under a single promoter and regulatory region.
 - Compare constitutively expressed operons with those that are inducible or repressible.
- **Analyze Specific Examples of Operon Regulation:**
 - Describe the mechanism of the trp operon, focusing on how the trp repressor binds to the operator in the presence of tryptophan to block transcription.
 - Explain the regulation of the lac operon, including the roles of the lac repressor, inducer molecules (e.g., allolactose, IPTG), and the CAP-cAMP complex in activating transcription when glucose is low.
- **Examine Global Regulatory Mechanisms in Bacteria:**
 - Define regulons and discuss how they allow the simultaneous control of multiple operons in response to environmental changes.
 - Explain the role of alarmones such as pppGpp in the stringent response, and how they adjust gene expression during nutrient stress.
- **Describe Quorum Sensing and Population-Level Regulation:**
 - Explain the basic components of quorum sensing systems (signal synthase, signal receptor, and signal molecules) and their role in coordinating bacterial behavior, including virulence, biofilm formation, and secondary metabolism.
 - Analyze how quorum sensing enables bacteria to alter gene expression based on population density.
- **Understand the Role of Alternate σ Factors:**
 - Describe how the use of different σ subunits by bacterial RNA polymerase allows for rapid and global shifts in gene expression in response to environmental signals.
 - Discuss examples of σ factor switching in processes such as sporulation.
- **Explore Post-Initiation Regulatory Mechanisms:**
 - Define attenuation and explain how leader sequences and the formation of alternative RNA secondary structures can control transcription termination.
 - Describe riboswitches, including how they bind small intracellular molecules to influence mRNA structure and regulate transcription or translation.

These learning goals are intended to provide you with a comprehensive framework for understanding how gene expression is dynamically regulated in response to both internal signals and external environmental cues.

27.1.1: Introduction

Each nucleated cell in a multicellular organism contains copies of the same DNA. Similarly, all cells in two pure bacterial cultures inoculated from the same starting colony contain the same DNA, except for changes that arise from spontaneous mutations. If each

cell in a multicellular organism has the same DNA, then how is it that cells in different parts of the organism's body exhibit different characteristics? Similarly, how is it that the same bacterial cells within two pure cultures exposed to different environmental conditions can exhibit different phenotypes? In both cases, each genetically identical cell does not "turn on" or express the same set of genes. Only a subset of proteins in a cell at a given time is expressed.

Genomic DNA contains both *structural genes*, which encode products that serve as cellular structures or enzymes, and *regulatory genes*, which encode products that regulate gene expression. The expression of a gene is a highly regulated process. Whereas regulating gene expression in multicellular organisms enables cellular differentiation, in single-celled organisms like prokaryotes, it primarily ensures that a cell's resources are not wasted on making proteins that the cell does not need at that time.

Elucidating the mechanisms that control gene expression is crucial to understanding human health. Malfunctions in this process in humans lead to the development of cancer and other diseases. Understanding the interaction between the gene expression of a pathogen and that of its human host is crucial for understanding a particular infectious disease. Gene regulation involves a complex web of interactions within a given cell among signals from the cell's environment, signaling molecules within the cell, and the cell's DNA. These interactions result in the expression of specific genes and the suppression of others, depending on the circumstances.

Prokaryotes and eukaryotes share some similarities in their mechanisms for regulating gene expression; however, gene expression in eukaryotes is more complex due to the temporal and spatial separation between the processes of transcription and translation. Thus, although most regulation of gene expression occurs through transcriptional control in prokaryotes, regulation of gene expression in eukaryotes occurs at the transcriptional level and post-transcriptionally (after the primary transcript has been made).

In bacteria and archaea, structural proteins with related functions are typically encoded together within the genome in a block known as an *operon*. They are transcribed together under the control of a single *promoter*, resulting in the formation of a **polycistronic transcript**, as shown in Figure [Math Processing Error]. In this way, the regulation of transcription for all the structural genes encoding the enzymes that catalyze the many steps in a single biochemical pathway can be controlled simultaneously, because they will either all be needed at the same time or none will be required. For example, in *E. coli*, all the structural genes that encode enzymes needed to utilize lactose as an energy source are clustered together in the lactose (or *lac*) operon, under the control of a single promoter, the *lac* promoter. French scientists François Jacob (1920–2013) and Jacques Monod, at the Pasteur Institute, were the first to demonstrate the organization of bacterial genes into operons through their studies on the *lac* operon of *E. coli*. For this work, they won the Nobel Prize in Physiology or Medicine in 1965.

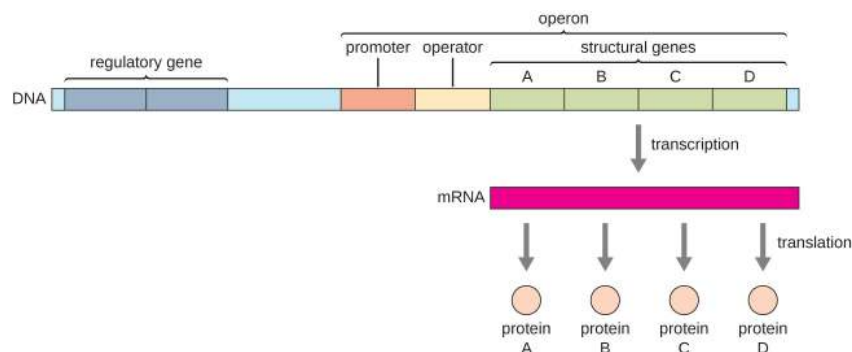


Figure [Math Processing Error]: **Schematic Representation of an Operon.** In prokaryotes, structural genes of related functions are often organized together on the genome and transcribed together under the control of a single promoter. The operon's regulatory region includes both the promoter and the operator. If a repressor binds to the operator, then the structural genes will not be transcribed. Alternatively, activators may bind to the regulatory region, enhancing transcription. [Parker, N., et. al. \(2019\) Microbiology. Openstax](#)

Each operon includes DNA sequences that influence its own transcription; these are located in a region called the regulatory region. The regulatory region encompasses the promoter and the surrounding region to which **transcription factors and** proteins encoded by regulatory genes can bind. Transcription factors influence the binding of *RNA polymerase* to the promoter, allowing it to progress and transcribe structural genes. A **repressor** is a transcription factor that suppresses the transcription of a gene in response to an external stimulus by binding to a DNA sequence within the regulatory region called the *operator*, which is located between the RNA polymerase binding site of the promoter and the transcriptional start site of the first structural gene. Repressor binding physically blocks RNA polymerase from transcribing structural genes. Conversely, an **activator** is a transcription factor that increases the transcription of a gene in response to an external stimulus by facilitating the binding of RNA polymerase to the

promoter. An *inducer*, a third type of regulatory molecule, is a small molecule that either activates or represses transcription by interacting with a repressor or an activator.

In prokaryotes, there are examples of operons whose gene products are required consistently and whose expression, therefore, is not regulated. Such operons are *constitutively expressed*, meaning they are transcribed and translated continuously to provide the cell with constant intermediate levels of the protein products. Such genes encode enzymes involved in housekeeping functions required for cellular maintenance, including DNA replication, repair, and expression, as well as enzymes involved in core metabolism. In contrast, other prokaryotic operons are expressed only when needed and are regulated by repressors, activators, and inducers.

Prokaryotic operons are commonly controlled by the binding of repressors to operator regions, thereby preventing the transcription of the structural genes. Such operons are classified as either *repressible operons* or *inducible operons*. Repressible operons, like the tryptophan (*trp*) operon, typically contain genes encoding enzymes required for a biosynthetic pathway. As long as the product of the pathway, like tryptophan, continues to be required by the cell, a repressible operon will continue to be expressed. However, when the product of the biosynthetic pathway begins to accumulate in the cell, eliminating the need for the cell to continue producing more, the expression of the operon is repressed. Conversely, *inducible operons*, such as the *lac* operon of *E. coli*, often contain genes encoding enzymes involved in a pathway that metabolizes a specific substrate, like lactose. These enzymes are only required when that substrate is available, thus expression of the operons is typically induced only in the presence of the substrate.

27.1.2: The *trp* Operon - A Repressible Operon

E. coli can synthesize tryptophan using enzymes that are encoded by five structural genes located next to each other in the *trp* operon, as shown in Figure [Math Processing Error]. When environmental tryptophan is low, the operon is turned on. This means that transcription is initiated, the genes are expressed, and tryptophan is synthesized. However, if tryptophan is present in the environment, the *trp* operon is turned off. Transcription does not occur, and tryptophan is not synthesized.

When tryptophan is not present in the cell, the repressor binds to the operator by itself; therefore, the operon is active, and tryptophan is synthesized. However, when tryptophan accumulates in the cell, two tryptophan molecules bind to the *trp* repressor molecule, which changes its shape, allowing it to bind to the *trp* operator. This binding of the active form of the *trp* repressor to the operator blocks RNA polymerase from transcribing the structural genes, stopping the expression of the operon. Thus, the actual product of the biosynthetic pathway controlled by the operon regulates the expression of the operon.

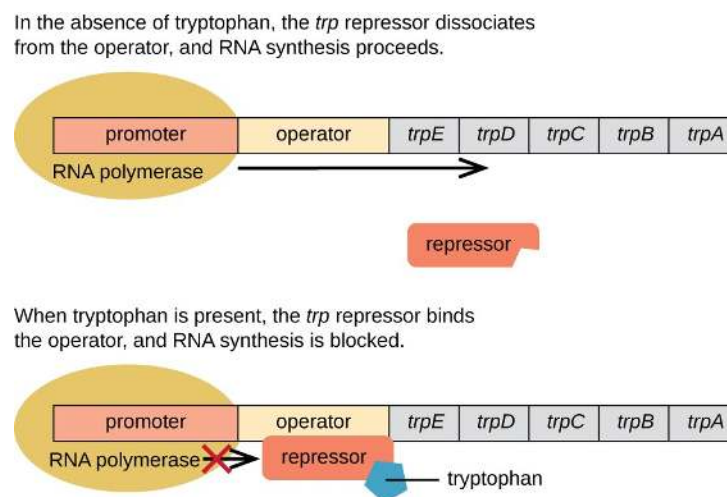


Figure [Math Processing Error]: The *Trp* Operon. [Parker, N., et. al. \(2019\) Microbiology. Openstax](#)

The five structural genes needed to synthesize tryptophan in *E. coli* are located next to each other in the *trp* operon. When tryptophan is absent, the repressor protein does not bind to the operator, and the genes are transcribed. When tryptophan is plentiful, tryptophan binds the repressor protein at the operator sequence. This physically blocks the RNA polymerase from transcribing the tryptophan biosynthesis genes.

Figure [Math Processing Error] shows an [interactive iCn3D model](#) of the *E. Coli* *Trp* repressor-operator complex (1TRO).

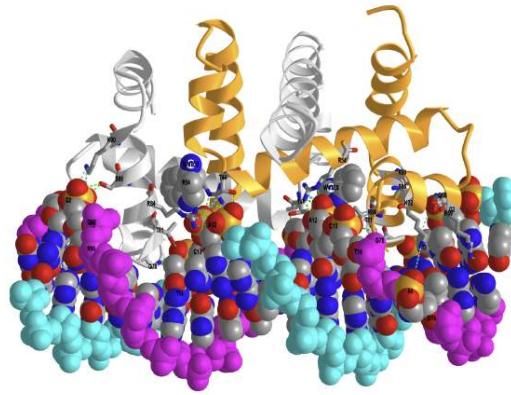


Figure [Math Processing Error]: E. Coli Trp repressor-operator complex (1TRO). (Copyright; author via source). Click the image for a popup or use this external link: <https://structure.ncbi.nlm.nih.gov/1...tjbYwr47gVZjn8>

The Trp repressor is depicted as a dimer, with one subunit shown in gray and the other in gold. The backbone of the two DNA strands is shown in spacefill magenta and cyan, except where the bases on the major groove interact with the Trp repressor. The tryptophan in each of the proton monomers is shown in spacefill with CPK colors and labeled. Noncovalent interactions (hydrogen bonds and salt bridges) between the protein and DNA are shown with dotted lines. 6 water-mediated hydrogen bonds to phosphate are not shown. Note that there are a few H bonds between the protein and base hydrogen bond donors and acceptors, suggesting that the repressor might bind specifically through geometric interactions with the backbone, along with the water-mediated hydrogen bonds.

27.1.3: The Lac Operon: An Inducible Operon

The *lac* operon is an example of an *inducible operon* that is also subject to activation in the absence of glucose. The *lac* operon encodes three structural genes, *lacZ*, *lacY*, and *lacA*, necessary to acquire and process the disaccharide lactose from the environment, as shown in Figure [Math Processing Error].

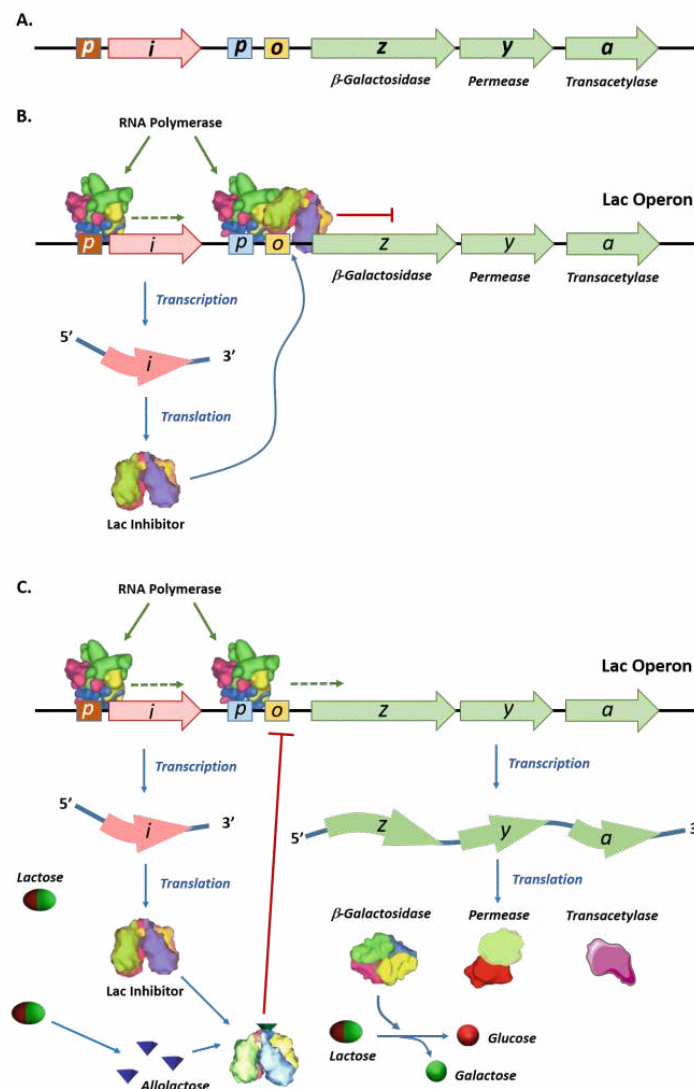


Figure [Math Processing Error]: Biological Activity of the *lac* Operon. Figure modified from [Esmaeili, A., et. al. \(2015\) BMC Bioinformatics 16:311](#)

Panel (A) shows a schematic representation of the *lac* operon in *E. coli*. The *lac* operon has three structural genes, *lacZ*, *lacY*, and *lacA*, that encode for β -galactosidase, permease, and galactoside acetyltransferase, respectively. The promoter (*p*) and operator (*o*) sequences that control the expression of the operon are shown. Upstream of the *lac* operon is the *lac* repressor gene, *lacI*, controlled by the *lacI* promoter (*p*).

Panel (B) shows the *lac* repressor inhibition of the *lac* operon gene expression in the absence of lactose. The *lac* repressor binds to the operator sequence of the operon and prevents the RNA polymerase enzyme, bound to the promoter (*p*), from initiating transcription.

Panel (C) shows that in the presence of lactose, some of the lactose is converted into allolactose, which binds and inhibits the activity of the *lac* repressor. The *lac* repressor-allolactose complex cannot bind to the operator region of the operon, thereby freeing RNA polymerase and allowing for the initiation of transcription. Expression of the *lac* operon genes enables the breakdown and utilization of lactose as a food source within the organism.

The *lacZ* gene encodes the β -galactosidase (β -gal) enzyme responsible for the hydrolysis of lactose into simple sugars glucose and galactose, as shown in Figure [Math Processing Error]. The β -gal enzyme can also mediate the breakdown of the alternate substrate 5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside (Xgal) (panel B). The breakdown product, 5-bromo-4-chloro-3-hydroxyindole – 1, spontaneously dimerizes to form the intensely blue product, 5,5'-dibromo-4,4'-dichloro-indigo – 2. Thus, Xgal has been a valuable research tool, not only in the study of β -gal enzymatic activity but also in the development of the commonly used blue-white DNA cloning system, which utilizes the β -gal enzyme as a marker in molecular cloning experiments.

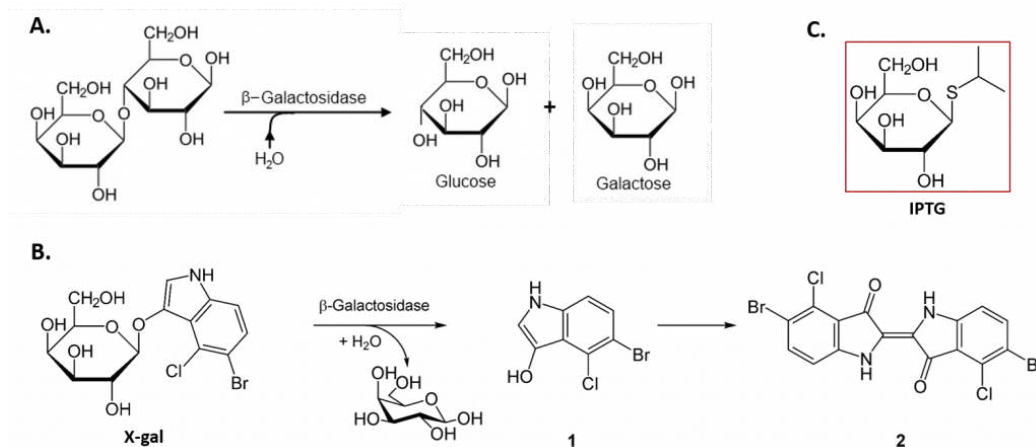


Figure [Math Processing Error]: Reactions Controlled by the Expression of the Lac Operon. Andreas Piehler, Yikrazuul, and NUR0tikerr Panel (A) shows the expression of the β -galactosidase enzyme enables the breakdown of lactose into simple sugars, glucose, and galactose for *E. coli* to use as a food resource. Panel (B) shows that the β -galactosidase enzyme also mediates the breakdown of the non-native substrate 5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside (Xgal). Breakdown product (1) 5-bromo-4-chloro-3-hydroxyindole quickly dimerizes into the intensely blue product (2) 5,5'-dibromo-4,4'-dichloro-indigo making it a useful tool for molecular biology. Panel (C) shows that β -D-1-thiogalactopyranoside (IPTG) can serve as a non-native inducer of the lac operon. It mimics the structure of lactose and binds with the Lac Repressor.

The lac operon contains two more genes, in addition to *lacZ* (Fig. 4). The *lacY* gene encodes a permease that increases the uptake of lactose into the cell, and *lacA* encodes a galactoside acetyltransferase (GAT) enzyme. The exact function of GAT during lactose metabolism has not been conclusively elucidated, but acetylation is thought to play a role in the transport of the modified sugars.

For the *lac* operon to be expressed, lactose must be present. This makes sense for the cell because it would be energetically wasteful to produce the enzymes necessary for lactose processing if lactose were not available.

In the absence of lactose, the *lacI* gene is constitutively expressed, resulting in the production of the *lac* repressor protein (Fig. 28.2.3B). The *lac* repressor binds to the operator region of the *lac* operon and physically prevents RNA polymerase from transcribing the structural genes (Fig. 28.2.3B). However, when lactose is present, the lactose inside the cell is converted to allolactose. Allolactose serves as an inducer molecule, binding to the repressor and altering its shape so that it can no longer bind to the operator DNA (Fig. 28.2.3 C). Removal of the repressor in the presence of lactose enables RNA polymerase to proceed through the operator region and initiate transcription of the *lac* structural genes. In addition to lactose, laboratory experiments have revealed that the non-natural compound Isopropyl β -D-1-thiogalactopyranoside (IPTG) can also bind to the *lac* repressor and induce the expression of the *lac* operon (panel C). Similar to Xgal, this compound has also been used as a research tool for molecular cloning.

Figure [Math Processing Error] shows an [interactive iCn3D model](#) of the lactose operon repressor and its complexes with DNA (1LBG).

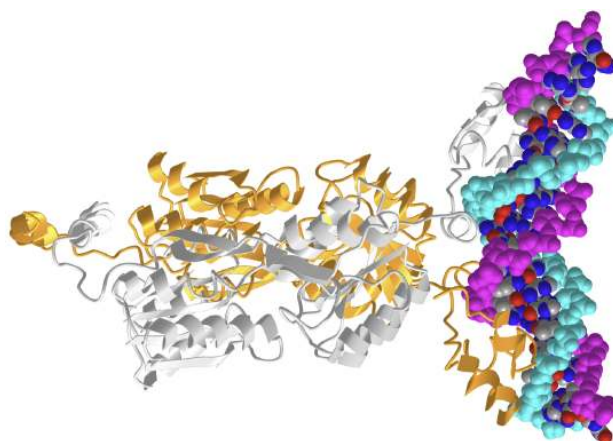


Figure [Math Processing Error]: Lactose operon repressor and its complexes with DNA (1LBG). (Copyright; author via source). Click the image for a popup or use this external link: <https://structure.ncbi.nlm.nih.gov/1...a79MvrAHGpJk99>

The resolution of the structure above was insufficient to show the amino acid side chains. Figure [\[Math Processing Error\]](#) shows an [interactive iCn3D model](#) of the [NMR solution structure](#) of the dimer of LAC repressor DNA-binding domain complexed to its natural operator O1 (2KEI).

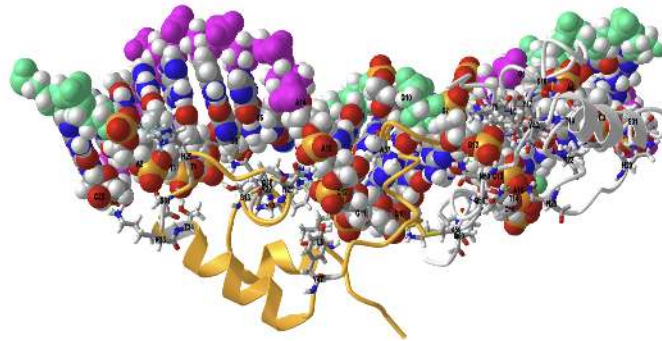


Figure [\[Math Processing Error\]](#): NMR solution structure of Dimer of LAC repressor DNA-binding domain complexed to its natural operator O1 (2KEI). (Copyright; author via source). Click the image for a popup or use this external link: <https://structure.ncbi.nlm.nih.gov/i...p48VrPFXg6WgD9>

Note the presence of white spheres representing hydrogen atoms (these do not appear in crystal structures but do in NMR structures). Color coding is the same as above. Zoom in to see specific interactions between the protein and the exposed DNA base hydrogen bond donors and acceptors. The complex of O1 and O2 exhibits similar specific and nonspecific contacts, which is expected given that the lambda repressor has a similar affinity for these two operator sites. In contrast, one side of the O3 complex exhibits a loss of protein-DNA interactions, consistent with its lower affinity for the operator O3.

27.1.4: The Lac Operon: Activation by Catabolite Activator Protein

Bacteria typically can use a variety of substrates as carbon sources. However, because glucose is usually preferable to other substrates, bacteria have mechanisms to ensure that alternative substrates are only used when glucose has been depleted. Additionally, bacteria have mechanisms to ensure that the genes encoding enzymes for using alternative substrates are expressed only when the alternative substrate is available. In the 1940s, Jacques Monod was the first to demonstrate a preference for certain substrates over others through his studies of *E. coli*'s growth when cultured in the presence of two different substrates simultaneously. Such studies generated diauxic growth curves, as shown in Figure [\[Math Processing Error\]](#). Although the preferred substrate glucose is used first, *E. coli* grows quickly, and the enzymes for lactose metabolism are absent. However, once glucose levels are depleted, growth rates slow, inducing the expression of the enzymes needed for the metabolism of the second substrate, lactose. Notice how the growth rate in lactose is slower, as indicated by the lower steepness of the growth curve.

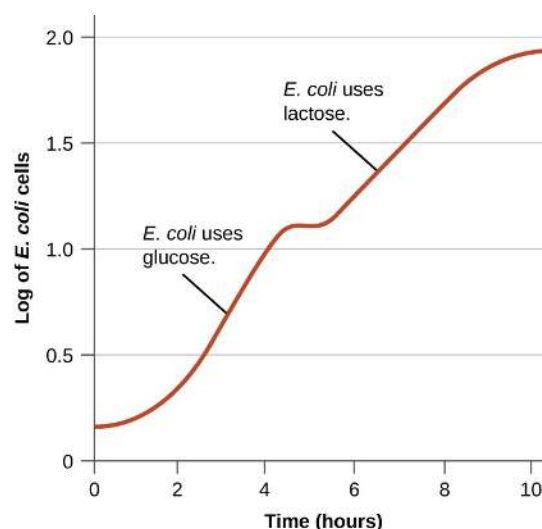


Figure [\[Math Processing Error\]](#): Utilization of Glucose in *E. Coli*. When grown in the presence of two substrates, *E. coli* uses the preferred substrate (in this case, glucose) until it is depleted. Then, enzymes needed for the metabolism of the second substrate are expressed and growth resumes, although at a slower rate. [Parker, N., et. al. \(2019\) Microbiology. Openstax](#)

The ability to switch from glucose use to another substrate, like lactose, is a consequence of the activity of an enzyme called Enzyme IIA (EIIA). When glucose levels drop, cells produce less ATP from catabolism, and EIIA becomes phosphorylated. Phosphorylated EIIA activates adenyl cyclase, an enzyme that converts some of the remaining ATP to cyclic AMP (cAMP), a cyclic derivative of AMP and an important signaling molecule involved in glucose and energy metabolism in *E. coli*, as shown in Figure [Math Processing Error]. As a result, cAMP levels begin to rise in the cell. This is an indicator to the cell that overall energy levels are low and that ATP is being depleted.

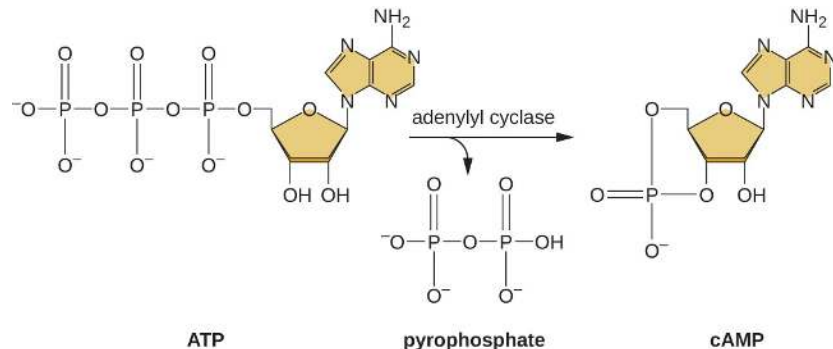


Figure [Math Processing Error]: Conversion of ATP to cAMP. When ATP levels decrease due to the depletion of glucose, some remaining ATP is converted to cAMP by adenyl cyclase. Thus, increased cAMP levels signal glucose depletion. Figure from: [Parker, N., et. al. \(2019\) Microbiology. Openstax](#)

The *lac* operon also plays a crucial role in this switch from glucose to lactose utilization. When glucose is scarce, the accumulating cAMP caused by increased adenyl cyclase activity binds to catabolite activator protein (CAP), also known as cAMP receptor protein (CRP). The complex binds to the promoter region of the *lac* operon, as shown in Figure [Math Processing Error]. In the regulatory regions of these operons, a CAP binding site is located upstream of the RNA polymerase binding site in the promoter. The binding of the CAP-cAMP complex to this site increases the binding ability of RNA polymerase to the promoter region to initiate the transcription of the structural genes. Thus, in the case of the *lac* operon, for transcription to occur, lactose must be present (removing the lac repressor protein), and glucose levels must be depleted (allowing the binding of an activating protein). When glucose levels are high, there is catabolite repression of operons encoding enzymes for the metabolism of alternative substrates. Because of low cAMP levels under these conditions, there is an insufficient amount of the CAP-cAMP complex to activate the transcription of these operons.

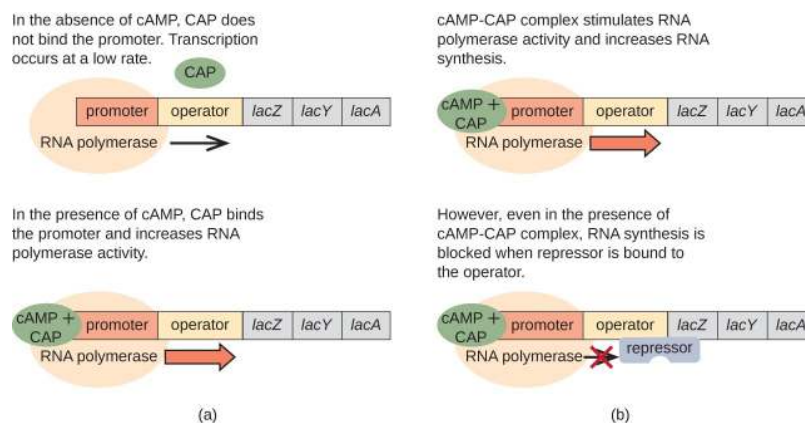


Figure [Math Processing Error]: Effect of CAP on the *Lac* Operon. (a) In the presence of cAMP, CAP binds to the promoters of operons, like the *lac* operon, that encode genes for enzymes for the use of alternative substrates. (b) For the *lac* operon to be expressed, there must be activation by cAMP-CAP as well as the removal of the lac repressor from the operator. [Parker, N., et. al. \(2019\) Microbiology. Openstax](#)

Figure [Math Processing Error] shows an [interactive iCn3D model](#) of the Catabolite activator protein CAP-DNA complex with bound cAMP (2CGP).

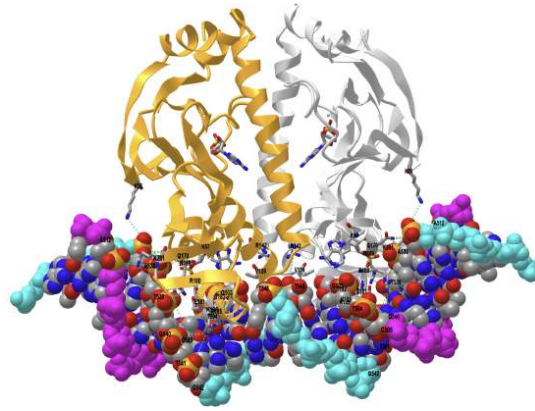


Figure [Math Processing Error]: Catabolite activator protein CAP-DNA complex with bound cAMP (2CGP). (Copyright; author via source). Click the image for a popup or use this external link: <https://structure.ncbi.nlm.nih.gov/1...z6mLhBpPmPMsC6>

27.1.5: Global Responses of Prokaryotes

In prokaryotes, several higher levels of gene regulation can control the transcription of many related operons simultaneously in response to an environmental signal. A group of operons all controlled simultaneously is called a *regulon*.

27.1.5.1: Alarmones

When sensing impending stress, prokaryotes alter the expression of a wide range of *operons* to respond in a coordinated manner. They do this through the production of *alarmones*, which are small intracellular nucleotide derivatives, such as guanosine pentaphosphate (pppGpp), as shown in Figure [Math Processing Error].

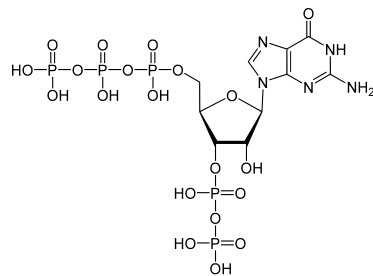


Figure [Math Processing Error]: Structure of Guanosine Pentaphosphate (pppGpp). [Yikrazuul](#)

Alarmones change which genes are expressed and stimulate the expression of specific stress-response genes. For example, pppGpp signaling is involved in the stringent response in bacteria, causing the inhibition of RNA synthesis when there is a shortage of amino acids present. This causes translation to decrease, and the amino acids present are therefore conserved. Furthermore, pppGpp causes the upregulation of many other genes involved in stress response, such as those for amino acid uptake from the surrounding media and biosynthesis.

The use of alarmones to alter gene expression in response to stress appears to be important in pathogenic bacteria, as well. On encountering host defense mechanisms and other harsh conditions during infection, many operons encoding virulence genes are upregulated in response to alarmone signaling. Knowledge of these responses is crucial for understanding the infection process of many pathogens and for developing therapies to counteract this process.

27.1.5.2: Quorum Sensing

Quorum sensing (QS) is an intercellular communication mechanism of bacteria used to coordinate the activities of individual cells at the population level in response to surroundings through the production and perception of diffusible signal molecules such as Acyl Homoserine Lactones or small signaling peptides, as shown in Figure [Math Processing Error]. The signal synthase, signal receptor, and signal molecules are three essential elements of the basic QS circuit machinery. Genes encoding signal-generating proteins are also included among the QS target genes. This forms an autoinduction feedback loop to modulate the generation of signal molecules. Several bacterial behaviors, including the expression of virulence factors, the production of secondary metabolites, biofilm formation, motility, and luminescence, are regulated by quorum sensing (QS). Through complex regulatory networks, bacteria are capable of expressing corresponding genes according to their population size and of behaving in a coordinated manner.

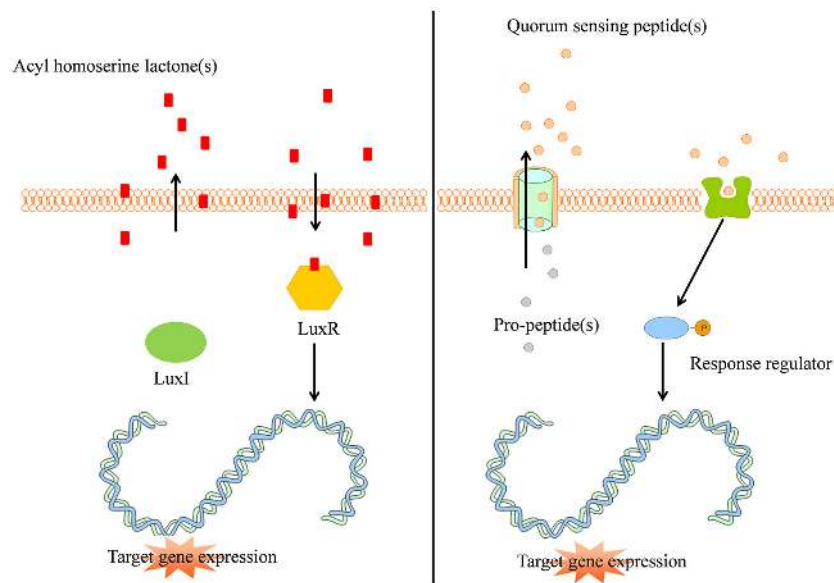


Figure [Math Processing Error]: Examples of Quorum Sensing Pathways. [Verbeke, F., et.al. \(2017\) Frontiers in Neuroscience 11:183.](#)

The left panel shows the typical Gram-negative quorum-sensing mechanism. Acyl homoserine lactone molecules, synthesized by LuxI, passively pass the bacterial cell membrane and when a sufficient concentration is reached (threshold level) activate the intracellular LuxR which subsequently activates target gene expression in a coordinated way. Note that a single cell is shown for simplicity. However, acyl homoserine lactones will commonly diffuse and target neighboring cells within the colony to mediate a communal or population response within the bacterial colony.

The right panel illustrates that quorum-sensing peptides are synthesized by bacterial ribosomes as propeptide proteins and undergo post-translational modifications during excretion via active transport. The quorum-sensing peptides bind to membrane-associated receptors, which become autophosphorylated and activate intracellular response regulators through phosphotransfer. These phosphorylated response regulators induce increased expression of their target genes.

For example, some microbial species, such as *Staphylococcus aureus*, can encase their community within a self-produced matrix of hydrated extracellular polymeric substances that include polysaccharides, proteins, nucleic acids, and lipid molecules. These encasements are known as biofilms. The formation of the *biofilm* on solid surfaces is a step-wise process comprising several stages, as shown in Figure [Math Processing Error]. It begins with the conditioning of the surface through coating with macromolecules from the surrounding aqueous environment, which enables the initial reversible adhesion of microorganisms. The next step involves the formation of stronger, irreversible attachments to the surface, followed by the proliferation and aggregation of microorganisms into multicellular and multilayered clusters that actively produce an extracellular matrix. Some cells in the mature biofilms continuously detach and separate from the aggregates, representing a continuous source of planktonic bacteria that can subsequently spread and form new microcolonies.

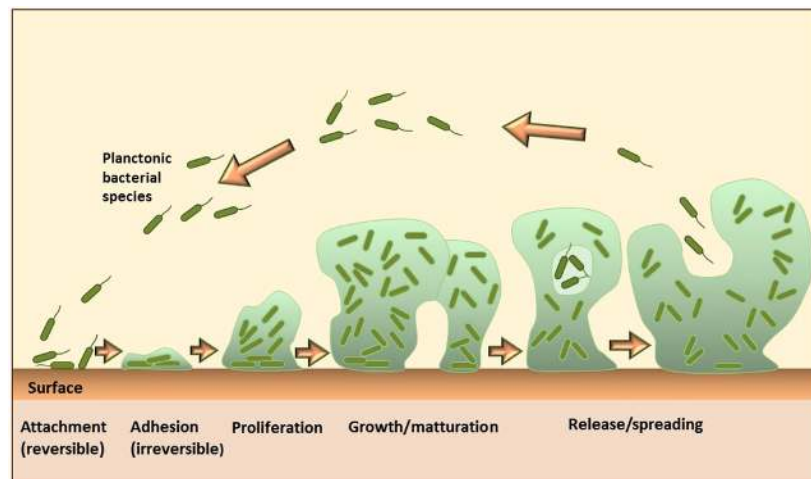


Figure [Math Processing Error] Schematic drawing of biofilm formation. [Rukaving, Z., and Vanic, Z. \(2016\) *Pharmaceutics* 8\(2\):18.](#)

Biofilms are a common cause of chronic, nosocomial (originating in a hospital), and medical device-related infections because they can develop on both vital and necrotic tissue, as well as on the inert surfaces of various implanted materials. Moreover, biofilms are linked with high-level resistance to antimicrobials, frequent treatment failures, and increased morbidity and mortality. As a consequence, biofilm infections and accompanying diseases have become a major health concern and a serious challenge for both modern medicine and pharmacy. A rough estimate suggests that more than 60% of hospital-associated infections are attributable to biofilms formed on indwelling medical devices, resulting in over one million cases of infected patients annually and exceeding \$1 billion in hospitalization costs per year in the USA.

Biofilm infections share several common characteristics: slow development in one or more focal points, delayed clinical manifestation, and persistence for months or years, typically with alternating periods of acute exacerbations and periods of clinical symptom absence. Although they are less aggressive than acute infections, their treatment is more challenging. There is up to a 1000-fold decrease in the susceptibility of biofilms to antimicrobial agents and disinfectants, as well as resistance to the host immune response. Thus, ways to reduce or inhibit biofilm formation are highly sought. The majority of the proposed biofilm-control methods focus on: (i) prevention and minimization of biofilm formation by selection and surface modifications of anti-adhesive materials; (ii) debridement techniques including ultrasound and surgical procedures; (iii) disruption of biofilm QS-signaling system; or (iv) achieving proper drug penetration and delivery to formed biofilms by the use of an electromagnetic field, ultrasound waves, photodynamic activation or specific drug delivery systems.

27.1.5.3: Alternate σ Factors

Since the σ subunit of bacterial RNA polymerase confers specificity on which promoters should be transcribed, altering the σ factor used is another way for bacteria to quickly and globally change which regulons are transcribed at a given time. The σ factor recognizes sequences within a bacterial promoter, so different σ factors will each recognize slightly different promoter sequences. In this way, when the cell senses specific environmental conditions, it may respond by changing which σ factor it expresses, degrading the old one, and producing a new one to transcribe the operons encoding genes whose products will be useful under the new environmental condition. For example, in sporulating bacteria of the genera *Bacillus* and *Clostridium* (which include many pathogens), a group of σ factors regulates the expression of the numerous genes required for sporulation in response to sporulation-stimulating signals.

27.1.6: Prokaryotic Attenuation and Riboswitches

Although most gene expression in prokaryotes is regulated at the level of transcription initiation, there are also mechanisms that control both the completion of transcription and translation concurrently. Since their discovery, these mechanisms have been shown to control the completion of transcription and translation of many prokaryotic operons. Because these mechanisms link the regulation of transcription and translation directly, they are specific to prokaryotes, because these processes are physically separated in eukaryotes.

One such regulatory system is attenuation, whereby secondary stem-loop structures formed within the 5' end of an mRNA, as it is being transcribed, determine whether transcription to complete the synthesis of this mRNA will occur and whether this mRNA will

be used for translation. Beyond the transcriptional repression mechanism already discussed, attenuation also controls the expression of the *trp* operon in *E. coli* as shown in Figure [Math Processing Error]. The *trp* operon regulatory region contains a leader sequence called *trpL* between the operator and the first structural gene, which has four stretches of RNA that can base pair with each other in different combinations. When a terminator stem-loop forms, transcription terminates, releasing RNA polymerase from the mRNA. However, when an antiterminator stem-loop forms, it prevents the formation of the terminator stem-loop, allowing RNA polymerase to transcribe the structural genes.

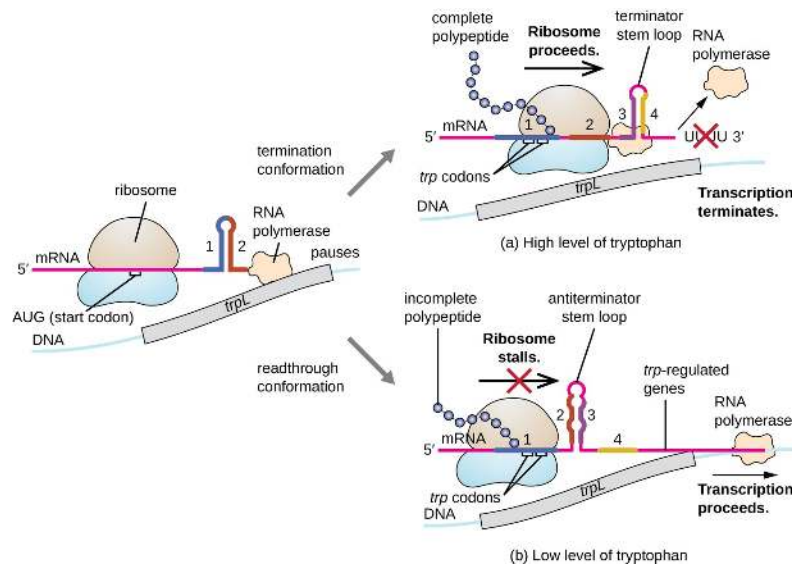


Figure [Math Processing Error]: Attenuation of Transcription and Translation. [Parker, N., et. al. \(2019\) Microbiology. Openstax](#)

When tryptophan is plentiful, translation of the short leader peptide encoded by *trpL* proceeds, the terminator loop between regions 3 and 4 forms, and transcription terminates. When tryptophan levels are depleted, translation of the short leader peptide stalls at region 1, allowing regions 2 and 3 to form an antiterminator loop, which enables RNA polymerase to transcribe the structural genes of the *trp* operon.

A related mechanism of concurrent regulation of transcription and translation in prokaryotes is the use of a riboswitch, a small region of noncoding RNA found within the 5' end of some prokaryotic mRNA molecules, as shown in Figure [Math Processing Error]. A riboswitch may bind to a small intracellular molecule to stabilize certain secondary structures of the mRNA molecule. The binding of the small molecule determines which stem-loop structure forms, thus influencing the completion of mRNA synthesis and protein synthesis.

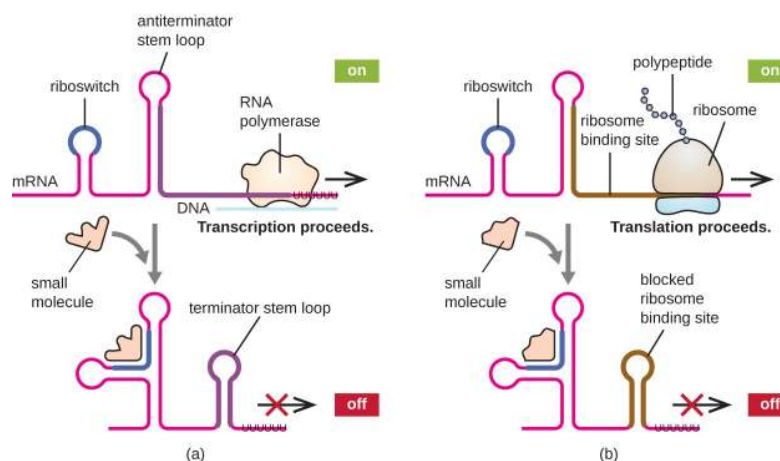
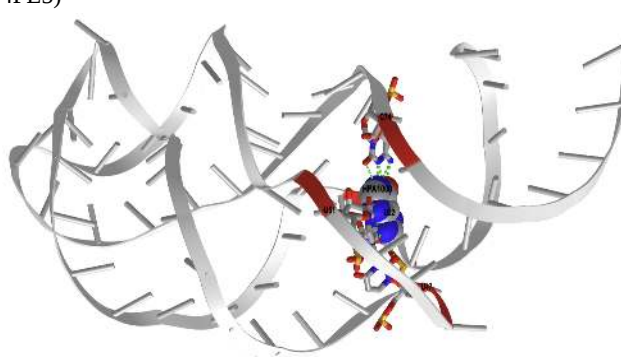
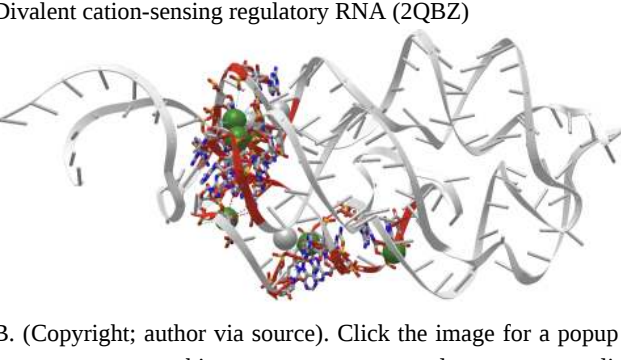
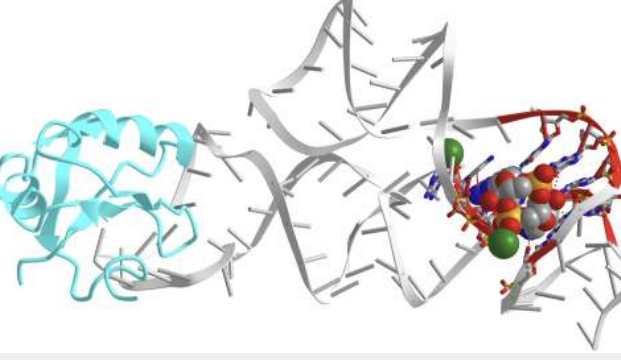
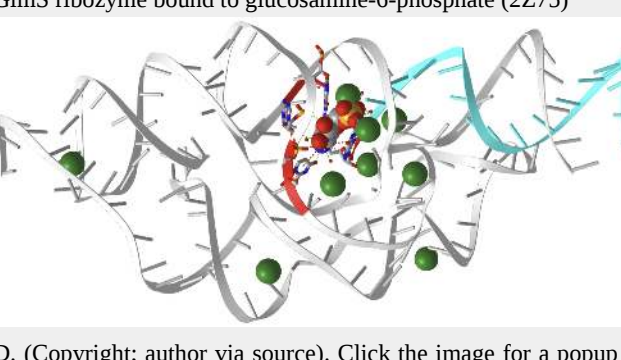


Figure [Math Processing Error]: Riboswitch Form and function. Riboswitches found within prokaryotic mRNA molecules can bind to small intracellular molecules, stabilizing certain RNA structures, and influencing either the completion of the synthesis of the mRNA molecule itself (left) or the protein made using that mRNA (right). [Parker, N., et. al. \(2019\) Microbiology. Openstax](#)

Riboswitches found within prokaryotic mRNA molecules can bind to small intracellular molecules, stabilizing specific RNA structures and influencing either the completion of the mRNA synthesis process itself (left) or the protein produced using that

mRNA (right).

Figure [Math Processing Error] shows [interactive iCn3D models](#) of a series of bacterial riboswitches. They are described in the legend below.

Guanine-responsive riboswitch bound to metabolite hypoxanthine (4FE5)  A. (Copyright; author via source). Click the image for a popup or use this external link: https://structure.ncbi.nlm.nih.gov/i...Nh1Sv7oSSyC6Z6	Divalent cation-sensing regulatory RNA (2QBZ)  B. (Copyright; author via source). Click the image for a popup or use this external link: https://structure.ncbi.nlm.nih.gov/i...VbU9Vf6DUA4VE8
Cyclid-di-GMP RNA riboswitch (3IRW)  C. (Copyright; author via source). Click the image for a popup or use this external link: https://structure.ncbi.nlm.nih.gov/i...wVntbdScP3VCF8	GlmS ribozyme bound to glucosamine-6-phosphate (2Z75)  D. (Copyright; author via source). Click the image for a popup or use this external link: https://structure.ncbi.nlm.nih.gov/i...w9T4K9ffK1NU27

A: Guanine-responsive riboswitch bound to metabolite hypoxanthine (4FE5) - Hypoxanthine, involved in purine metabolism, is shown bound to RNA representing the 5' untranslated region of the xanthine phosphoribosyltransferase (xnt)/ xanthine-specific purine permease (pbux) genes that lead to transcription termination.

B: The M-box in mycobacterial genes regulating Mg^{2+} transport binds divalent cations. They are transcribed under low Mg^{2+} concentrations. Salt bridges (ion-ion interactions) are shown in cyan, and pi-cation interactions in red dotted lines

C: Bis-(3'-5')-cyclic dimeric guanosine monophosphate (c-di-GMP) is a second messenger in bacteria and regulates many cellular processes, including the formation of biofilms. The riboswitch shown here is from *Vibrio cholerae*. The U1 small nuclear ribonucleoprotein A is shown in cyan. Figure [Math Processing Error] shows a cartoon of the actual riboswitch in the 5' untranslated region of target genes

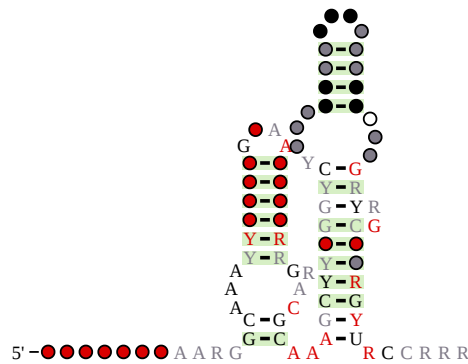


Figure [Math Processing Error]" Consensus secondary structure and sequence conservation of Cyclic di-GMP-I riboswitch

D. This ribozyme is in the 5' untranslated region of glucosamine-6-phosphate synthase mRNA. The protein enzyme, 2. This protein enzyme catalyzes the conversion of fructose 6-phosphate and glutamine to glucosamine 6-phosphate (GlcN6P) and glutamate. The *glmS* ribozyme in the 5'-untranslated region cleaves itself on binding GlcN6P. This self-cleavage is inhibited by glucose 6-phosphate (Glc6P). Hence, high levels of the gene product for the synthase lead to cleavage of its own mRNA. The *glmS* ribozyme RNA is shown in gray, and its substrate RNA is shown in cyan.

27.1.7: Summary

This chapter provides an in-depth exploration of the mechanisms by which cells regulate gene expression, addressing how genetically identical cells can exhibit distinct phenotypes. The material highlights the importance of both local (operon-level) and global regulatory strategies, integrating insights from prokaryotic and eukaryotic systems.

Key Concepts and Themes

- **Foundations of Gene Regulation:**

Although every cell in an organism contains the same DNA, the selective expression of only a subset of genes at any given time enables cells to differentiate and respond to environmental cues. This regulation is crucial for both multicellular development and the efficient use of resources in single-celled organisms.

- **Operon Organization in Prokaryotes:**

In bacteria, genes encoding proteins with related functions are often organized into operons—clusters of structural genes controlled by a single promoter and regulatory region. This organization allows for the coordinated expression of entire biochemical pathways.

- **Repressible Operons:** The *trp* operon serves as a model where gene expression is turned off when tryptophan levels are high, as tryptophan binds to the *trp* repressor, enabling it to block transcription.
- **Inducible Operons:** Conversely, the *lac* operon is activated only in the presence of lactose (and in the absence of glucose), with the *lac* repressor preventing transcription in the absence of the inducer. The CAP-cAMP complex further enhances transcription when energy levels are low.

- **Regulatory Elements and Transcription Factors:**

Regulatory regions, including promoters and operators, serve as binding sites for transcription factors such as repressors and activators. These factors modulate the ability of RNA polymerase to initiate transcription, either by hindering its binding or by facilitating its recruitment.

- **Global Regulatory Mechanisms:**

Beyond the control of individual operons, bacteria coordinate the expression of multiple genes and operons through regulons. Key regulatory molecules, such as alarmones (e.g., pppGpp), enable cells to adjust gene expression in response to stress (the stringent response). Additionally, quorum sensing allows bacteria to communicate via diffusible signals, coordinating collective behaviors like biofilm formation and virulence.

- **Alternative σ Factors:**

Bacterial RNA polymerase employs different σ subunits to recognize distinct promoter sequences. By switching σ factors, cells can rapidly reprogram gene expression in response to environmental changes, such as during sporulation or stress responses.

• Post-Initiation Regulation – Attenuation and Riboswitches:

Some operons are further regulated by mechanisms that modulate transcription elongation. Attenuation, exemplified by the *trp* operon, uses alternative RNA secondary structures in a leader sequence to either terminate or allow transcription. Riboswitches, on the other hand, are small regulatory RNA elements that bind specific metabolites; their conformational changes can affect either transcription termination or translation initiation.

Conclusion

Overall, this chapter underscores that gene regulation is a multifaceted process critical for cellular function and adaptation. By coordinating the expression of structural and regulatory genes through operon organization, transcription factor binding, global signaling molecules, and RNA-based regulatory mechanisms, cells can finely tune their proteomic output in response to internal and external signals. Understanding these processes is fundamental not only to basic biology but also to addressing health-related issues such as cancer, infectious diseases, and antibiotic resistance.

27.1.8: References

1. Parker, N., Schneegurt, M., Thi Tu, A-H., Lister, P., Forster, B.M. (2019) Microbiology. *Openstax*. Available at: <https://opentextbc.ca/microbiologyopenstax/>
2. Chan, K-G., Liu, Y-C., and Chang C-Y. (2015) Inhibiting N-acyl-homoserine lactone synthesis and quenching *Pseudomonas* quinolone quorum sensing to attenuate virulence. *Front. Microbiol.* 6:1173. Available at: <https://www.frontiersin.org/articles/10.3389/fmicb.2015.01173/full>
3. Rukavina, Z., and Vanic Zeljka. (2016) Current trends in development of liposomes targeting bacterial biofilms. *Pharmaceutics* 8(2):18. Available at: <https://www.mdpi.com/1999-4923/8/2/18/html>
4. Wikipedia contributors. (2020, April 18). Guanosine pentaphosphate. In *Wikipedia, The Free Encyclopedia*. Retrieved 16:26, August 23, 2020, from https://en.Wikipedia.org/w/index.php?title=Guanosine_pentaphosphate&oldid=951778776
5. Verbeke, F., De Craemer, S., Debunne, N., Janssens, Y., Wynendaele, E., Van de Wiele, C., and De Spiegeleer, B. (2017) Peptides as quorum sensing molecules: measurement techniques and obtained levels *in vitro* and *in vivo*. *Frontiers in Neuroscience* 11:183. Available at: https://www.researchgate.net/publication/316055402_Peptides_as_Quorum_Sensing_Molecules_Measurement_Techniques_and_Obtained_Levels_In_vitro_and_In_vivo
6. Yesudhas, D., Batool, M., Anwar, M.A., Panneerselvam, S., and Choi, S. (2017) Proteins recognizing DNA: Structural uniqueness and versatility of DNA-binding domains in Stem Cell Transcription Factors. *Genes* 8(8):192. Available at: <https://www.mdpi.com/2073-4425/8/8/192/html>
7. Castellanos, M., Mothi, N., and Muñoz, V. (2020) Eukaryotic transcription factors can track and control their target genes using DNA antennas. *Nature Comm.* 11:540. Available at: <https://www.nature.com/articles/s41467-019-14217-8>
8. Neideracher, G., Klopff, E., and Schüller, C. (2011) Interplay of dynamic transcription and chromatin remodeling: Lessons from yeast. *Int J Mol Sci* 12(8):4758-4769. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3179130/>
9. Kim, S., and Kaang, B-K. (2017) Epigenetic regulation and chromatin remodeling in learning and memory. *Exp. & Mol. Med.* 49:e281. Available at: <https://www.nature.com/articles/emm2016140#Fig1>
10. Tvardovskiy, A., Schwämmle, V., Kempf, S., Rogowska-Wrzesinka, A., and Jensen, O.N. (2016) Accumulation of histone variant H3.3 with age is associated with profound changes in the histone methylation landscape. *Nuc. Acids Res.* 45(16):1093. Available at: https://www.researchgate.net/publication/318987684_Accumulation_of_histone_variant_H33_with_age_is_associated_with_profound_changes_in_the_histone_methylation_landscape
11. Cipolletti, M., Fernandez, V.S., Montalesi, E., Marino, M., Fiochetti, M. (2018) Beyond the antioxidant activity of dietary polyphenols in cancer: The modulation of estrogen receptors (ERs) signaling. *Int J. Mol Sci* 19(9):2624. Available at: <https://www.mdpi.com/1422-0067/19/9/2624/html>
12. Griekspoor, A., Zward, W., Neeffjes, J., and Michalides, R. (2007) Visualizing the action of steroid hormone receptors in living cells. *Nucl. Recept. Signal.* 5:e003 Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1853070/>
13. Mitsis, T., Papageorgiou, L., Efthimiadou, A., Bacopoulou, F., Vlachakis, D., Chrousos, G.P., Eliopoulos, E. (2020) A comprehensive structural and functional analysis of the ligand binding domain of the nuclear receptor superfamily reveals highly conserved signaling motifs and two distinct canonical forms through evolution. *World Acad Sci J* 1: 264-274, 2019. Available at: <https://www.spandidos-publications.com/10.3892/wasj.2020.30>
14. Reed, S.M., and Quelle, D.E. (2015) p53 Acetylation: Regulation and consequences. *Cancers* 7(1):30-69. Available at: <https://www.mdpi.com/2072-6694/7/1/30/html>.

15. MacLaine, N.J., and Hupp, T.R. (2009) The regulation of p53 by phosphorylation: a model for how distinct signals integrate into the p53 pathway. *Aging* 1(5):490-502. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2806026/>
16. Wikipedia contributors. (2020, August 1). Estrogen. In *Wikipedia, The Free Encyclopedia*. Retrieved 01:28, September 6, 2020, from <https://en.Wikipedia.org/w/index.php?title=Estrogen&oldid=970560042>
17. Kluska, K., Adamczyk, J., and Krezel, A. (2018) Metal binding properties, stability, and reactivity of zinc fingers. *Coord. Chem Rev.* 367:18-64. Available at: <https://www.sciencedirect.com/science/article/pii/S0010854517305441>
18. Wikipedia contributors. (2020, July 4). Leucine zipper. In *Wikipedia, The Free Encyclopedia*. Retrieved 07:00, September 7, 2020, from https://en.Wikipedia.org/w/index.php?title=Leucine_zipper&oldid=965962667
19. Wikipedia contributors. (2020, April 15). Zinc finger. In *Wikipedia, The Free Encyclopedia*. Retrieved 18:28, September 7, 2020, from https://en.Wikipedia.org/w/index.php?title=Zinc_finger&oldid=951116840
20. Horsthemke, B. (2018) A critical view on transgenerational epigenetic inheritance in humans. *Nat. Comm.* 9:2973. Available at: <https://www.nature.com/articles/s41467-018-05445-5#rightslink>

This page titled **27.1: Regulation of Gene Expression in Bacteria** is shared under a **not declared** license and was authored, remixed, and/or curated by **Henry Jakubowski and Patricia Flatt**.

- **Current page** by **Henry Jakubowski and Patricia Flatt** has no license indicated.
- **11.2: Diffusion Across a Membrane - Passive and Facilitated Diffusion** by **Henry Jakubowski and Patricia Flatt** has no license indicated.

27.2: Regulation of Gene Expression in Eukaryotes

Learning Goals (ChatGPT o3-mini)

- **Understand Transcription Initiation and Enhancer Function:**
 - Explain how transcription initiation requires the coordinated assembly of multiple transcription factors (TFs) at promoter regions.
 - Describe the role of distant enhancers and the looping of DNA in stabilizing RNA polymerase binding and enhancing transcription initiation.
- **Comprehend Transcription Factor Activation:**
 - Detail how TFs are maintained in inactive states and become activated by cellular signals via post-translational modifications or allosteric changes (e.g., via cell cycle signals or steroid hormones).
 - Evaluate the pleiotropic effects of TF activation on downstream gene expression and cellular physiology.
- **Analyze the Role of p53 in Gene Regulation and Cell Cycle Control:**
 - Describe p53's functions as a transcription factor and its role as the “guardian of the genome” by mediating cell cycle arrest, apoptosis, and DNA repair.
 - Understand how post-translational modifications (phosphorylation, acetylation, ubiquitination, etc.) regulate p53 stability, tetramerization, nuclear localization, and target gene specificity.
- **Examine Steroid Hormone Receptor Mechanisms:**
 - Illustrate the modular structure of steroid hormone receptors (SHRs) including the DNA binding domain (DBD) and ligand binding domain (LBD).
 - Explain how ligand binding induces receptor dimerization, nuclear translocation, and binding to Hormone Response Elements (HREs) to regulate gene transcription.
 - Differentiate between receptor agonists, partial antagonists (SERMs), and full antagonists (SERDs), and their impact on receptor conformation and function.
- **Explore Mechanisms of TF Target Search and DNA Binding:**
 - Describe the multiple mechanisms (sliding, hopping, intersegmental transfer) by which transcription factors locate their specific DNA binding sites amidst a background of nonspecific interactions.
 - Analyze how clusters of degenerate binding sites (e.g., short tandem repeats) surrounding target elements can increase effective binding affinity and local concentration of TFs.
- **Investigate Chromatin Remodeling and Epigenetic Regulation:**
 - Explain the roles of histone post-translational modifications (acetylation, methylation) in dictating chromatin states (euchromatin vs. heterochromatin) and gene accessibility.
 - Discuss ATP-dependent chromatin remodeling complexes and their functions in nucleosome sliding, eviction, and histone variant exchange, and how these processes impact transcriptional regulation.
- **Assess the Concepts and Evidence of Transgenerational Epigenetic Inheritance:**
 - Define epigenetic mechanisms (e.g., DNA methylation, histone modifications, non-coding RNAs) that contribute to cellular memory and the maintenance of gene expression states across cell divisions.
 - Evaluate the experimental strategies and challenges involved in distinguishing epigenetic inheritance from genetic, ecological, and cultural inheritance.
 - Critically analyze case studies (e.g., effects of environmental stress on sperm RNA and DNA methylation, fetal programming) that suggest the potential for epigenetic marks to be transmitted across generations.

These learning goals provide a framework to deepen your understanding of the complex layers of gene regulation—from transcription factor activation and chromatin remodeling to the potential for epigenetic changes to influence inheritance and cellular phenotypes.

As seen in Chapter 26, the initiation of transcription requires the assembly of a multitude of transcription factors (TF) localized at the promoter region. Transcription can also utilize far-reaching interactions of enhancers, which bind at a distant DNA site and loop back around to stabilize the RNA polymerase at the promoter. Control of transcriptional initiation depends on the activation of TF factors, their binding to specific DNA recognition sequences, and chromatin remodeling.

27.2.1: Transcription Factor (TF) Activation

Many transcription factors (TFs) are expressed within cells and remain in an inactive conformation until the proper environmental stimulus is present. Cellular signaling pathways can cause post-translational protein modifications, leading to transcription factor (TF) activation, or small molecules may physically bind and allosterically modify the protein structure to mediate activation. Here, we will use examples from the cell cycle signaling cascade and steroid hormone receptor pathways to highlight some mechanisms of transcription factor (TF) activation. A key takeaway from this section is that transcription factor activation is often highly pleiotropic, exhibiting numerous cellular effects. Depending on the cell type and the environmental conditions, different combinations of downstream target genes may be activated or inactivated. Teasing apart these intricacies and the physiological effects that they have within an organism is a major goal of ongoing research.

27.2.1.1: Cell Cycle Regulation by p53

p53 is one of the most studied proteins in science. To date, over 68,000 papers appear in PubMed containing p53 or TP53 in the title and/or abstract. Originally described as an oncogene (since a mutated, functionally altered form of the protein was first characterized), p53 is now recognized as the most frequently inactivated tumor suppressor in human cancers. It is a transcription factor that controls the expression of genes and miRNAs affecting many important cellular processes, including proliferation, DNA repair, programmed cell death (apoptosis), autophagy, metabolism, and cell migration, as shown in Figure 27.2.1. Many of those processes are critical to a variety of human pathologies and conditions extending beyond cancer, including ischemia, neurodegenerative diseases, stem cell renewal, aging, and fertility. Notably, p53 also has non-transcriptional functions, ranging from intrinsic nuclease activity to activation of mitochondrial Bak (Bcl-2 homologous antagonist killer) and caspase-independent apoptosis.

As a transcription factor, p53 responds to various genotoxic insults and cellular stresses (e.g., DNA damage or oncogene activation) by inducing or repressing the expression of over a hundred different genes. p53 transcriptional regulation plays a dominant role in causing the arrest of damaged cells, facilitating their repair and survival, or inducing cell death when DNA damage is irreparable. p53 can also cause cells to become permanently growth arrested, and there is compelling *in vivo* evidence that these “senescent” cells secrete factors that enhance their clearance by the immune system, leading to tumor regression. Through these mechanisms, p53 helps maintain genomic stability within an organism, thereby justifying its long-held nickname, “the guardian of the genome.” Other p53 gene targets are involved in inhibiting tumor cell angiogenesis, migration, metastasis, and other important processes (such as metabolic reprogramming) that normally promote tumor formation and progression.

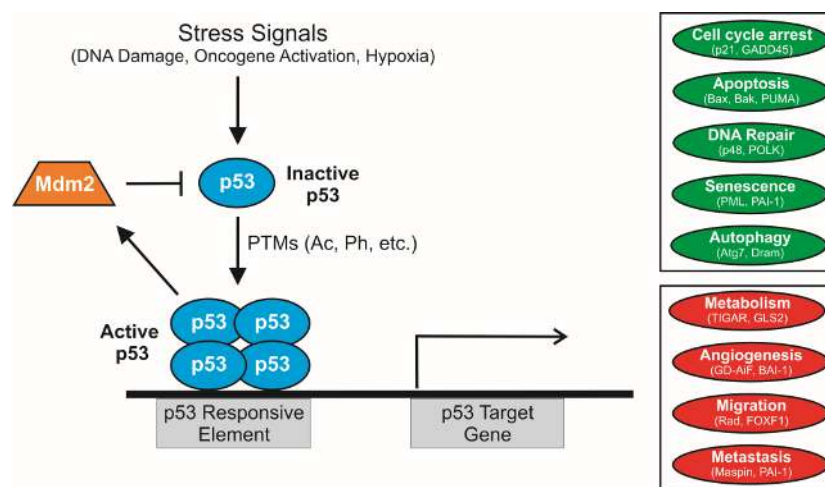


Figure 27.2.1: Cellular stress leads to p53 transcriptional activation of downstream targets. [Reed, S.M., and Quelle, D.E. \(2015\) 7\(1\):30-69.](#)

Figure 27.2.2 shows an [interactive iCn3D model](#) of the human p53 tetramer bound to the natural CDKN1A(p21) p53-response element (3TS8).

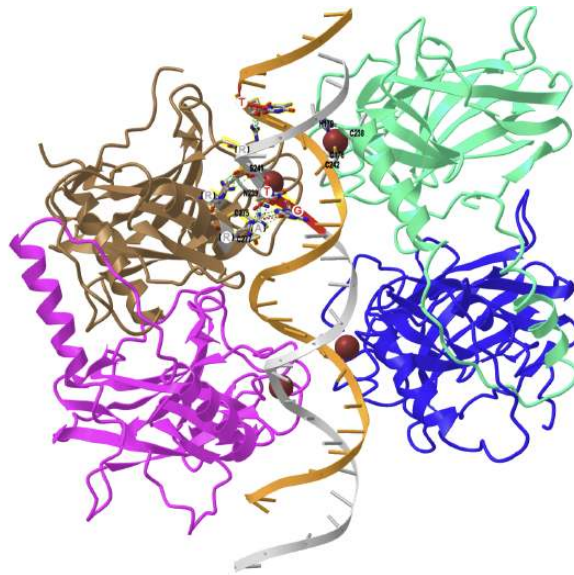


Figure 27.2.2: Human p53 tetramer bound to the natural CDKN1A(p21) p53-response element (3TS8). (Copyright; author via source). Click the image for a popup or use this external link: <https://structure.ncbi.nlm.nih.gov/...cvfiRBEzfxJuk6>

Each monomer of the p53 tetramer is shown in a different color. The noncovalent interactions of the brown monomer with DNA are illustrated, with key amino acids and nucleic acid bases depicted in CPK-colored sticks. They interact with the Zn^{2+} ions, shown in the light green monomer.

Typically, p53 levels are maintained at low levels by its major antagonist, Mdm2, an E3 ubiquitin ligase that is also a transcriptional target of p53. Stress signals, such as DNA damage, oncogene activation, and hypoxia, promote p53 stability and activity by inducing post-translational modifications (PTMs) and tetramerization of the p53 protein. p53 functions as a transcription factor that binds to specific p53 response elements upstream of its target genes. p53 affects many important cellular processes linked to tumor suppression, including the induction of senescence (green), apoptosis, and DNA repair, as well as the inhibition of metabolism (red), angiogenesis, and cell migration. These functions are primarily mediated through transcriptional regulation of their targets (examples given).

p53 protein function is regulated post-translationally by coordinated interaction with signaling proteins, including protein kinases, acetyltransferases, methyltransferases, and ubiquitin-like modifying enzymes, as shown in Figure 27.2.3. The majority of covalent modification sites occur at intrinsically unstructured linear peptide docking motifs that flank the DNA-binding domain of p53, which plays a role in anchoring or allosterically activating the enzymes that mediate the covalent modification of p53. In undamaged cells, the p53 protein has a relatively short half-life and is degraded by a ubiquitin-proteasome-dependent pathway through the action of E3 ubiquitin ligases, such as MDM2 (Fig 28.3.1). Following stress, p53 is phosphorylated at multiple residues, thereby modifying its biochemical functions and increasing its activity as a transcription factor. Post-translational modifications help stabilize the tetramer formation of the protein and enhance its translocation from the cytoplasm into the nucleus. The tetrameric form of p53 is then functional in binding to DNA in a sequence-specific manner and either activating or repressing transcription, depending on the target sequence. Some post-translational modifications, such as acetylation, are DNA-dependent and can play a role in chromatin remodeling and activation of p53 target gene expression.

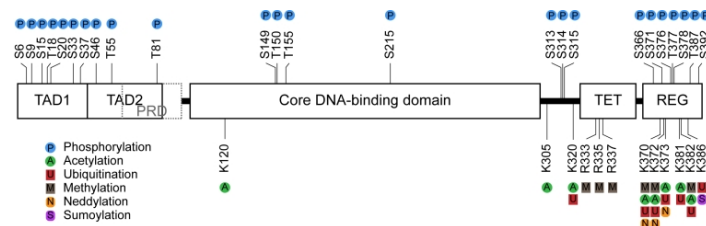


Figure 27.2.3: Sites of Post-Translational Modification on p53. Schematic representation of the 393 amino acid domain structure of human p53 showing the sites of post-translational modification including phosphorylation, acetylation, ubiquitination, methylation, neddylation, and sumoylation. Abbreviations: N-terminal transactivation domain (TAD); proline-rich domain (PRD); tetramerization domain (TET); C-terminal regulatory domain (REG); arginine (R); lysine (K); serine (S); threonine (T). [MacLaine, N.J., and Hupp T.R. \(2009\) Aging 1\(5\):490-502](#)

It should be noted that single-point mutations that modify the ability of the protein to be phosphorylated in one position typically do not show a decrease in the stabilization or activation of the protein following a damage or stress event. Thus, multiple modifications likely allow for redundancy within this pathway, ensuring the activation of the protein following a stress event. Furthermore, the environment within the cell can lead to different p53 phenotypes, such as the activation of growth arrest and DNA repair processes (ie if there is not a lot of damage) or it can lead to the activation of apoptosis or programmed cell death pathways (ie if the damage is too extensive to be repaired).

27.2.1.2: Steroid Hormone Receptors

Steroid hormone receptors (SHRs) belong to the superfamily of *nuclear receptors (NRs)*, which are one of the essential classes of transcriptional factors. NRs play a critical role in all aspects of human development, metabolism, and physiology. Since they generally act as ligand-activated transcription factors, they are an essential component of cell signaling. NRs form an ancient and conserved family that arose early in the metazoan lineage. Major events of gene duplication and gene loss characterize the molecular evolution of NR. Phylogenetic analysis revealed a distinct separation of NR ligand binding domains (LBDs) into four monophyletic branches, the steroid hormone receptor-like cluster, the thyroid hormone-like receptors cluster, the retinoid X-like and steroidogenic factor-like receptor cluster, and the nerve growth factor-like/HNF4 receptor cluster, as shown in Figure 27.2.4

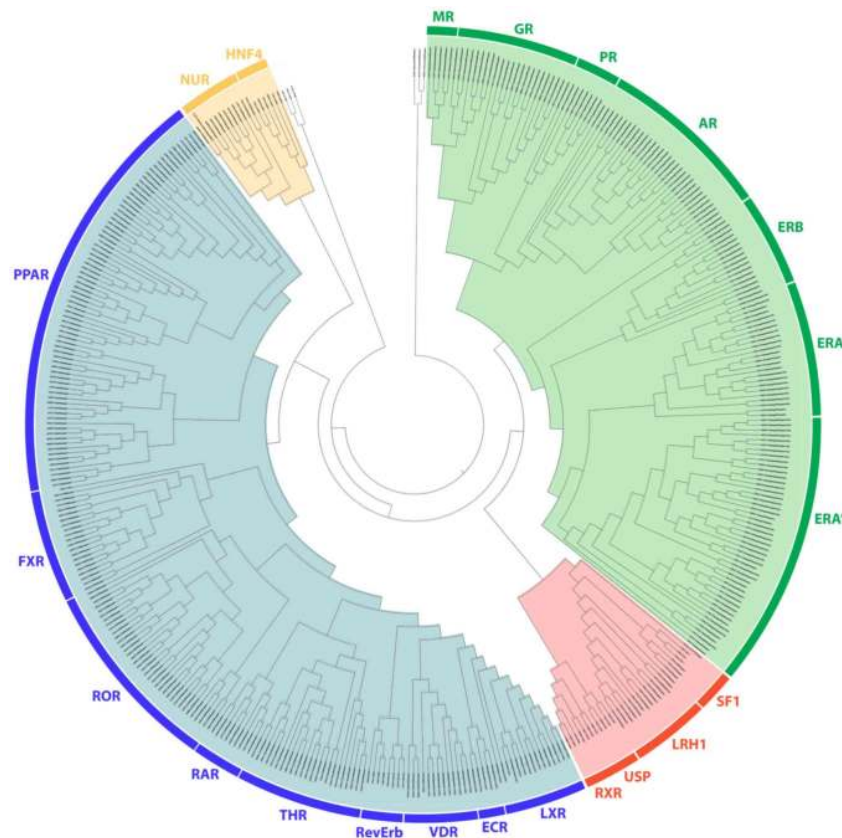


Figure 27.2.4: **Phylogenetic tree of the nuclear receptors' ligand binding domain.** Four distinct monophyletic branches are visible. Those monophyletic branches are divided into subcategories. The phylogenetic trees confidently separate the steroid hormone-like (branch colored green), the retinoid X-like and steroidogenic factor-like receptors cluster (branch colored orange), the thyroid hormone-like receptors cluster (branch colored blue), and the nerve growth factor-like/hepatocyte nuclear factor-4 receptors cluster (branch colored yellow). [Mitsis, T., et. al. \(2020\) World Acad Sci J 1: 264-274, 2019](#)

Here, we will focus on the Steroid Hormone-Like Receptors (SHRs) branch. SHRs play a crucial role in various important physiological processes, including organ development, metabolite homeostasis, and response to external stimuli. The estrogen receptor comes in two major forms, ER α and ER β . Other members of this subgroup include the cortisol-binding glucocorticoid receptor (GR), the aldosterone-binding mineralocorticoid receptor (MR), the progesterone receptor (PR), and the dihydrotestosterone (DHT) binding androgen receptor (AR), as shown in Figure \(\PageIndex{5}\) below.

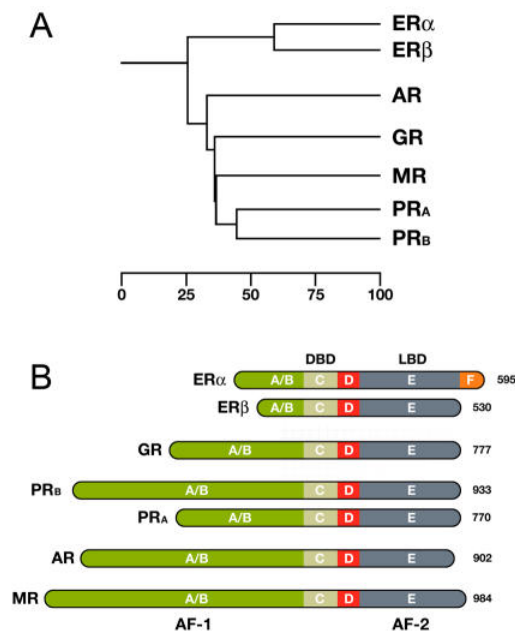


Figure 27.2.5: Overview of Steroid Hormone Receptor Family (SHR). [riekspoor, A., et. al. \(2007\) Nucl Recept Signal. 5:e003_A](#). Phylogenetic tree of the Steroid Hormone Receptor (SHR) family showing the evolutionary interrelationships and distance between the various receptors. Based on alignments available at The NucleaRDB [Horn et al., 2001]. B. All steroid receptors are composed of a variable N-terminal domain (A/B) containing the AF-1 transactivation region, a highly conserved DNA-binding domain (DBD), a flexible hinge region (D), and a C-terminal ligand-binding domain (LBD, E) containing the AF-2 transactivation region. The estrogen receptor α is unique in that it contains an additional C-terminal F domain. Numbers represent the length of the receptor in amino acids.

Panel A displays a phylogenetic tree of the Steroid Hormone Receptor (SHR) family, illustrating the evolutionary relationships and distances between the various receptors. Based on alignments available at The NucleaRDB [Horn et al., 2001].

Panel B shows that all steroid receptors are composed of a variable N-terminal domain (A/B) containing the AF-1 transactivation region, a highly conserved DNA-binding domain (DBD), a flexible hinge region (D), and a C-terminal ligand-binding domain (LBD, E) containing the AF-2 transactivation region. The estrogen receptor α is unique in that it contains an additional C-terminal F domain. Numbers represent the length of the receptor in amino acids.

The members of the Steroid Hormone Receptor family share a similar, modular architecture, consisting of several independent functional domains (Fig. 5B above). Most conserved is the centrally located DNA-binding domain (DBD), which contains the characteristic zinc-finger motifs. The DBD is followed by a flexible hinge region and a moderately conserved Ligand Binding Domain (LBD), located at the carboxy-terminal end of the receptor. The estrogen receptor α is unique in that it contains an additional F domain, the function of which is unclear. The LBD is composed of twelve α -helices (H1-H12) that together fold into a canonical α -helical sandwich. Besides its ligand-binding capability, the LBD also plays a crucial role in nuclear translocation, chaperone binding, receptor dimerization, and coregulator recruitment through its potent ligand-dependent transactivation domain, referred to as AF-2. A second, ligand-independent, transactivation domain is located in the more variable N-terminal part of the receptor, designated as AF-1. To date, no crystal structure of a full-length SHR exists, though structures of the DBD and LBD regions of most SHRs are available. These have helped significantly in understanding the molecular aspects of DNA and ligand binding, but have also, to some extent, led to biased attention to these parts of the receptor alone. For example, many coregulator interaction studies are still performed with the LBD only, while numerous studies have demonstrated that the AF-2 domain often tells only part of the story. With the help of biophysical techniques, however, it is feasible to study the full-length receptor in its native environment.

Most SHRs remain in the cytoplasm of the cell until they are bound with the appropriate steroid, as shown in Figure 27.2.6 Steroid binding causes the dimerization of SHRs and localization to the cell nucleus, where the SHRs interact with the DNA at sequence-specific motifs known as Hormone Response Elements (HREs) (Step 5). Many SHRs can also interact with membrane-bound receptors and affect cellular signaling pathways, in addition to activating gene expression (step 6).

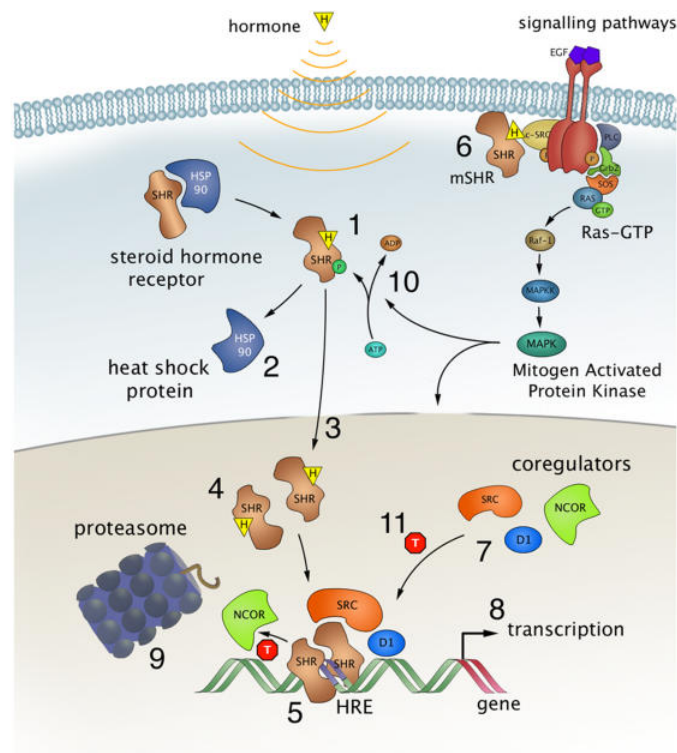


Figure 27.2.6: Steroid Hormone Receptors (SHR) act as hormone-dependent nuclear transcription factors. Upon entering the cell by passive diffusion, the hormone (H) binds the receptor, which is subsequently released from heat shock proteins, and translocates to the nucleus. There, the receptor dimerizes, binds specific sequences in the DNA, called Hormone Responsive Elements or HREs, and recruits a number of coregulators that facilitate gene transcription. [Griekspoor, A., et. al. \(2007\) Nucl Recept Signal. 5:e003](#)

Steroid hormones, such as estrogens, reach their target cells via the blood, where they are bound to carrier proteins. Naturally occurring estrogens, including estradiol, estrone, and estriol, differ primarily in structure in the presence of hydroxyl groups (Fig. 28.3.6). Estradiol is the predominant estrogen during reproductive years, both in terms of absolute serum levels as well as in terms of estrogenic activity. During menopause, estrone is the predominant circulating estrogen, and during pregnancy, estriol is the predominant circulating estrogen in terms of serum levels. Another type of estrogen called estetrol (E4) is also produced predominantly during pregnancy, as shown in Figure 27.2.7. Estrogens play a crucial role in various physiological processes, including the regulation of the menstrual cycle and reproduction, maintenance of bone density, brain function, cholesterol mobilization, and the maturation of reproductive organs during development, as well as in controlling inflammation.

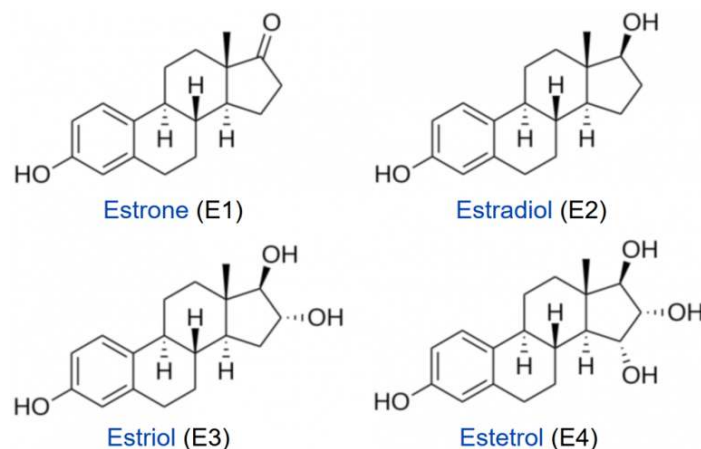


Figure 27.2.7: Naturally Occurring Estrogens. [Wikipedia \(2020\) Estrogen.](#)

Due to their lipophilic nature, it is believed that steroid hormones, such as estrogen, cross the cell membrane by simple diffusion. However, some evidence suggests that they can also be actively taken up through the endocytosis of carrier protein-bound

hormones. For a long time, it has been assumed that binding of the ligand resulted in a simple on/off switch of the receptor (Fig. 6, step 1). While this is likely the case for typical agonists like estrogen and progesterone, this is not always correct for receptor antagonists, used in drug therapy. These antagonists come in two types: so-called partial antagonists (for the estrogen receptors, known as SERMs, or Selective Estrogen Receptor Modulators) and full antagonists. The partial antagonist can, depending on the cell type, act as either an SHR agonist or antagonist. In contrast, full antagonists (for ER, known as SERDs, or Selective Estrogen Receptor Downregulators) always inhibit the receptor, independent of cell type, in part by targeting the receptor for degradation. Binding of either type of antagonist results in major conformational changes within the LBD and in the release from heat shock proteins that thus far had protected the unliganded receptor from unfolding and aggregation (Fig. 6 step 2).

Figure 27.2.8 shows an [interactive iCn3D model](#) of the Androgen Receptor DNA-Binding Domain Bound to a Direct Repeat Response Element (1R4I).

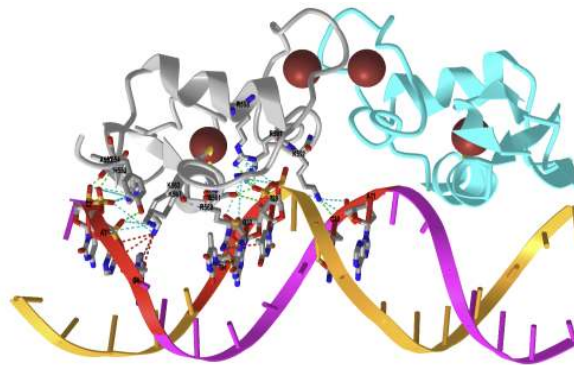


Figure 27.2.8: Androgen Receptor DNA-Binding Domain Bound to a Direct Repeat Response Element (1R4I). (Copyright; author via source). Click the image for a popup or use this external link: <https://structure.ncbi.nlm.nih.gov/i...FKLa7aaZiq6Hz8>

27.2.2: Transcription Factor (TF) Recognition and Binding to DNA

TF controls gene expression by binding to their target DNA site to recruit, or block, the transcription machinery onto the promoter region of the gene of interest. Their function relies on the ability to find their target site quickly and selectively. In living cells, transcription factors (TFs) are present in nM concentrations and bind the target site with comparable affinity. However, they also bind to any DNA sequence (nonspecific binding), resulting in millions of low-affinity (i.e., $> 10^{-6}$ M) competing sites. Nonspecific binding facilitates the search for the target site by three major mechanisms, as shown in Figure 27.2.9

The second scenario is a 'hopping' mechanism, in which a TF might hop from one site to another in 3D space by dissociating from its original site and subsequently binding to the new site. This may occur within the same chain, and re-association takes place adjacent to the former dissociated site. A third search mechanism is described as "intersegmental transfer." In this scenario, the protein moves between two sites via an intermediate 'loop' formed by the DNA and subsequently binds at two different DNA sites. This mechanism applies to transcription factors (TFs) with two DNA-binding sites. Proteins with two DNA-binding sites can occasionally bind non-specifically to two locations situated far apart within the DNA strand, which are brought into close contact through the formation of these loops. Such TFs transfer across a point of close contact without dissociating from the DNA.

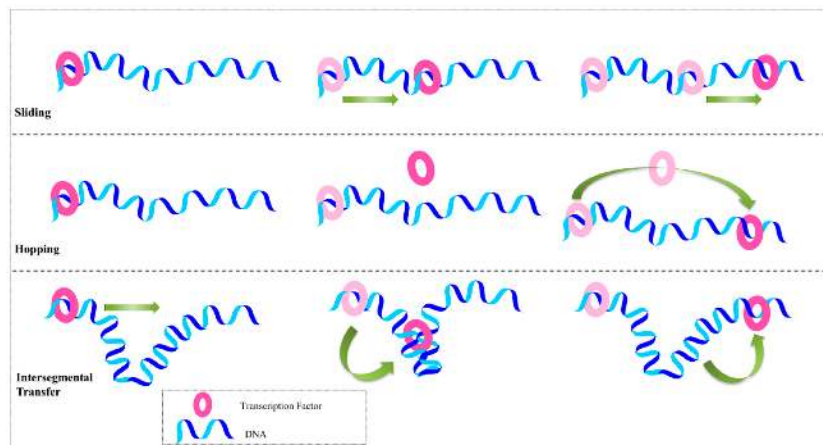


Figure 27.2.9: **Protein-DNA recognition mechanisms.** The main three protein-DNA recognition mechanisms are shown.
[Yesudhas, D., et.al. \(2017\) Genes 8\(8\):192](#)

Top: When the transcription factor (pink ring) moves from one site to another by sliding along the DNA and is transferred from one base pair to another without dissociating from the DNA, this mechanism is called sliding.

Center: Hopping occurs when the transcription factor moves along the DNA by dissociating from one site and re-associating with another site.

Bottom: Intersegmental transfer describes the mechanism by which the transcription factor gets transferred through DNA bending or the formation of a DNA loop, resulting in the protein being bound transiently to both sides and subsequently moving from one site to the other.

One of the main scenarios involves a 'sliding' mechanism, in which the protein moves from its initial non-specific site to its actual target site by sliding along the DNA (also known as 1-dimensional (1D) sliding). When the TF starts to move and shift counterions from the phosphate backbone, the same number of counterions bind to the site left free by the protein. The sliding rate is also dependent on the hydrodynamic radius of the protein; the required rotational movement over the DNA backbone is greater for larger proteins, which tend to slide slowly.

Recent Updates: 9/25/23

The sliding model that Von Hippel proposed and Berg suggests that a DNA-binding protein exists in two interconverting conformations. One is a specific form (O) that can bind to a target DNA sequence, such as an operator in DNA, through specific hydrogen bonds (along with electrostatic interaction) characterized by a low K_D . The other is a nonspecific form (D) that binds mainly with weaker affinity through electrostatic interactions and a high K_D . Nonspecific binding brings the protein to the DNA surface. Dynamic conformational changes from the O to D allow sampling of hydrogen bonds between donors and acceptors in the protein and in the major groove of the protein. The protein can diffuse much more quickly along the DNA to find its target site, as the search for the specific target site is now effectively one-dimensional instead of three-dimensional. There is no thermodynamic barrier to sliding since counterions that leave the DNA when bound to the protein rebind behind it as the protein slides. These processes are illustrated in Figure 27.2.10 below.

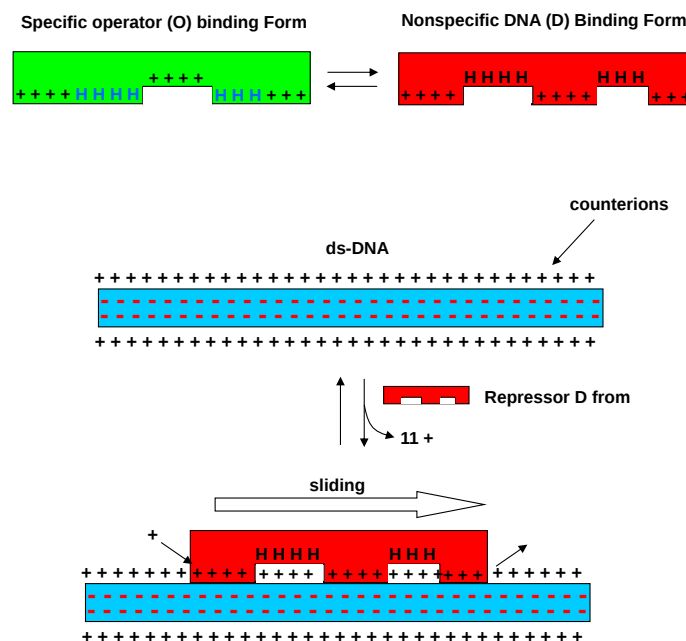


Figure 27.2.10 Two-state model sliding model of DNA protein binding to its specific binding site (operator)

There are a large number of overlapping nonspecific binding sites (let's say each is six base pairs in length), which also help drive the nonspecific binding of the protein to the DNA through entropy increases (i.e., consider the probability of binding to 1 site on the DNA versus multitudes of overlapping sites). Experiments show that the on-rate for a DNA-binding protein, which is the rate at which it finds its target site, increases with the increasing length of the DNA molecule in which the specific site is embedded. The opposite would be expected given the diffusion rate of large molecules. The k_{on} was found to be greater than diffusion-controlled limits, which can be explained by the reduced dimensionality of the search for the specific site when the protein is loosely bound through nonspecific electrostatic interactions that enable the 1D search.

Each eukaryotic transcription factor (TF) controls tens to hundreds of genes scattered throughout the genome, and expressing each gene requires various TFs simultaneously binding to their sites to form the transcription complex, an extremely rare event in probabilistic terms. As a result, the *in vivo* site occupancy patterns of eukaryotic TFs are more complex than predicted by their *in vitro* site-specific binding profiles and do not strongly correlate with the actual levels of gene expression. A notable feature highlighted by genome analysis is the accumulation of potential transcription factor (TF) binding sites in regions flanking eukaryotic genes. Such clusters of degenerate recognition sites are assumed to be key for transcription control and thus are generally classified as gene regulatory regions (RR). For example, the affinity of the *Drosophila* transcription factor Engrailed to the regulatory regions (RRs) of its target genes is strongly amplified by long tracts of degenerate consensus repeats that are present in these regions.

Role of Short Tandem Repeats (STRs) in the Genome

As we mentioned previously, only about 1.5% of the human genome encodes genes that produce actual proteins. Much of the genome is transcribed at low levels into RNAs, some of which have clearly defined functions, such as rRNA, mRNAs, and regulatory RNAs. A significant portion is presumably involved in facilitating the 3D organization of the genome and its dynamic architecture, which determines its replicative and transcriptional access. One poorly understood feature of genomic DNA is the presence of **short tandem repeats (STRs)**. These repeats stretch up to 100 nucleotides in length, with each repetitive tandem repeat running from just 1 to 6 bases long. They comprise approximately 6% of the genome (compared to 1.5% for protein-coding genes) and are found in abundance in chromatin that is transcribed into mRNA for protein production. For example, a specific DNA sequence might be GTCACGTGAC, while a small STR would be (CG)₆C(CG)₁₁.

Specifically, STRs surround sites where classic transcription factors (TF) bind. As described in this and the previous chapter sections, TFs bind to specific DNA binding motifs (such as helix-loop-helix or Zn²⁺ fingers) on consensus sequences (like response elements and enhancers) in the DNA, thereby controlling transcription. The binding of transcription factors or other proteins to target sequences occurs initially through nonspecific electrostatic interactions, followed by a dimensionally restricted diffusion along the DNA as the protein finds its specific sequence target.

In contrast, the STRs offer little sequence uniqueness for high-affinity, low K_D binding sites for specific protein interactions, so the question remains as to how they express function, which their omnipresence suggests they have. Studies by Horton et al (Science 381, 1304 (2023)) have shown that classic TFs do indeed bind STRs, albeit at lower affinity (higher K_D) compared to their binding to classic TF-specific sequences. They bind with higher affinity than to nonspecific DNA sequences. Just as nonspecific interactions facilitate TF binding to promoter sites, so do the multiple STRs that straddle the DNA binding element.

If a TF binds its isolated target DNA with a low effective K_D (high affinity) and $-\Delta G^0$ value, a target DNA surrounded by multiple STRs would have an even lower effective K_D (higher affinity) and even more $-\Delta G^0$ value. They do so by increasing the effective on rate (k_{on}) for protein binding. (Remember that $K_D = k_{off}/k_{on}$.) The effective size of the target for the TF becomes greater when it is an “island” in the middle of a STR “sea”. The increased affinity stems in large part from the more favorable entropy of having the TF bind not only to one site but effectively to multiple overlapping sites. This also increases the localized concentration of TF near the specific site, which drives binding. The k_{off} is not expected to change.

Another effect of the STRs on TF binding is that multiple DNA-binding proteins can bind to the same site through their interactions with STRs at the site, leading to new ways to regulate gene transcription. The group studies just two TFs, but sequence analyses suggest that many TF would use a similar mechanism.

27.2.3: Histone Modification and Chromatin Remodeling

Regulation of transcription involves dynamic rearrangements of chromatin structure. Recall that eukaryotic DNA is complexed with histone octamers, which are composed of dimers of the core histones H2A, H2B, H3, and H4. 147 bp of DNA are wrapped 1.65 times around each octamer, forming nucleosomes, the basic packaging units of chromatin. Nucleosomes, connected by linker DNA of variable length as “beads on a string”, generate the 11 nm linear structure. The linker histone H1 is positioned at the top of the core histone octamer, enabling the higher-organized compaction of DNA into transcriptionally inactive 30 nm fibers.

To understand the role of chromatin in transcription regulation, it is essential to know where nucleosomes are positioned and how this positioning is achieved. There are four groups of activities that change chromatin structure during transcription: (1) histone modifications, (2) eviction and repositioning of histones, (3) chromatin remodeling, and (4) histone variant exchange. Histone modifiers introduce post-translational, covalent modifications to histone tails and thereby change the contact between DNA and histones. These modifications govern access to regulatory factors. Histone chaperones aid in the eviction and positioning of histones. A third class of chromatin restructuring factors is ATP-dependent chromatin remodelers. These multi-subunit complexes utilize energy from ATP hydrolysis for various chromatin remodeling activities, including nucleosome sliding, nucleosome displacement, and the incorporation and exchange of histone variants.

Post-translational modifications (PTMs) of histone proteins are a primary mechanism that controls chromatin architecture. Over 20 distinct types of histone PTMs have been described, among which the most abundant ones are acetylation and methylation of lysine residues. Histone PTMs can be deposited on and removed from chromatin by different enzymes, known as histone PTM ‘writers’ and ‘erasers’. Histone post-translational modifications (PTMs) exert their regulatory effects through two primary mechanisms. First, histone PTMs serve as docking sites for various nuclear proteins—histone PTM ‘readers’—that specifically recognize modified histone residues through their modification-binding domains. The recruitment of these proteins to specific genomic loci promotes key chromatin processes, including transcriptional regulation and DNA damage repair. Second, some histone post-translational modifications (PTMs), such as acetylation, directly affect chromatin's higher-order structure and compaction, thereby controlling chromatin accessibility to protein machinery, including those involved in transcription. Chromatin may adopt one of two major states, switching between them interchangeably. These states are *heterochromatin* and *euchromatin*. *Heterochromatin* is a compact form that is resistant to the binding of various proteins, such as transcriptional machinery. In contrast, *euchromatin* is a relaxed form of chromatin that is open to modifications and transcriptional processes, as shown in Figure 27.2.10. Histone methylation promotes the formation of Heterochromatin, whereas histone acetylation promotes euchromatin.

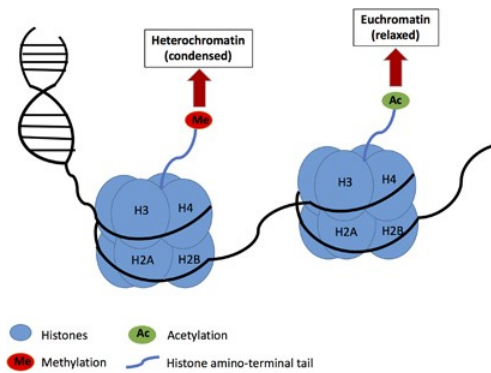


Figure 27.2.11: Schematic drawing of histone methylation and acetylation in relation to chromatin remodeling. [Kim, S., and Kaang, B-K. \(2017\) Exp & Mol. Med. 49:e281](#)

The addition of methyl groups to the tails of histone core proteins leads to histone methylation, which in turn leads to the adoption of a condensed state of chromatin called 'heterochromatin.' Heterochromatin blocks the transcription machinery from binding to DNA, resulting in transcriptional repression. The addition of acetyl groups to lysine residues in the N-terminal tails of histones causes histone acetylation, which leads to the adoption of a relaxed state of chromatin called 'euchromatin.' In this state, transcription factors and other proteins can bind to their specific DNA binding sites and initiate active transcription.

Chromatin remodeling can also be an ATP-dependent process and involve histone dimer ejection, full nucleosome ejection, nucleosome sliding, and histone variant exchange as shown in Figure 27.2.12 ATP-dependent chromatin remodeling complexes bind to nucleosome cores and the surrounding DNA, utilizing energy from ATP hydrolysis to disrupt DNA-histone interactions, slide or eject nucleosomes, alter nucleosome structures, and modulate the access of transcription factors to the DNA (Figure 28.3.9). In addition to modulating gene expression, some of the complexes are involved in nucleosome assembly and organization, following transcription at locations in which nucleosomes have been ejected, packing of DNA following replication, and DNA repair.

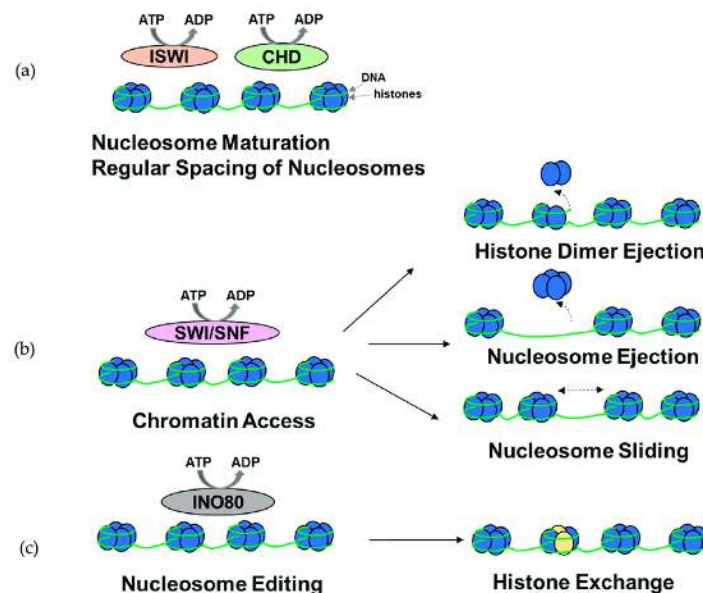


Figure 27.2.12: Overview of the functions of ATP-dependent chromatin remodeling complexes. [Hasan, N., and Ahuja, N. \(2019\) Cancers 11\(12\):1859](#)

Panel (a) shows a subset of ISWI and CHD complexes that are involved in nucleosome assembly, maturation, and spacing.

Panel (b) shows SWI/SNF complexes are primarily involved in histone dimer ejection, nucleosome ejection, and nucleosome repositioning through sliding, thus modulating chromatin access.

Panel (c) shows INO80 complexes are involved in histone exchange. It is worth noting that the complexes may be involved in other chromatin remodeling functions.

Figure 27.2.13 shows the effects of Histone Variant H3.3 on *C. elegans* Lifespan

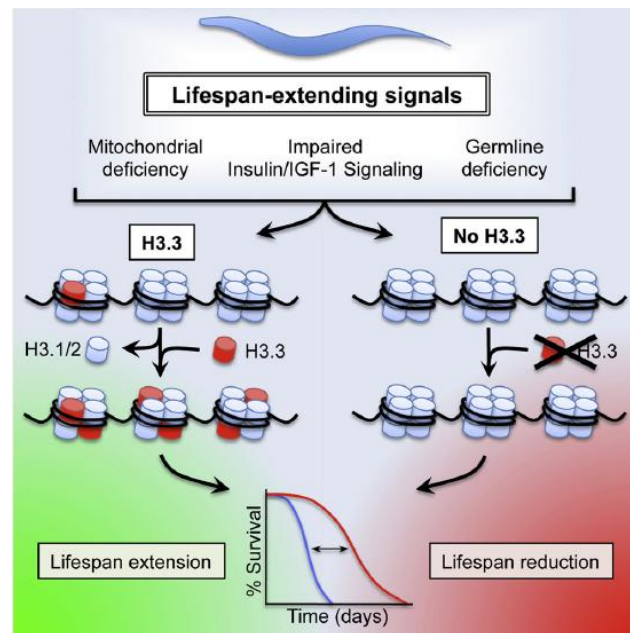


Figure 27.2.13: The Effects of Histone Variant H3.3 on *C. elegans* Lifespan. H3.3 expression increases over time in *C. elegans* during their normal lifespan. In organisms with impaired Insulin/IGF-1 signaling, germline deficiency of H3.3 resulted in significant decreases in lifespan. [Piazzesi, A., et. al. \(2016\) Cell Rep 17\(4\):987-996.](#)

27.2.4: Protein-DNA Interactions

Proteins utilize a diverse range of DNA-binding structural motifs, including the homeodomain (HD), helix-turn-helix (HTH), and high-mobility group box (HMG), to recognize DNA. HTH is the most common binding motif and can be found in several repressor and activator proteins, as shown in Figure 27.2.14. Despite their structural diversity, these domains participate in a variety of functions, including acting as substrate interaction mediators, enzymes that operate on DNA, and transcriptional regulators. Several proteins also contain flexible segments outside the DNA-binding domain to facilitate specific and non-specific interactions. For example, many HD proteins use N-terminal arms and a linker region to interact with DNA. The Encyclopedia of DNA Elements (ENCODE) data suggest that about 99.8% of putative binding motifs of TFs are not bound by their respective TFs in the genome. It is, therefore, clear that the presence of a single binding motif per TF is not adequate for TF binding.

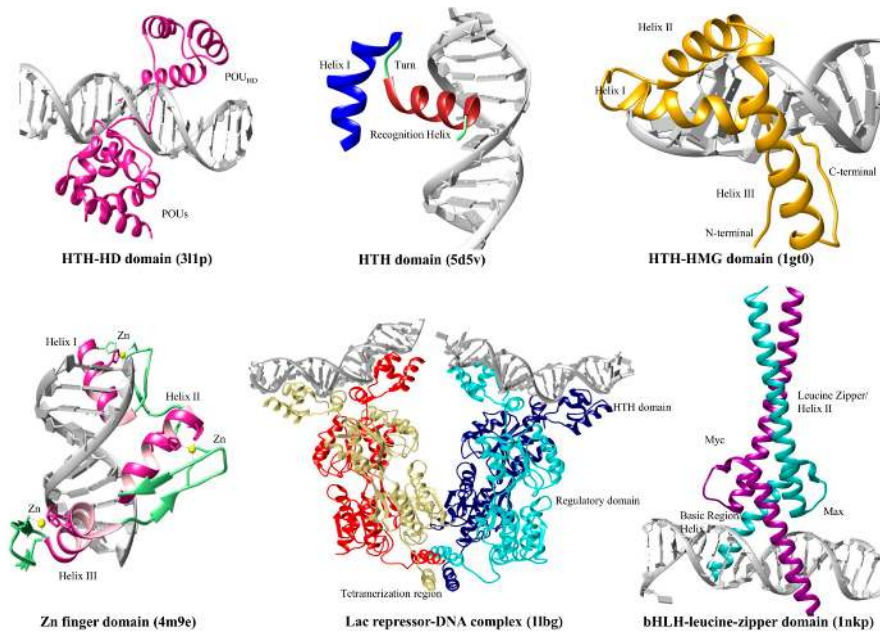


Figure 27.2.14 Representative figures of the transcription factor binding domains. The figure shows the crystal structures of different types of TF domains (3l1p, 4m9e, 5d5v, 1lbg, 1gt0, and 1nkp). The structures were obtained from the Protein Data Bank (PDB) and redrawn using Chimera. The respective domains and important regions have been labeled. HTH stands for the helix-turn-helix domain. bHLH stands for the basic helix-loop-helix motif. HD and HMG stand for homeodomain and high-mobility group box domain, respectively. [Yesudhas, D., et.al. \(2017\) Genes 8\(8\):192](#)

Figure 27.2.14 shows [interactive iCn3D models](#) of the transcription factor binding domains as depicted in the figure above. (Copyright; author via source)

Human Hsf1 with Satellite III repeat DNA - HTH Domain (5d5v) U P r o t e i n Click the image for a popup or use this external link: https://structure.ncbi.nlm.nih.gov/i...qkHMwpH7LdDmr7 N A c o m p	POU-HMG-DNA ternary complex - HTH-HMG domain (1gt0)  Click the image for a popup or use this external link: https://structure.ncbi.nlm.nih.gov/i...b9X7DEGSK2n2V7
---	---

l
e
x
H
T
H
-
H
D
d
o
m
a
i
n
(
3
l
1
p
)
C
l
i
c
k
t
h
e
i
m
a
g
e
f
o
r
a
p
o
p

u
p
o
r
u
s
e
t
h
i
s
e
x
t
e
r
n
a
l
l
i
n
k
:
[h
t
t
p
s
:
/
/
s
t
r
u
c
t
u
r
e
.](#)

n
c
b
i
.
n
l
m
.
n
i
h
.
g
o
v
/
i
.
.
.
f
H
m
W
M
v
U
C
i
J
u
s
7
7

Lactose operon repressor and its complexes with
DN A and inducer (1lbg)

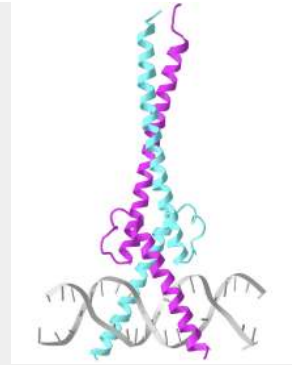
f
4
z
i
n

Myc-Max and Mad-Max recognizing DNA(1nkp)

c
f
i
n
g
e
r
D
N
A
b
i
n
d
i
n
g
d
o
m
a
i
n
i
n
c
o
m
p
l
e
x
w
i
t
h
m
e
t
h
y
l



Click the image for a popup or use this external link: <https://structure.ncbi.nlm.nih.gov/...6v3y24U7N2qYC8>



Click the image for a popup or use this external link: <https://structure.ncbi.nlm.nih.gov/...qcTBN4UwxPFH99>

a
t
e
d
D
N
A
(
4
m
9
e
)

C
l
i
c
k
t
h
e
i
m
a
g
e
f
o
r
a
p
o
p
u
l
a
r
u
s
e
t

h
i
s
e
x
t
e
r
n
a
l
l
i
n
k
:
[h
t
t
p
:
/
/
s
t
r
u
c
t
u
r
e
.
n
c
b
i
.
n
l
m](#)

.
n
i
h
.
g
o
v
/
i
.
.
.
T
7
Z
X
x
M
h
Q
6
6
W
N
L
9

Most search mechanistic studies that attempt to determine how transcription factors (TFs) find their binding sites are limited to naked DNA-protein complexes, which do not accurately reflect the actual crowded environment of a cell. Studies using naked DNA and transcription factors have shown that many DNA-binding proteins travel a long distance through one-dimensional diffusion. However, the search process for eukaryotes must occur in the presence of chromatin, which can hinder protein mobility. In this case, the protein must dissociate from the DNA, enter a 3D mode of diffusion state, and continue the target site searching process. The sliding and intersegmental transfer mechanisms can be explained through the example of the *lac* repressor. The *lac* repressor consists of four identical monomers, forming a dimer of dimers for its DNA binding. The binding sequence of these dimers is symmetric or pseudo-symmetric, and these identical monomers identify each half. The HTH domain of the *lac* repressor is the DNA-binding domain that facilitates the interaction with its target site on DNA, as shown in Figure 27.2.15

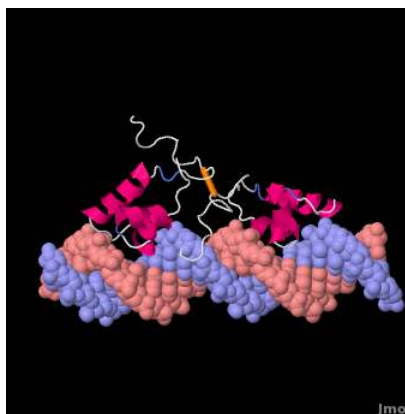


Figure 27.2.15: The Helix-Turn-Helix Motif of the Lac Repressor. Lac repressor binds to DNA non-specifically, enabling it to slide rapidly along the DNA double helix until it encounters the lac operator sequence. The DNA-binding domain employs a helix-turn-helix (HTH) motif (Alpha Helices, Turns). During non-specific binding, the hinge region is disordered. The DNA double helix is depicted as straight in the model when the Lac Repressor binds non-specifically. Upon recognizing the specific operator sequence, the non-specific binding converts to specific binding. During this conversion, the hinge region changes from disordered loops to Alpha Helices, which bind to the minor groove of the DNA. As explained below, this binding stabilizes a kinked (“bent”) DNA double helix conformation. [Protopedia – Life in 3D](#)

As a result of a rapid search (sliding) along the DNA molecule and intersegmental transfer between distant DNA sequences, the lactose repressor finds its target sites faster than the diffusion limit. The section, comprising residues 1–46 of the HTH protein domain, characterized by three α -helices, maintains its secondary structure through both specific and non-specific binding. When the repressor binds to a non-specific site, the HTH domain interacts with the DNA backbone and maintains its interaction with the helix region in the major groove. This arrangement facilitates the interaction of the recognition helix with the edges of the DNA bases, enabling the repressor to walk or search for its specific site on the DNA. The C-terminal residues of the DNA-binding domain, residues 47–62, form the hinge region and are normally disordered during non-specific recognition; however, during specific site recognition, residues 50–58 acquire an α -helix configuration (hinge helix) (Fig. 15 above). The disordered hinge region and the flexibility of the HTH domain enable the protein to move freely along the DNA, allowing it to search for its target site. In specific binding complexes, the hinge helix of each monomer is located at the symmetrical center of the binding site, allowing the hinge helices to interact with each other (intersegmental transfer) and thereby enhancing stability. Moreover, DNA bends at the symmetrical center of the specific binding site (37° angle), thereby supporting monomer-monomer interactions.

In addition to the helix-turn-helix structure, the zinc finger motif is also very common, especially in eukaryotic TFs, as shown in Figure 27.2.16. Proteins that contain zinc fingers (zinc finger proteins) are classified into several different structural families. Unlike many other clearly defined supersecondary structures, such as Greek keys or β hairpins, there are several types of zinc fingers, each with a unique three-dimensional architecture. This three-dimensional structure determines a particular zinc finger protein's class. Still, it can also be recognized based on the primary structure of the protein or the identity of the ligands coordinating the zinc ion. Despite the large variety of these proteins, however, the vast majority typically function as interaction modules that bind DNA, RNA, proteins, or other small, useful molecules. Variations in structures serve primarily to alter the binding specificity of a particular protein. The most common type of zinc finger motif utilizes two Cys and two His residues (CCHH) coordinating the Zn(II) ion to adopt a $\beta\beta\alpha$ fold with three hydrophobic residues responsible for the formation of a small hydrophobic core, which offers additional stabilization of the zinc finger domain.

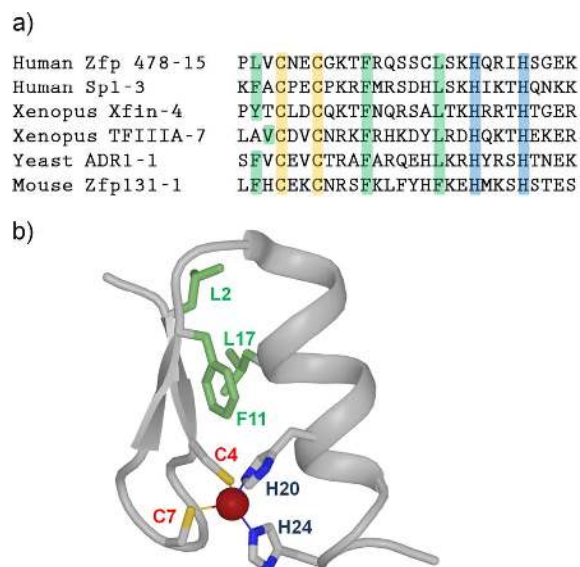


Figure 27.2.16: Sequence alignments of the CCHH zinc fingers and a representative structure. [Kluska, K., Adamczyk, J., and Krezel, A. \(2018\) Coord Chem Rev 367:18-64](#)

Panel (a) shows the alignment of the TFIIIA-like zinc finger domains from different organisms. The green color denotes residues that are responsible for the hydrophobic core formation in most CCHH zinc fingers (L17, F11, and L2). Yellow and blue indicate the coordinating Cys and His residues, respectively.

Panel (b) shows the 3D NMR structure of 15-th ZF from zinc finger protein 478 [PDB: [2YRH](#)]

Figure 27.2.17 shows an [interactive iCn3D model](#) of C2H2-type zinc finger domain (699-729) from zinc finger protein 473 (2YRH).

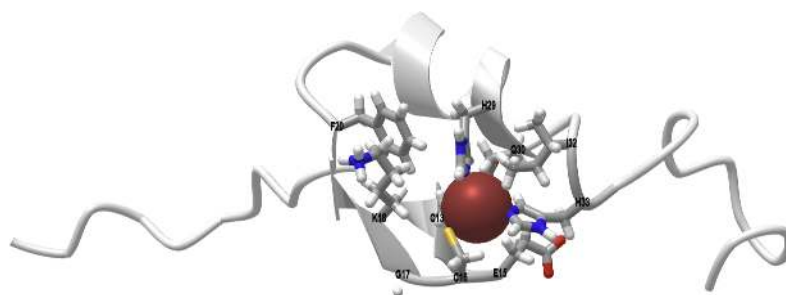


Figure 27.2.17: C2H2-type zinc finger domain (699-729) from zinc finger protein 473 (2YRH). (Copyright; author via source). Click the image for a popup or use this external link: <https://structure.ncbi.nlm.nih.gov/i...M5jGWvYU9NHP78>

27.2.4.0.1:

Overall, zinc finger motifs exhibit considerable versatility in binding modes, even among members of the same class (e.g., some bind DNA, while others bind protein), suggesting that they are stable scaffolds that have evolved specialized functions. For example, zinc finger-containing proteins are involved in various processes, including gene transcription, translation, mRNA trafficking, cytoskeleton organization, epithelial development, cell adhesion, protein folding, chromatin remodeling, and zinc sensing, among others. Zinc-binding motifs are stable structures, and they rarely undergo conformational changes upon binding their target.

The last binding domain that we will consider in detail here is the helix-loop-helix domain found in Leucine zipper-containing proteins. Specifically, bZIPs (Basic-region leucine zippers) are a class of eukaryotic transcription factors. The bZIP domain is approximately 60 to 80 amino acids in length, featuring a highly conserved DNA-binding basic region and a more diverse leucine zipper dimerization region. The two regions form α -helical structures that are connected via a looped region. This forms a core helix-loop-helix (HLH) structure within each monomer of the protein. Two monomers then join through the formation of a leucine zipper junction, forming a heterodimeric protein structure. The resulting heterodimer can bind with DNA in a sequence-specific manner through the basic α -helices as shown in Figure 27.2.18

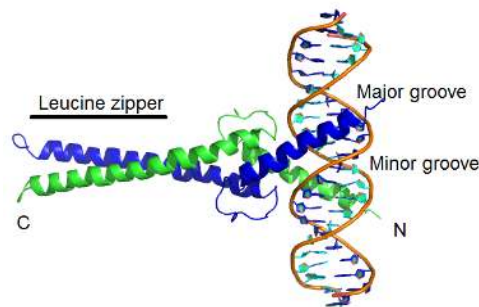


Figure 27.2.18: Leucine Zipper Transcription Factors from the bZIP family. The monomer subunits of a heterodimeric bZIP protein contain a Helix-loop-Helix (HLH) core structure, where one helix forms the leucine zipper with the other monomer, and the basic helices of each monomer interact with the major groove of the target DNA. A flexible loop region holds the helices together. (One monomer is shown in blue and one monomer is shown in green). [Latacca](#)

Specifically, basic residues, such as lysines and arginines, interact in the major groove of the DNA, forming sequence-specific interactions. Most bZIP proteins show a high binding affinity for the ACGT motifs. The bZIP heterodimers exist in a variety of eukaryotes and are more common in organisms with higher evolutionary complexity.

Figure 27.2.19 shows an [interactive iCn3D model](#) of the GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted alpha helices (1YSA).

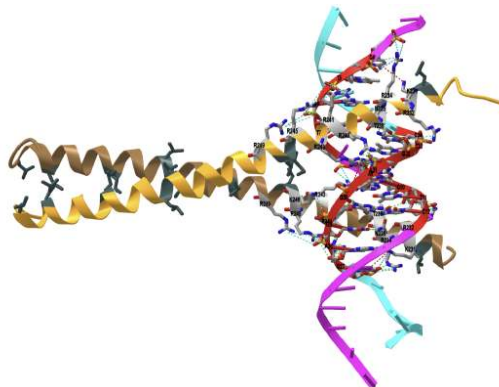


Figure 27.2.19: GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted alpha helices (1YSA). (Copyright; author via source). Click the image for a popup or use this external link: <https://structure.ncbi.nlm.nih.gov/i...hTPDCEQJeLm4z6>

27.2.5:

27.2.6: Epigenetics and Transgenerational Inheritance

Even though all somatic cells of a multicellular organism have the same genome, different cell types have different transcriptomes (sets of all expressed RNA molecules), different proteomes (sets of all proteins), and, hence, different functions. Cell differentiation during embryonic development requires the activation and repression of specific gene sets by the action of cell lineage-defining transcription factors. Within a cell lineage, gene activity states are often maintained over several rounds of cell division (a phenomenon known as “cellular memory” or “cellular inheritance”).

Since the rediscovery of epigenetics some 30 years ago (Conrad Hal Waddington originally proposed it in the early 1940s), cellular inheritance has been attributed to gene regulatory feedback loops, chromatin modifications (including DNA methylation and histone modifications), as well as long-lived non-coding RNA molecules, collectively referred to as the “*epigenome*”. Among the various chromatin modifications, DNA methylation and polycomb-mediated silencing are probably the most stable, endowing genomes with the ability to impose silencing of transcription for specific sequences even in the presence of all the factors required for their expression.

27.2.7: Defining Transgenerational Epigenetic Inheritance

The metastability of the epigenome explains why development is both plastic and canalized, as proposed initially by Waddington. Although epigenetics deals only with the cellular inheritance of chromatin and gene expression states, it has been proposed that epigenetic features could also be transmitted through the germline and persist in subsequent generations. The widespread interest in **“transgenerational epigenetic inheritance”** is nourished by the hope that epigenetic mechanisms might provide a basis for the inheritance of acquired traits. Yes, Lamarck has never been dead and every so often raises his head, this time with the help of epigenetics.

Although acquired traits related to body or brain functions can be encoded in the epigenome of a cell, they cannot be easily transmitted from one generation to the next. For this to occur, these epigenetic changes would have to manifest in the germ cells as well, which in mammals are separated from somatic cells by the so-called Weismann barrier. Further, the chromatin is extensively reshaped during germ cell differentiation as well as during the development of totipotent cells after fertilization, even though some loci appear to escape epigenetic reprogramming in the germline. Long-lived RNA molecules appear to be less affected by these barriers and are therefore more likely to carry epigenetic information across generations, although the mechanisms underlying this process are largely unresolved.

27.2.8: Evidence for Transgenerational Epigenetic Inheritance

Over the past decade, numerous studies on transgenerational responses to environmental or metabolic factors in mice and rats have been published. The factors include endocrine disruptors, high-fat diet, obesity, diabetes, undernourishment, and trauma. These studies investigated DNA methylation, sperm RNA, or both. For example, when male mice are made prediabetic by treatment with streptozotocin, it affects the DNA methylation patterns in their resulting sperm, as well as the pancreatic islets of F1 and F2 of the resulting offspring. Furthermore, studies have shown that traumatic stress in early life altered behavioral and metabolic processes in the progeny and that injection of sperm RNAs from traumatized males into fertilized wild-type oocytes reproduced the alterations in the resulting offspring.

In humans, epidemiological studies have linked food supply in the grandparental generation to health outcomes in the grandchildren. An indirect study based on DNA methylation and polymorphism analyses has suggested that sporadic imprinting defects in Prader–Willi syndrome are due to the inheritance of a grandmaternal methylation imprint through the male germline. Due to the uniqueness of these human cohorts, these findings still require independent replication. Most cases of segregation of abnormal DNA methylation patterns in families with rare diseases, however, turned out to be caused by an underlying genetic variant. Thus, studies of this nature must rule out the effects of traditional genetic inheritance as a factor in the observed phenotypes.

Genetic inheritance alone cannot fully explain why we resemble our parents. In addition to genes, we inherit from our parents the environment and culture, which in part have been shaped by previous generations, as shown in Figure 27.2.20 A specific form of the environment is our mother’s womb, to which we were exposed during the first 9 months of our lives. The maternal environment can have a profound and lasting impact on our health. In the Dutch hunger winter, for example, severe undernourishment affected pregnant women, their unborn offspring, and the offspring’s fetal germ cells. The increased incidence of cardiovascular and metabolic disease observed in F1 adults is not due to the transmission of epigenetic information through the maternal germline, but a direct consequence of the exposure in utero, a phenomenon called “fetal programming” or—if fetal germ cells and F2 offspring are affected—“intergenerational inheritance”.

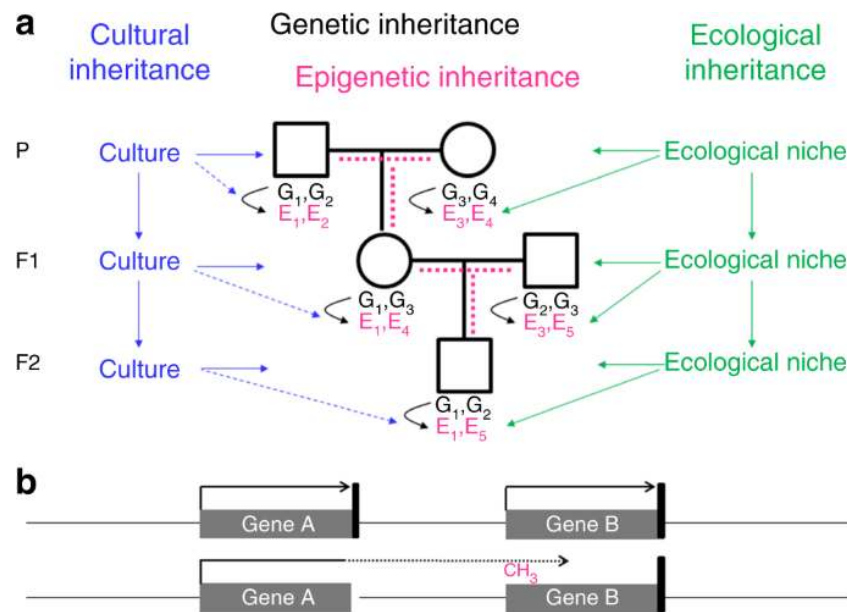


Figure 27.2.20: Transgenerational inheritance systems. [Horsthemke, B. \(2018\) Nat Comm. 9:2973](#)

Panel a illustrates that offspring inherit genes from their parents (black), the environment (green), and culture (blue). Genes and the environment affect the epigenome (magenta) and the phenotype. Culture also affects the phenotype; however, there is currently no evidence of a direct effect of culture on the epigenome (broken blue lines). It is a matter of debate how much epigenetic information is inherited through the germline (broken magenta lines). G genetic variant, E epigenetic variant.

Panel b shows that an epimutation (promoter methylation and silencing of gene B in this example) often results from aberrant read-through transcription from a mutant neighboring gene, either in sense orientation as shown here or in antisense orientation. The presence of such a secondary epimutation in several generations of a family mimics transgenerational epigenetic inheritance, although it represents a form of genetic inheritance. Black arrow, transcription; black vertical bar, transcription termination signal; broken arrow, read-through transcription

27.2.9: Roadmap to Proving Transgenerational Epigenetic Inheritance

Here are some steps to demonstrate that epigenetics, rather than classical genetics determine inheritance.

- Rule out genetic, ecological, and cultural inheritance. For studies in mice and rats, inbred strains and strictly controlled environments are necessary. When a pregnant female animal is exposed to a specific environmental stimulus, F3 offspring and subsequent generations must be studied to exclude a direct effect of the stimulus on the somatic cells and germ cells of the embryos. Even more desirable is the use of *in vitro* fertilization (IVF), embryo transfer, and foster mothers. When a male animal is exposed to an environmental stimulus, F2 offspring must be studied to exclude transient effects on germ cells. To ensure that any phenotype is exclusively transmitted via gametes, IVF must be used, controlling for potential artifacts associated with IVF.
- In contrast to laboratory animals, it is impossible to rule out ecological and cultural inheritance in humans; however, genetic effects can be excluded. If an epimutation follows Mendelian inheritance patterns, be cautious: you are more likely looking at a secondary epimutation and genetic inheritance. Study the haplotype background of the epimutation: if it is always found on the same haplotype in a given family, you are most likely dealing with a secondary epimutation. Do whole genome sequencing to search for a genetic variant that might have caused the epimutation, and be aware that this variant might be distantly located. Good spots to start looking at are the two neighboring genes, where a mutation might cause transcriptional read-through in sense or antisense orientation into the locus under investigation. Unfortunately, if you don't find anything, you still cannot be 100% sure that a genetic variant does not exist.
- Identify the responsible epigenetic factor in the germ cells. Admittedly, this is easier said than done, especially in female germ cells, which are scarce or unavailable. Be aware that germ cell preparations may be contaminated with somatic cells or somatic DNA. Use swim-up (sperm) or micromanipulation techniques to purify germ cells to the highest purity. Exclude the presence of somatic cells and somatic DNA by molecular testing, for example, by methylation analysis of imprinted genes, which are fully methylated or fully unmethylated only in germ cells.

- Demonstrate that the epigenetic factor in the germ cells is responsible for the phenotypic effect in the next generation. If possible, remove the factor from the affected germ cells and demonstrate that the effect is lost. Add the factor to control germ cells and demonstrate that the effect is gained. While RNA molecules can and have been extracted from the sperm of exposed animals and injected into control zygotes, DNA methylation, and histone modifications cannot easily be manipulated (although CRISPR/Cas9-based epigenome editors are being developed and used for this purpose), and all of these experiments can hardly be done in humans. In light of these problems, it may currently be too much to ask to prove transgenerational epigenetic inheritance in humans, but it should nevertheless be kept in mind and discussed.

27.2.10: Summary

This chapter provides an integrative overview of the multiple layers of gene regulation that determine cellular phenotype, despite the uniformity of the genomic DNA. It emphasizes both the regulation of transcription initiation by transcription factors (TFs) and the epigenetic mechanisms that contribute to cellular memory and potentially transgenerational inheritance.

Transcription Initiation and Enhancer Function

- **TF Assembly at Promoters:**

The initiation of transcription requires the coordinated binding of numerous TFs at the promoter region. These factors not only help recruit RNA polymerase but also direct its activity by interacting with specific DNA recognition sequences.

- **Enhancer-Mediated Regulation:**

Distant enhancer elements can loop back to interact with promoter regions, stabilizing the transcription complex and modulating gene expression. This dynamic interplay between promoters and enhancers underscores the complexity of transcriptional control.

- **Chromatin Remodeling:**

Chromatin structure plays a pivotal role in regulating transcription. Transcriptional control is achieved through:

- Activation of TFs by cellular signaling (often via post-translational modifications or allosteric binding of small molecules).
- Specific binding of TFs to DNA regulatory elements.
- Remodeling of chromatin, which involves histone modifications and the action of ATP-dependent chromatin remodelers to alter nucleosome positioning and accessibility.

Transcription Factor Activation and Function

- **p53 – The Guardian of the Genome:**

p53 is a key transcription factor activated by various cellular stresses (e.g., DNA damage, oncogene activation). Once activated—often through phosphorylation, acetylation, and other post-translational modifications—p53 forms a tetramer that binds specific response elements to regulate a multitude of target genes. These targets control processes such as cell cycle arrest, DNA repair, apoptosis, and senescence, and are critical for tumor suppression as well as various physiological processes.

- **Steroid Hormone Receptors (SHRs):**

As members of the nuclear receptor superfamily, SHRs serve as ligand-activated TFs. They possess modular domains—a variable N-terminal domain (with AF-1), a highly conserved DNA-binding domain, a flexible hinge, and a C-terminal ligand-binding domain (with AF-2). Binding of steroid hormones (or antagonists) triggers receptor dimerization, nuclear translocation, and subsequent binding to hormone response elements (HREs) to regulate gene expression.

Mechanisms of TF-DNA Recognition

- **Target Search Strategies:**

TFs utilize a combination of sliding (1D diffusion), hopping, and intersegmental transfer along the DNA to locate their specific binding sites rapidly. This is exemplified by the lac repressor, whose helix-turn-helix domain facilitates both nonspecific and specific DNA interactions.

- **Enhancement by Short Tandem Repeats (STRs):**

STRs surrounding TF target sites increase the effective binding affinity and local concentration of TFs, thereby enhancing specificity and transcriptional regulation.

Epigenetics and Transgenerational Inheritance

- **Epigenetic Regulation and Cellular Memory:**

Despite identical genomic sequences, cells exhibit diverse transcriptomes and proteomes through differential gene expression, controlled in part by epigenetic mechanisms. Key epigenetic marks include DNA methylation, histone modifications, and the influence of long-lived noncoding RNAs.

- **Transgenerational Inheritance:**

The chapter explores the concept of transgenerational epigenetic inheritance—the possibility that epigenetic information, acquired in response to environmental stimuli, can be passed through the germline to subsequent generations. It highlights the experimental challenges in distinguishing these effects from classical genetic, ecological, and cultural inheritance.

Conclusion

In summary, this chapter elucidates how transcription is finely tuned through the interplay of TF activation, DNA binding dynamics, enhancer interactions, and chromatin remodeling. Additionally, it considers the emerging field of epigenetics, which seeks to explain how gene expression states are maintained and, in some cases, transmitted across generations. Together, these processes form a complex regulatory network that not only drives cellular differentiation and adaptation but also has profound implications for human health and disease.

27.2.11: References

1. Parker, N., Schneegurt, M., Thi Tu, A.-H., Lister, P., Forster, B.M. (2019) Microbiology. *Openstax*. Available at: <https://opentextbc.ca/microbiologyopenstax/>
2. Chan, K.-G., Liu, Y.-C., and Chang C.-Y. (2015) Inhibiting N-acyl-homoserine lactone synthesis and quenching *Pseudomonas* quinolone quorum sensing to attenuate virulence. *Front. Microbiol.* 6:1173. Available at: <https://www.frontiersin.org/articles/10.3389/fmicb.2015.01173/full>
3. Rukavina, Z., and Vanic Zeljka. (2016) Current trends in the development of liposomes targeting bacterial biofilms. *Pharmaceutics* 8(2):18. Available at: <https://www.mdpi.com/1999-4923/8/2/18/html>
4. Wikipedia contributors. (2020, April 18). Guanosine pentaphosphate. In *Wikipedia, The Free Encyclopedia*. Retrieved 16:26, August 23, 2020, from https://en.Wikipedia.org/w/index.php?title=Guanosine_pentaphosphate&oldid=951778776
5. Verbeke, F., De Craemer, S., Debunne, N., Janssens, Y., Wynendaele, E., Van de Wiele, C., and De Spiegeleer, B. (2017) Peptides as quorum sensing molecules: measurement techniques and obtained levels *in vitro* and *in vivo*. *Frontiers in Neuroscience* 11:183. Available at: https://www.researchgate.net/publication/316055402_Peptides_as_Quorum_Sensing_Molecules_Measurement_Techniques_and_Obtained_Levels_In_vitro_and_In_vivo
6. Yesudhas, D., Batool, M., Anwar, M.A., Panneerselvam, S., and Choi, S. (2017) Proteins recognizing DNA: Structural uniqueness and versatility of DNA-binding domains in Stem Cell Transcription Factors. *Genes* 8(8):192. Available at: <https://www.mdpi.com/2073-4425/8/8/192/html>
7. Castellanos, M., Mothi, N., and Muñoz, V. (2020) Eukaryotic transcription factors can track and control their target genes using DNA antennas. *Nature Comm.* 11:540. Available at: <https://www.nature.com/articles/s41467-019-14217-8>
8. Neideracher, G., Klopff, E., and Schüller, C. (2011) Interplay of dynamic transcription and chromatin remodeling: Lessons from yeast. *Int J Mol Sci* 12(8):4758–4769. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3179130/>
9. Kim, S., and Kaang, B.-K. (2017) Epigenetic regulation and chromatin remodeling in learning and memory. *Exp. & Mol. Med.* 49:e281. Available at: <https://www.nature.com/articles/emm2016140#Fig1>
10. Tvardovskiy, A., Schwämmle, V., Kempf, S., Rogowska-Wrzesinska, A., and Jensen, O.N. (2016) Accumulation of histone variant H3.3 with age is associated with profound changes in the histone methylation landscape. *Nuc. Acids Res.* 45(16):1093. Available at: https://www.researchgate.net/publication/318987684_Accumulation_of_histone_variant_H33_with_age_is_associated_with_profound_changes_in_the_histone_methylation_landscape
11. Cipolletti, M., Fernandez, V.S., Montalesi, E., Marino, M., Fiochetti, M. (2018) Beyond the antioxidant activity of dietary polyphenols in cancer: The modulation of estrogen receptors (ERs) signaling. *Int J. Mol Sci* 19(9):2624. Available at: <https://www.mdpi.com/1422-0067/19/9/2624/html>
12. Griekspoor, A., Zward, W., Neeffjes, J., and Michalides, R. (2007) Visualizing the action of steroid hormone receptors in living cells. *Nucl. Recept. Signal.* 5:e003 Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1853070/>
13. Mitsis, T., Papargorgiou, L., Efthimiadou, A., Bacopoulou, F., Vlachakis, D., Chrousos, G.P., Eliopoulos, E. (2020) A comprehensive structural and functional analysis of the ligand binding domain of the nuclear receptor superfamily reveals

highly conserved signaling motifs and two distinct canonical forms through evolution. *World Acad Sci J* 1: 264-274, 2019.

Available at: <https://www.spandidos-publications.com/10.3892/wasj.2020.30>

14. Reed, S.M., and Quelle, D.E. (2015) p53 Acetylation: Regulation and consequences. *Cancers* 7(1):30-69. Available at: <https://www.mdpi.com/2072-6694/7/1/30/html>.
15. MacLaine, N.J., and Hupp, T.R. (2009) The regulation of p53 by phosphorylation: a model for how distinct signals integrate into the p53 pathway. *Aging* 1(5):490-502. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2806026/>
16. Wikipedia contributors. (2020, August 1). Estrogen. In *Wikipedia, The Free Encyclopedia*. Retrieved 01:28, September 6, 2020, from <https://en.Wikipedia.org/w/index.php?title=Estrogen&oldid=970560042>
17. Kluska, K., Adamczyk, J., and Krezel, A. (2018) Metal binding properties, stability, and reactivity of zinc fingers. *Coord. Chem Rev.* 367:18-64. Available at: <https://www.sciencedirect.com/science/article/pii/S0010854517305441>
18. Wikipedia contributors. (2020, July 4). Leucine zipper. In *Wikipedia, The Free Encyclopedia*. Retrieved 07:00, September 7, 2020, from https://en.Wikipedia.org/w/index.php?title=Leucine_zipper&oldid=965962667
19. Wikipedia contributors. (2020, April 15). Zinc finger. In *Wikipedia, The Free Encyclopedia*. Retrieved 18:28, September 7, 2020, from https://en.Wikipedia.org/w/index.php?title=Zinc_finger&oldid=951116840
20. Horsthemke, B. (2018) A critical view on transgenerational epigenetic inheritance in humans. *Nat. Comm.* 9:2973. Available at: <https://www.nature.com/articles/s41467-018-05445-5#rightslink>

This page titled [27.2: Regulation of Gene Expression in Eukaryotes](#) is shared under a [not declared](#) license and was authored, remixed, and/or curated by [Henry Jakubowski and Patricia Flatt](#).

- [Current page](#) by Henry Jakubowski and Patricia Flatt has no license indicated.
- [11.2: Diffusion Across a Membrane - Passive and Facilitated Diffusion](#) by Henry Jakubowski and Patricia Flatt has no license indicated.

Index

A

alarmones
27.1: Regulation of Gene Expression in Bacteria
alkylation
24.2: DNA Mutations, Damage, and Repair
allosteric regulation
27.1: Regulation of Gene Expression in Bacteria
alternate sigma factors
27.1: Regulation of Gene Expression in Bacteria
alternative splicing
25.2: RNA Processing
amino acids
26.1: The Genetic Code
aneuploidy
23.1: Gene Mapping and Chromosomal Karyotypes
attenuation
27.1: Regulation of Gene Expression in Bacteria

B

bacterial gene regulation
27.1: Regulation of Gene Expression in Bacteria
bacterial operons
27.1: Regulation of Gene Expression in Bacteria
Barbara McClintock
23.2: DNA Transposable Elements
base excision repair
24.2: DNA Mutations, Damage, and Repair
beneficial mutations
24.2: DNA Mutations, Damage, and Repair
biochemical processes
26.3: Translational Regulation and Protein Degradation
biochemistry
23.2: DNA Transposable Elements
24.2: DNA Mutations, Damage, and Repair
26.1: The Genetic Code
biofilms
27.1: Regulation of Gene Expression in Bacteria
biotechnology
24.3: DNA Recombination

C

cancer
24.2: DNA Mutations, Damage, and Repair
cancer research
24.3: DNA Recombination
catabolite activator protein
27.1: Regulation of Gene Expression in Bacteria
cell cycle
23.3: Chromosome Packaging
24.1: DNA Replication
cell cycle regulation
24.2: DNA Mutations, Damage, and Repair
27.2: Regulation of Gene Expression in Eukaryotes
cellular homeostasis
26.3: Translational Regulation and Protein Degradation
centiMorgans
23.1: Gene Mapping and Chromosomal Karyotypes
centromere
23.3: Chromosome Packaging
chiasma
24.3: DNA Recombination
chromatid
23.3: Chromosome Packaging

chromatin compaction
23.3: Chromosome Packaging
chromatin remodeling
23.3: Chromosome Packaging
24.1: DNA Replication
27.2: Regulation of Gene Expression in Eukaryotes
chromatin structure
23.3: Chromosome Packaging
27.2: Regulation of Gene Expression in Eukaryotes
chromosomal aberrations
23.1: Gene Mapping and Chromosomal Karyotypes
chromosomal alterations
24.2: DNA Mutations, Damage, and Repair
chromosomal crossover
23.1: Gene Mapping and Chromosomal Karyotypes
chromosomal karyotypes
23.1: Gene Mapping and Chromosomal Karyotypes
chromosomal organization
23.3: Chromosome Packaging
chromosomal segregation
24.3: DNA Recombination
chromosomal translocation
23.1: Gene Mapping and Chromosomal Karyotypes
chromosome
23.3: Chromosome Packaging
chromosome territories
23.3: Chromosome Packaging
class I mobile elements
23.2: DNA Transposable Elements
class II mobile elements
23.2: DNA Transposable Elements
codons
26.1: The Genetic Code
crossover
24.3: DNA Recombination
cytogenetics
23.1: Gene Mapping and Chromosomal Karyotypes

D

direct repair
24.2: DNA Mutations, Damage, and Repair
DNA crosslinking
24.2: DNA Mutations, Damage, and Repair
DNA damage
24.2: DNA Mutations, Damage, and Repair
DNA methylation
23.2: DNA Transposable Elements
DNA mutations
24.2: DNA Mutations, Damage, and Repair
DNA packaging
23.3: Chromosome Packaging
DNA polymerases
24.1: DNA Replication
24.2: DNA Mutations, Damage, and Repair
DNA recombination
24.3: DNA Recombination
DNA repair
23.3: Chromosome Packaging
24.2: DNA Mutations, Damage, and Repair
24.3: DNA Recombination
DNA replication
23.3: Chromosome Packaging
24.1: DNA Replication
DNA replication errors
24.2: DNA Mutations, Damage, and Repair

DNA strand breaks
24.2: DNA Mutations, Damage, and Repair
DNA synthesis
24.2: DNA Mutations, Damage, and Repair
DNA transcription
25.1: DNA-Dependent Synthesis of RNA
DNA transposons
23.2: DNA Transposable Elements
Down syndrome
24.2: DNA Mutations, Damage, and Repair

E

E. coli
27.1: Regulation of Gene Expression in Bacteria
E. coli replication
24.1: DNA Replication
elongation
26.2: Protein Synthesis
elongation factors
26.2: Protein Synthesis
environmental mutagens
24.2: DNA Mutations, Damage, and Repair
environmental response
27.1: Regulation of Gene Expression in Bacteria
environmental signals
26.3: Translational Regulation and Protein Degradation
enzyme catalysis
25.3: RNA-Dependent Synthesis of RNA and DNA
epigenetic regulation
23.2: DNA Transposable Elements
epigenetics
27.2: Regulation of Gene Expression in Eukaryotes
eukaryotic cells
25.2: RNA Processing
26.3: Translational Regulation and Protein Degradation
eukaryotic gene regulation
27.2: Regulation of Gene Expression in Eukaryotes
eukaryotic genomes
23.2: DNA Transposable Elements
eukaryotic initiation
26.2: Protein Synthesis
eukaryotic replication
24.1: DNA Replication
eukaryotic transcription
25.1: DNA-Dependent Synthesis of RNA
eukaryotic translation
26.1: The Genetic Code
evolution
24.3: DNA Recombination

F

frameshift mutations
24.2: DNA Mutations, Damage, and Repair

G

gene expression
25.1: DNA-Dependent Synthesis of RNA
25.2: RNA Processing
26.3: Translational Regulation and Protein Degradation
27.1: Regulation of Gene Expression in Bacteria
gene mapping
23.1: Gene Mapping and Chromosomal Karyotypes

gene regulation

23.3: Chromosome Packaging

gene targeting

24.3: DNA Recombination

genetic code

26.1: The Genetic Code

genetic code universality

26.1: The Genetic Code

genetic diseases

24.2: DNA Mutations, Damage, and Repair

genetic diversity

23.2: DNA Transposable Elements

24.3: DNA Recombination

genetic information

25.1: DNA-Dependent Synthesis of RNA

genetic linkage

23.1: Gene Mapping and Chromosomal Karyotypes

genetic map

23.1: Gene Mapping and Chromosomal Karyotypes

genetic markers

23.1: Gene Mapping and Chromosomal Karyotypes

genetic mutations

26.1: The Genetic Code

genome evolution

23.2: DNA Transposable Elements

genome organization

23.1: Gene Mapping and Chromosomal Karyotypes

genome rearrangements

23.2: DNA Transposable Elements

genome stability

23.2: DNA Transposable Elements

24.1: DNA Replication

genomic integrity

23.3: Chromosome Packaging

genomic stability

24.2: DNA Mutations, Damage, and Repair

24.3: DNA Recombination

germline mutations

24.2: DNA Mutations, Damage, and Repair

H

harmful mutations

24.2: DNA Mutations, Damage, and Repair

helicase

24.1: DNA Replication

histone modification

27.2: Regulation of Gene Expression in Eukaryotes

histone modifications

23.3: Chromosome Packaging

histone octamer

23.3: Chromosome Packaging

Holliday junction

24.3: DNA Recombination

homologous recombination

24.3: DNA Recombination

horizontal transposon transfer

23.2: DNA Transposable Elements

host genome interaction

23.2: DNA Transposable Elements

I

initiation factors

26.2: Protein Synthesis

insertional mutagens

23.2: DNA Transposable Elements

K

karyogram

23.1: Gene Mapping and Chromosomal Karyotypes

karyosystematics

23.1: Gene Mapping and Chromosomal Karyotypes

karyotyping

23.1: Gene Mapping and Chromosomal Karyotypes

Kozak sequence

26.2: Protein Synthesis

L

lac operon

27.1: Regulation of Gene Expression in Bacteria

lagging strand

24.1: DNA Replication

leading strand

24.1: DNA Replication

linkage maps

23.1: Gene Mapping and Chromosomal Karyotypes

M

meiosis

24.3: DNA Recombination

meiosis stages

24.3: DNA Recombination

messenger RNA

25.1: DNA-Dependent Synthesis of RNA

microRNA

26.3: Translational Regulation and Protein Degradation

microRNAs

26.3: Translational Regulation and Protein Degradation

mismatch repair

24.2: DNA Mutations, Damage, and Repair

mitochondrial DNA

24.1: DNA Replication

mitosis

23.3: Chromosome Packaging

mobile genetic elements

23.2: DNA Transposable Elements

molecular biology

27.1: Regulation of Gene Expression in Bacteria

mRNA

26.1: The Genetic Code

26.2: Protein Synthesis

mRNA processing

25.2: RNA Processing

mRNA structural elements

26.3: Translational Regulation and Protein Degradation

mutation examples

24.2: DNA Mutations, Damage, and Repair

N

nucleosome

23.3: Chromosome Packaging

nucleotide excision repair

24.2: DNA Mutations, Damage, and Repair

nucleotides

25.1: DNA-Dependent Synthesis of RNA

O

Okazaki fragments

24.1: DNA Replication

operon model

27.1: Regulation of Gene Expression in Bacteria

origin of replication

24.1: DNA Replication

oxidative damage

24.2: DNA Mutations, Damage, and Repair

P

p53

27.2: Regulation of Gene Expression in Eukaryotes

Peptide bond

26.1: The Genetic Code

peptidyl transferase

26.2: Protein Synthesis

Peptidyl Transferase Center

26.3: Translational Regulation and Protein Degradation

point mutations

24.2: DNA Mutations, Damage, and Repair

polyadenylation

25.2: RNA Processing

polycistronic transcript

27.1: Regulation of Gene Expression in Bacteria

polymerase domains

25.3: RNA-Dependent Synthesis of RNA and DNA

polymerase structure

25.3: RNA-Dependent Synthesis of RNA and DNA

primase

24.1: DNA Replication

primosome

24.1: DNA Replication

prokaryotic gene regulation

27.1: Regulation of Gene Expression in Bacteria

prokaryotic initiation

26.2: Protein Synthesis

prokaryotic transcription

25.1: DNA-Dependent Synthesis of RNA

prokaryotic translation

26.1: The Genetic Code

promoter regions

25.1: DNA-Dependent Synthesis of RNA

proteasome

26.3: Translational Regulation and Protein Degradation

protein degradation

26.3: Translational Regulation and Protein Degradation

protein factors

24.3: DNA Recombination

protein folding

26.3: Translational Regulation and Protein Degradation

protein synthesis

26.1: The Genetic Code

26.2: Protein Synthesis

26.3: Translational Regulation and Protein Degradation

protein synthesis fidelity

26.1: The Genetic Code

proteostasis

26.3: Translational Regulation and Protein Degradation

Q

quorum sensing

27.1: Regulation of Gene Expression in Bacteria

R

reading frame

26.1: The Genetic Code

recombinase proteins

24.3: DNA Recombination

recombination frequency

23.1: Gene Mapping and Chromosomal Karyotypes

recombination pathways

24.3: DNA Recombination

regulatory genes

27.1: Regulation of Gene Expression in Bacteria

regulatory proteins

27.1: Regulation of Gene Expression in Bacteria

repressor

27.1: Regulation of Gene Expression in Bacteria

retrotransposition

23.2: DNA Transposable Elements

retrotransposons

23.2: DNA Transposable Elements

reverse transcription

25.3: RNA-Dependent Synthesis of RNA and DNA

ribosomal RNA

25.1: DNA-Dependent Synthesis of RNA

ribosomal structure

26.1: The Genetic Code

ribosomal subunits

26.2: Protein Synthesis

ribosome

26.1: The Genetic Code

26.2: Protein Synthesis

ribosome function

26.1: The Genetic Code

ribosome heterogeneity

26.3: Translational Regulation and Protein Degradation

ribosome structure

26.3: Translational Regulation and Protein Degradation

ribosomes

26.3: Translational Regulation and Protein Degradation

riboswitches

27.1: Regulation of Gene Expression in Bacteria

RNA interference

25.1: DNA-Dependent Synthesis of RNA

RNA modifications

25.2: RNA Processing

RNA Polymerase

25.1: DNA-Dependent Synthesis of RNA

RNA processing

25.1: DNA-Dependent Synthesis of RNA

25.2: RNA Processing

RNA proofreading fidelity

25.3: RNA-Dependent Synthesis of RNA and DNA

RNA regulation

25.1: DNA-Dependent Synthesis of RNA

RNA splicing

25.2: RNA Processing

RNA stability

25.1: DNA-Dependent Synthesis of RNA

25.2: RNA Processing

RNA structure

25.1: DNA-Dependent Synthesis of RNA

RNA synthesis

25.1: DNA-Dependent Synthesis of RNA

RNA synthesis mechanisms

25.3: RNA-Dependent Synthesis of RNA and DNA

RNA virus pathogenesis

25.3: RNA-Dependent Synthesis of RNA and DNA

RNA virus replication

25.3: RNA-Dependent Synthesis of RNA and DNA

RNA virus taxonomy

25.3: RNA-Dependent Synthesis of RNA and DNA

RNA viruses

25.3: RNA-Dependent Synthesis of RNA and DNA

rolling circle replication

24.1: DNA Replication

S

semiconservative replication

24.1: DNA Replication

short tandem repeats

27.2: Regulation of Gene Expression in Eukaryotes

signal integration

27.1: Regulation of Gene Expression in Bacteria

silent mutation

24.2: DNA Mutations, Damage, and Repair

sister chromatids

24.3: DNA Recombination

somatic mutations

24.2: DNA Mutations, Damage, and Repair

spectral karyotyping

23.1: Gene Mapping and Chromosomal Karyotypes

spliceosome

25.2: RNA Processing

steroid hormone receptors

27.2: Regulation of Gene Expression in Eukaryotes

strand invasion

24.3: DNA Recombination

structural genes

27.1: Regulation of Gene Expression in Bacteria

supercoiling

23.3: Chromosome Packaging

synaptonemal complex

24.3: DNA Recombination

T

telomerase

23.3: Chromosome Packaging

24.1: DNA Replication

telomeres

23.3: Chromosome Packaging

termination

26.2: Protein Synthesis

therapeutic strategies

24.2: DNA Mutations, Damage, and Repair

topoisomerase

23.3: Chromosome Packaging

topoisomerases

24.1: DNA Replication

transcription

25.2: RNA Processing

transcription elongation

25.1: DNA-Dependent Synthesis of RNA

transcription factor

27.1: Regulation of Gene Expression in Bacteria

transcription factors

25.1: DNA-Dependent Synthesis of RNA

27.2: Regulation of Gene Expression in Eukaryotes

transcription initiation

25.1: DNA-Dependent Synthesis of RNA

27.2: Regulation of Gene Expression in Eukaryotes

transcription regulation

27.1: Regulation of Gene Expression in Bacteria

transcription termination

25.1: DNA-Dependent Synthesis of RNA

transcriptional fidelity

25.1: DNA-Dependent Synthesis of RNA

transesterification

26.3: Translational Regulation and Protein Degradation

transfer RNA

25.1: DNA-Dependent Synthesis of RNA

transgenerational inheritance

27.2: Regulation of Gene Expression in Eukaryotes

translation

25.2: RNA Processing

26.1: The Genetic Code

26.2: Protein Synthesis

translation differences

26.2: Protein Synthesis

translation process

26.1: The Genetic Code

translational control

26.3: Translational Regulation and Protein Degradation

translational regulation

26.3: Translational Regulation and Protein Degradation

translesion synthesis

24.2: DNA Mutations, Damage, and Repair

transposable elements

23.2: DNA Transposable Elements

transposition

23.2: DNA Transposable Elements

transposition mechanisms

23.2: DNA Transposable Elements

tRNA

26.1: The Genetic Code

26.2: Protein Synthesis

tRNA modifications

26.1: The Genetic Code

Trp operon

27.1: Regulation of Gene Expression in Bacteria

tumor suppressor

27.2: Regulation of Gene Expression in Eukaryotes

types of mutations

24.2: DNA Mutations, Damage, and Repair

U

ubiquitination

26.3: Translational Regulation and Protein Degradation

V

viral evolution

25.3: RNA-Dependent Synthesis of RNA and DNA

viral genome replication

25.3: RNA-Dependent Synthesis of RNA and DNA

viral polymerase inhibitors

25.3: RNA-Dependent Synthesis of RNA and DNA

W

wobble base

26.1: The Genetic Code

Glossary

Anaplerotic reactions | Chemical reactions that form intermediates of a metabolic pathway.

citric acid cycle | The citric acid cycle (CAC) – also known as the TCA cycle (tricarboxylic acid cycle) or the Krebs cycle – is a series of chemical reactions used by all aerobic organisms to release stored energy through the oxidation of acetyl-CoA derived from carbohydrates, fats, and proteins. In addition, the cycle provides precursors of certain amino acids, as well as the reducing agent NADH, that are used in numerous other reactions. [[Wikipedia](#)]

FISH | Fluorescence in situ hybridization is a cytogenetic technique used to detect and localize the specific DNA or RNA sequences. [boundless]

fluorescence | The emission of light (or other electromagnetic radiation) by a material when stimulated by the absorption of radiation or of a subatomic particle [boundless]

GABA | Gamma-Aminobutyric acid is the chief inhibitory neurotransmitter in the developmentally mature mammalian central nervous system. Its principal role is reducing neuronal excitability throughout the nervous system. [[Wikipedia](#)]

GABA shunt | The GABA shunt a closed loop that is involved in the production and conservation of the pool of GABA.

glyoxylate Shunt | The glyoxylate shunt is a two-step metabolic pathway (via isocitrate lyase, aceA and malate synthase, glcB) as an alternative to the citric acid cycle.

hybridize | To combine complementary subunits of multiple biological macromolecules. [boundless]

Krebs cycle | The citric acid cycle (CAC) – also known as the TCA cycle (tricarboxylic acid cycle) or the Krebs cycle – is a series of chemical reactions used by all aerobic organisms to release stored energy through the oxidation of acetyl-CoA derived from carbohydrates, fats, and proteins. In addition, the cycle provides precursors of certain amino acids, as well as the reducing agent NADH, that are used in numerous other reactions. [[Wikipedia](#)]

Pyruvate Dehydrogenase | Pyruvate dehydrogenase is an enzyme that catalyzes the reaction of pyruvate and a lipoamide to give the acetylated dihydrolipoamide and carbon dioxide. The conversion requires the coenzyme thiamine pyrophosphate. [[Wikipedia](#)]

TCA cycle | The citric acid cycle (CAC) – also known as the TCA cycle (tricarboxylic acid cycle) or the Krebs cycle – is a series of chemical reactions used by all aerobic organisms to release stored energy through the oxidation of acetyl-CoA derived from carbohydrates, fats, and proteins. In addition, the cycle provides precursors of certain amino acids, as well as the reducing agent NADH, that are used in numerous other reactions. [[Wikipedia](#)]

- [Glossary](#) has no license indicated.

Detailed Licensing

Overview

Title: Fundamentals of Biochemistry Vol. III - Information Pathways

Webpages: 31

All licenses found:

- [Undeclared](#): 90.3% (28 pages)
- [CC BY-SA 4.0](#): 6.5% (2 pages)
- [CC BY 4.0](#): 3.2% (1 page)

By Page

- Fundamentals of Biochemistry Vol. III - Information Pathways - [CC BY 4.0](#)
 - Front Matter - [Undeclared](#)
 - [TitlePage](#) - [Undeclared](#)
 - [InfoPage](#) - [Undeclared](#)
 - [Table of Contents](#) - [Undeclared](#)
 - [Licensing](#) - [Undeclared](#)
 - [About the Authors](#) - [Undeclared](#)
 - 23: Chromosome Structure - [Undeclared](#)
 - 23.1: Gene Mapping and Chromosomal Karyotypes - [CC BY-SA 4.0](#)
 - 23.2: DNA Transposable Elements - [Undeclared](#)
 - 23.3: Chromosome Packaging - [CC BY-SA 4.0](#)
 - 24: DNA Metabolism - [Undeclared](#)
 - 24.1: DNA Replication - [Undeclared](#)
 - 24.2: DNA Mutations, Damage, and Repair - [Undeclared](#)
 - 24.3: DNA Recombination - [Undeclared](#)
 - 25: RNA Metabolism - [Undeclared](#)
 - 25.1: DNA-Dependent Synthesis of RNA - [Undeclared](#)
 - 25.2: RNA Processing - [Undeclared](#)
 - 25.3: RNA-Dependent Synthesis of RNA and DNA - [Undeclared](#)
 - 25.4: 26.4 References - [Undeclared](#)
 - 26: Protein Metabolism - [Undeclared](#)
 - 26.1: The Genetic Code - [Undeclared](#)
 - 26.2: Protein Synthesis - [Undeclared](#)
 - 26.3: Translational Regulation and Protein Degradation - [Undeclared](#)
 - 27: Regulation of Gene Expression - [Undeclared](#)
 - 27.1: Regulation of Gene Expression in Bacteria - [Undeclared](#)
 - 27.2: Regulation of Gene Expression in Eukaryotes - [Undeclared](#)
 - Back Matter - [Undeclared](#)
 - [Index](#) - [Undeclared](#)
 - [Glossary](#) - [Undeclared](#)
 - [Detailed Licensing](#) - [Undeclared](#)