

Transcription (genetics)

From Wikipedia, the free encyclopedia

Transcription is the process of creating a complementary RNA copy of a sequence of DNA.^[1] Both RNA and DNA are nucleic acids, which use base pairs of nucleotides as a complementary language that can be converted back and forth from DNA to RNA by the action of the correct enzymes. During transcription, a DNA sequence is read by RNA polymerase, which produces a complementary, antiparallel RNA strand. As opposed to DNA replication, transcription results in an RNA complement that includes uracil (U) in all instances where thymine (T) would have occurred in a DNA complement.

Transcription can be explained easily in 4 or 5 steps, each moving like a wave along the DNA.

1. Helicase unwinds/"unzips" the DNA by breaking the hydrogen bonds between complementary nucleotides.
2. RNA nucleotides are paired with complementary DNA bases.
3. RNA sugar-phosphate backbone forms with assistance from RNA polymerase.
4. Hydrogen bonds of the untwisted RNA+DNA helix break, freeing the newly synthesized RNA strand.
5. If the cell has a nucleus, the RNA is further processed (addition of a 3' poly-A tail and a 5' cap) and exits through to the cytoplasm through the nuclear pore complex.

Transcription is the first step leading to gene expression. The stretch of DNA transcribed into an RNA molecule is called a *transcription unit* and encodes at least one gene. If the gene transcribed encodes a protein, the result of transcription is messenger RNA (mRNA), which will then be used to create that protein via the process of translation. Alternatively, the transcribed gene may encode for either ribosomal RNA (rRNA) or transfer RNA (tRNA), other components of the protein-assembly process, or other ribozymes.^[citation needed]

A DNA transcription unit encoding for a protein contains not only the sequence that will eventually be directly translated into the protein (the *coding sequence*) but also *regulatory sequences* that direct and regulate the synthesis of that protein. The regulatory sequence before (upstream from) the coding sequence is called the five prime untranslated region (5'UTR), and the sequence following (downstream from) the coding sequence is called the three prime untranslated region (3'UTR).^[citation needed]

Transcription has some proofreading mechanisms, but they are fewer and less effective than the controls for copying DNA; therefore, transcription has a lower copying fidelity than DNA replication.^[2]

As in DNA replication, DNA is read from 3' → 5' during transcription. Meanwhile, the complementary RNA is created from the 5' → 3' direction. This means its 5' end is created first in base pairing. Although DNA is arranged as two antiparallel strands in a double helix, only one of the two DNA strands, called the template strand, is used for transcription. This is because RNA is only single-stranded, as opposed to double-stranded DNA. The other DNA strand is called the coding strand, because its sequence is the same as the newly created RNA transcript (except for the substitution of uracil for thymine). The use of only the 3' → 5' strand eliminates the need for the Okazaki fragments seen in DNA replication.^[citation needed]

Transcription is divided into 5 stages: *pre-initiation*, *initiation*, *promoter clearance*, *elongation* and *termination*.^[citation needed]

Contents

- 1 Major steps
 - 1.1 Pre-initiation
 - 1.2 Initiation
 - 1.3 Promoter clearance
 - 1.4 Elongation
 - 1.5 Termination
- 2 Measuring and detecting transcription
- 3 Transcription factories
- 4 History
- 5 Reverse transcription
- 6 See also
- 7 References
- 8 External links

Major steps

Pre-initiation

In eukaryotes, RNA polymerase, and therefore the initiation of transcription, requires the presence of a core promoter sequence in the DNA. Promoters are regions of DNA that promote transcription and, in eukaryotes, are found at -30, -75, and -90 base pairs upstream from the start site of transcription. Core promoters are sequences within the promoter that are essential for transcription initiation. RNA polymerase is able to bind to core promoters in the presence of various specific transcription factors.^[*citation needed*]

The most common type of core promoter in eukaryotes is a short DNA sequence known as a TATA box, found 25-30 base pairs upstream from the start site of transcription.^[*citation needed*] The TATA box, as a core promoter, is the binding site for a transcription factor known as TATA-binding protein (TBP), which is itself a subunit of another transcription factor, called Transcription Factor II D (TFIID). After TFIID binds to the TATA box via the TBP, five more transcription factors and RNA polymerase combine around the TATA box in a series of stages to form a preinitiation complex. One transcription factor, DNA helicase, has helicase activity and so is involved in the separating of opposing strands of double-stranded DNA to provide access to a single-stranded DNA template. However, only a low, or basal, rate of transcription is driven by the preinitiation complex alone. Other proteins known as activators and repressors, along with any associated coactivators or corepressors, are responsible for modulating transcription rate.^[*citation needed*]

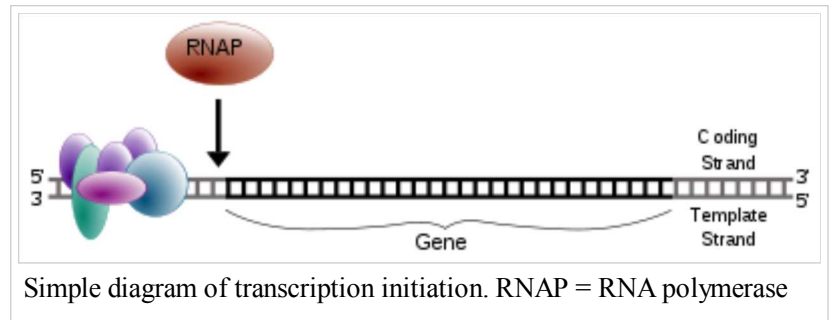
Thus, preinitiation complex contains.^[*citation needed*] 1. Core Promoter Sequence 2. Transcription Factors 3. DNA Helicase 4. RNA Polymerase 5. Activators and Repressors The transcription preinitiation in archaea is, in essence, homologous to that of eukaryotes, but is much less complex.^[3] The archaeal preinitiation complex assembles at a TATA-box binding site; however, in archaea, this complex is composed of only RNA polymerase II, TBP, and TFB (the archaeal homologue of eukaryotic transcription factor II B (TFIIB)).^{[4][5]}

Initiation

In bacteria, transcription begins with the binding of RNA polymerase to the promoter in DNA. RNA polymerase is a core enzyme consisting of five subunits: 2 α subunits, 1 β subunit, 1 β' subunit, and 1 ω subunit. At the start of initiation, the core enzyme is associated with a sigma factor that aids in finding the appropriate -35 and -10

base pairs downstream of promoter sequences.^[6] }

Transcription initiation is more complex in eukaryotes. Eukaryotic RNA polymerase does not directly recognize the core promoter sequences. Instead, a collection of proteins called transcription factors mediate the binding of RNA polymerase and the initiation of transcription. Only after certain transcription factors are attached to the promoter does the RNA polymerase bind to it. The completed assembly of transcription factors and RNA polymerase bind to the promoter, forming a transcription initiation complex. Transcription in the archaea domain is similar to transcription in eukaryotes.^[7]



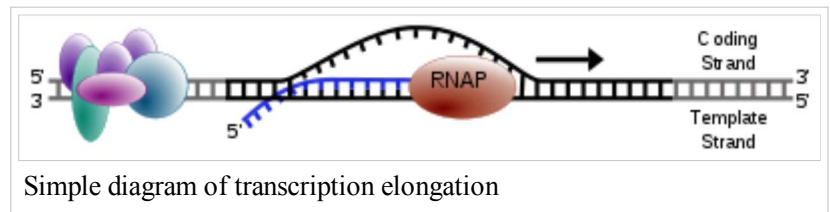
Promoter clearance

After the first bond is synthesized, the RNA polymerase must clear the promoter. During this time there is a tendency to release the RNA transcript and produce truncated transcripts. This is called *abortive initiation* and is common for both eukaryotes and prokaryotes.^[8] Abortive initiation continues to occur until the σ factor rearranges, resulting in the transcription elongation complex (which gives a 35 bp moving footprint). The σ factor is released before 80 nucleotides of mRNA are synthesized.^[9] Once the transcript reaches approximately 23 nucleotides, it no longer slips and elongation can occur. This, like most of the remainder of transcription, is an energy-dependent process, consuming adenosine triphosphate (ATP).^[citation needed]

Promoter clearance coincides with phosphorylation of serine 5 on the carboxy terminal domain of RNA Pol in eukaryotes, which is phosphorylated by TFIIF.^[citation needed]

Elongation

One strand of the DNA, the *template strand* (or noncoding strand), is used as a template for RNA synthesis. As transcription proceeds, RNA polymerase traverses the template strand and uses base pairing complementarity with the DNA template to create an RNA copy. Although RNA polymerase traverses the template strand from 3' \rightarrow 5', the coding (non-template) strand and newly-formed RNA can also be used as reference points, so transcription can be described as occurring 5' \rightarrow 3'. This produces an RNA molecule from 5' \rightarrow 3', an exact copy of the coding strand (except that thymines are replaced with uracils, and the nucleotides are composed of a ribose (5-carbon) sugar where DNA has deoxyribose (one less oxygen atom) in its sugar-phosphate backbone).^[citation needed]



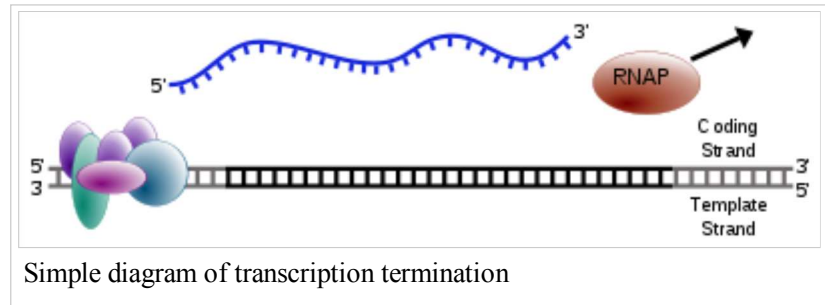
Unlike DNA replication, mRNA transcription can involve multiple RNA polymerases on a single DNA template and multiple rounds of transcription (amplification of particular mRNA), so many mRNA molecules can be rapidly produced from a single copy of a gene.^[citation needed]

Elongation also involves a proofreading mechanism that can replace incorrectly incorporated bases. In eukaryotes, this may correspond with short pauses during transcription that allow appropriate RNA editing factors to bind. These pauses may be intrinsic to the RNA polymerase or due to chromatin structure.

[*citation needed*]

Termination

Bacteria use two different strategies for transcription termination. In Rho-independent transcription termination, RNA transcription stops when the newly synthesized RNA molecule forms a G-C-rich hairpin loop followed by a run of Us. When the hairpin forms, the mechanical stress breaks the weak rU-dA bonds, now filling the DNA-RNA hybrid. This pulls the poly-U transcript out of the active site of the RNA polymerase, in effect, terminating transcription. In the "Rho-dependent" type of termination, a protein factor called "Rho" destabilizes the interaction between the template and the mRNA, thus releasing the newly synthesized mRNA from the elongation complex.^[10]

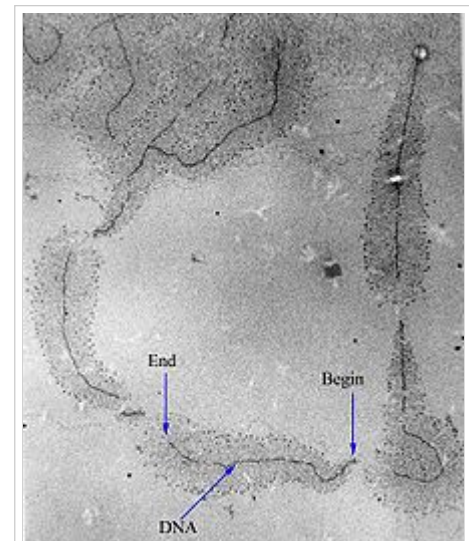


Transcription termination in eukaryotes is less understood but involves cleavage of the new transcript followed by template-independent addition of *As* at its new 3' end, in a process called polyadenylation.^[11]

Measuring and detecting transcription

Transcription can be measured and detected in a variety of ways.^[*citation needed*]

- Nuclear Run-on assay: measures the relative abundance of newly formed transcripts
- RNase protection assay and ChIP-Chip of RNAP: detect active transcription sites
- RT-PCR: measures the absolute abundance of total or nuclear RNA levels, which may however differ from transcription rates
- DNA microarrays: measures the relative abundance of the global total or nuclear RNA levels; however, these may differ from transcription rates
- In situ hybridization: detects the presence of a transcript
- MS2 tagging: by incorporating RNA stem loops, such as MS2, into a gene, these become incorporated into newly synthesized RNA. The stem loops can then be detected using a fusion of GFP and the MS2 coat protein, which has a high affinity, sequence-specific interaction with the MS2 stem loops. The recruitment of GFP to the site of transcription is visualised as a single fluorescent spot. This remarkable new approach has revealed that transcription occurs in discontinuous bursts, or pulses (see Transcriptional bursting). With the notable exception of in situ techniques, most other methods provide cell population averages, and are not capable of detecting this fundamental property of genes.^[12]
- Northern blot: the traditional method, and until the advent of RNA-Seq, the most quantitative
- RNA-Seq: applies next-generation sequencing techniques to sequence whole transcriptomes, which allows the measurement of relative abundance of RNA, as well as the detection of additional variations such as fusion genes, post-translational edits and novel splice sites



Electron micrograph of the ribosomal transcription process. The forming mRNA strands are visible as branches from the main DNA strand.

[*citation needed*]

Transcription factories

Main article: Transcription factories

Active transcription units are clustered in the nucleus, in discrete sites called transcription factories or euchromatin. Such sites can be visualized by allowing engaged polymerases to extend their transcripts in tagged precursors (Br-UTP or Br-U) and immuno-labeling the tagged nascent RNA. Transcription factories can also be localized using fluorescence in situ hybridization or marked by antibodies directed against polymerases. There are ~10,000 factories in the nucleoplasm of a HeLa cell, among which are ~8,000 polymerase II factories and ~2,000 polymerase III factories. Each polymerase II factory contains ~8 polymerases. As most active transcription units are associated with only one polymerase, each factory usually contains ~8 different transcription units. These units might be associated through promoters and/or enhancers, with loops forming a 'cloud' around the factor.^[*citation needed*]

History

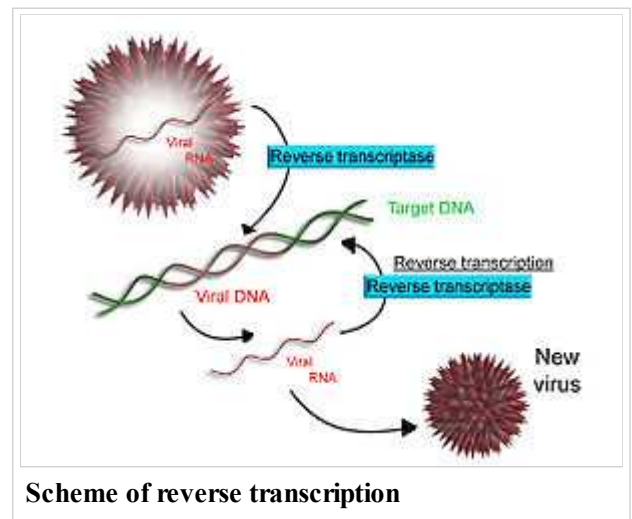
A molecule that allows the genetic material to be realized as a protein was first hypothesized by François Jacob and Jacques Monod. RNA synthesis by RNA polymerase was established *in vitro* by several laboratories by 1965; however, the RNA synthesized by these enzymes had properties that suggested the existence of an additional factor needed to terminate transcription correctly.^[*citation needed*]

In 1972, Walter Fiers became the first person to actually prove the existence of the terminating enzyme.

Roger D. Kornberg won the 2006 Nobel Prize in Chemistry "for his studies of the molecular basis of eukaryotic transcription".^[13]

Reverse transcription

Some viruses (such as HIV, the cause of AIDS), have the ability to transcribe RNA into DNA. HIV has an RNA genome that is duplicated into DNA. The resulting DNA can be merged with the DNA genome of the host cell. The main enzyme responsible for synthesis of DNA from an RNA template is called reverse transcriptase. In the case of HIV, reverse transcriptase is responsible for synthesizing a complementary DNA strand (cDNA) to the viral RNA genome. An associated enzyme, ribonuclease H, digests the RNA strand, and reverse transcriptase synthesises a complementary strand of DNA to form a double helix DNA structure. This cDNA is integrated into the host cell's genome via another enzyme (integrase) causing the host cell to generate viral proteins that reassemble into new viral particles. In HIV, subsequent to this, the host cell undergoes programmed cell death, apoptosis of T cells.^[14] However, in other retroviruses, the host cell remains intact as the virus buds out of the cell.



Some eukaryotic cells contain an enzyme with reverse transcription activity called telomerase. Telomerase is a reverse transcriptase that lengthens the ends of linear chromosomes. Telomerase carries an RNA template from

which it synthesizes DNA repeating sequence, or "junk" DNA. This repeated sequence of DNA is important because, every time a linear chromosome is duplicated, it is shortened in length. With "junk" DNA at the ends of chromosomes, the shortening eliminates some of the non-essential, repeated sequence rather than the protein-encoding DNA sequence farther away from the chromosome end. Telomerase is often activated in cancer cells to enable cancer cells to duplicate their genomes indefinitely without losing important protein-coding DNA sequence. Activation of telomerase could be part of the process that allows cancer cells to become *immortal*. However, the true *in vivo* significance of telomerase has still not been empirically proven.^[*citation needed*]

See also

- RNA Polymerase
- Translation - process of decoding RNA to form polypeptides
- Splicing - process of removing introns from precursor messenger RNA (pre-mRNA) to make messenger RNA (mRNA)
- Reverse transcription - process viruses use to make DNA from RNA
- Crick's central dogma - DNA is transcribed to RNA, which is translated to polypeptides, never the other way around.
- Gene regulation

References

- ↑ MedicineNet.com. "Transcription definition" (<http://www.medterms.com/script/main/art.asp?articlekey=5835>) . <http://www.medterms.com/script/main/art.asp?articlekey=5835> Retrieved 11 October 2009.
- ↑ Berg J, Tymoczko JL, Stryer L (2006). *Biochemistry* (6th ed.). San Francisco: W. H. Freeman. ISBN 0716787245.
- ↑ Littlefield, O., Korkhin, Y., and Sigler, P.B. (1999). "The structural basis for the oriented assembly of a TBP/TFB /promoter complex". *PNAS* **96** (24): 13668–13673. doi:10.1073/pnas.96.24.13668 (<http://dx.doi.org/10.1073%2Fpnas.96.24.13668>) PMC 24122 (<http://www.pubmedcentral.gov/articlerender.fcgi?tool=pmcentrez&artid=24122>) PMID 10570130 (<http://www.ncbi.nlm.nih.gov/pubmed/10570130>) .
- ↑ Hausner, W; Thomm, M (2001). "Events during Initiation of Archaeal Transcription: Open Complex Formation and DNA-Protein Interactions". *Journal of Bacteriology* **183** (10): 3025–3031. doi:10.1128/JB.183.10.3025-3031.2001 (<http://dx.doi.org/10.1128%2FJB.183.10.3025-3031.2001>) PMC 95201 (<http://www.pubmedcentral.gov/articlerender.fcgi?tool=pmcentrez&artid=95201>) PMID 11325929 (<http://www.ncbi.nlm.nih.gov/pubmed/11325929>) .
- ↑ Qureshi, SA; Bell, SD; Jackson, SP (1997). "Factor requirements for transcription in the archaeon *Sulfolobus shibatae*". *EMBO Journal* **16** (10): 2927–2936. doi:10.1093/emboj/16.10.2927 (<http://dx.doi.org/10.1093%2Femboj%2F16.10.2927>) PMC 1169900 (<http://www.pubmedcentral.gov/articlerender.fcgi?tool=pmcentrez&artid=1169900>) PMID 9184236 (<http://www.ncbi.nlm.nih.gov/pubmed/9184236>) .
- ↑ {{cite book|last=Raven|first=Peter H.|title=Biology: ninth edition|year=2011|publisher=McGraw-Hill|location=New York|isbn=9780073532226|pages=278-301}}
- ↑ Mohamed Ouhammouch, Robert E. Dewhurst, Winfried Hausner, Michael Thomm, and E. Peter Geiduschek (2003). "Activation of archaeal transcription by recruitment of the TATA-binding protein". *Proceedings of the National Academy of Sciences of the United States of America* **100** (9): 5097–5102. doi:10.1073/pnas.0837150100 (<http://dx.doi.org/10.1073%2Fpnas.0837150100>) PMC 154304 (<http://www.pubmedcentral.gov/articlerender.fcgi?tool=pmcentrez&artid=154304>) PMID 12692306 (<http://www.ncbi.nlm.nih.gov/pubmed/12692306>) .
- ↑ Goldman, R.; Ebright, H.; Nickels, E. (May 2009). "Direct detection of abortive RNA transcripts in vivo". *Science* **324** (5929): 927–928. Bibcode 2009Sci...324..927G (<http://adsabs.harvard.edu/abs/2009Sci...324..927G>) doi:10.1126/science.1169237 (<http://dx.doi.org/10.1126%2Fscience.1169237>) ISSN 0036-8075 (<http://www.worldcat.org/issn/0036-8075>) PMC 2718712 (<http://www.pubmedcentral.gov>) .

- /articlerender.fcgi?tool=pmcentrez&artid=2718712) . PMID 19443781 (<http://www.ncbi.nlm.nih.gov/pubmed/19443781>) .
9. ^ Dvir, A (Sep 2002). "Promoter escape by RNA polymerase II". *Biochimica et Biophysica Acta* **1577** (2): 208–223. ISSN 0006-3002 (<http://www.worldcat.org/issn/0006-3002>) . PMID 12213653 (<http://www.ncbi.nlm.nih.gov/pubmed/12213653>) .
 10. ^ Richardson J. Rho-dependent termination and ATPases in transcript termination. *Biochimica et Biophysica Acta* (BBA) - Gene Structure and Expression. 2002;1577(2):251-260. Available at: [http://dx.doi.org/10.1016/S0167-4781\(02\)00456-6](http://dx.doi.org/10.1016/S0167-4781(02)00456-6) [Accessed March 5, 2011].
 11. ^ Lykke-Andersen S, Jensen TH. Overlapping pathways dictate termination of RNA polymerase II transcription. *Biochimie*. 2007;89(10):1177-82. Available at: <http://dx.doi.org/10.1016/j.biochi.2007.05.007> [Accessed August 5, 2010].
 12. ^ Raj, A. and van Oudenaarden, A. (2008). Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* 135, 216-26.
 13. ^ "Chemistry 2006" (http://nobelprize.org/nobel_prizes/chemistry/laureates/2006/) . *Nobel Foundation*. http://nobelprize.org/nobel_prizes/chemistry/laureates/2006/ . Retrieved 2007-03-29.
 14. ^ Kolesnikova I. N. (2000 г.). "Some patterns of apoptosis mechanism during HIV-infection" (<http://www.dissercat.com/content/nekotorye-osobennosti-mekhanizmov-apoptoza-pri-vich-infektsii>) (in ru). *Dissertation*. <http://www.dissercat.com/content/nekotorye-osobennosti-mekhanizmov-apoptoza-pri-vich-infektsii> . Retrieved 2011-02-20.

External links

- Interactive Java simulation of transcription initiation. (<http://cmol.nbi.dk/models/dynamtrans/dynamtrans.html>) From Center for Models of Life (<http://cmol.nbi.dk/>) at the Niels Bohr Institute.
- Interactive Java simulation of transcription interference--a game of promoter dominance in bacterial virus. (<http://cmol.nbi.dk/models/dna/rnap.html>) From Center for Models of Life (<http://cmol.nbi.dk/>) at the Niels Bohr Institute.
- Biology animations about this topic under Chapter 15 and Chapter 18 (<http://highered.mcgraw-hill.com/sites/dl/free/0072437316/120060/ravenanimation.html>)
- Virtual Cell Animation Collection, Introducing Transcription (<http://vcell.ndsu.nodak.edu/animations/transcription/index.htm>)

Retrieved from "[http://en.wikipedia.org/wiki/Transcription_\(genetics\)](http://en.wikipedia.org/wiki/Transcription_(genetics))"

Categories: Gene expression | Molecular biology

- This page was last modified on 5 June 2011 at 19:36.
 - Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. See Terms of Use for details.
- Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.