RNA Structure, Function, and Synthesis

RNA

RNA differs from DNA in both structural and functional respects. RNA has two major structural differences: each of the ribose rings contains a 2´-hydroxyl, and RNA uses **uracil** in place of **thymine**. RNA molecules are capable of base pairing, but generally will not form large regions of stable RNA-RNA double helix. RNA can act as a genetic material (although this role, at least for current organisms, seems to be restricted to viruses).

Unlike DNA, RNA can form complex three-dimensional structures. As a result, RNA can also exhibit catalytic activity. The combination of the ability to store genetic information with the ability to catalyze reactions has resulted in a proposal for the origin of life: the "**RNA World**". The RNA world hypothesis proposes that RNA molecules once filled all of the roles of protein and nucleic acid macromolecules, and acted in both an information storage capacity and as the source of the enzymatic activity required for metabolic reactions.

In general, RNA is less suited to acting as genetic material than DNA, and is less suited to forming efficient catalysts than proteins. Assuming that the RNA world once existed, nearly all of its functions have been taken over by other biological molecules. However, some vestiges of the RNA world may still exist.

The vast majority of RNA functions are concerned with protein synthesis.

The major types of RNA:

Ribosomal RNA (rRNA)

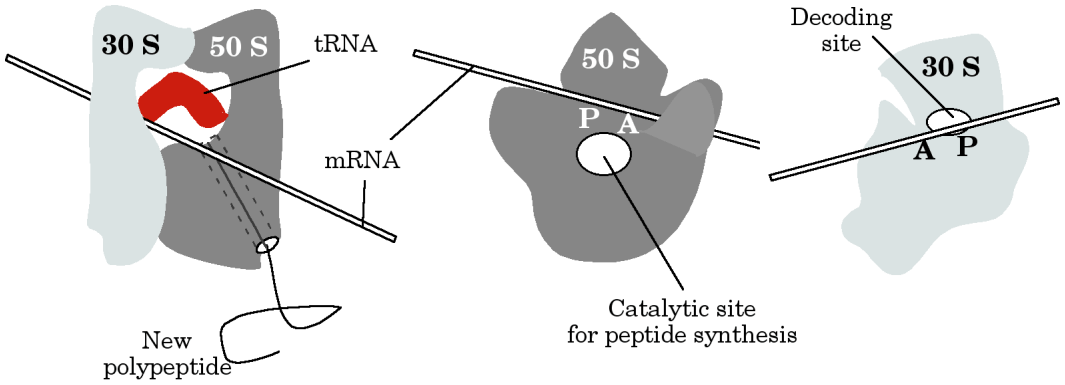
Ribosomal RNA molecules comprise 65 to 70% of the mass of the ribosome (the machinery responsible for protein synthesis). Ribosomes are very large objects; prokaryotic ribosomes have molecular weights of about 2.5 million, while eukaryotic ribosomes have molecular weights of about 4 million.

> The original studies on ribosomes used relatively crude techniques that were unable to measure size in terms of molecular weight. Instead the size of the ribosomal particles and their components were measured by their rate of sedimentation (movement driven by gravitational acceleration or centrifugal acceleration). Sedimentation is a function of size, shape, and density, with larger objects tending to sediment faster than smaller ones.
>
> Object sizes are measured in Svedberg units. Prokaryotic ribosomes are 70 S particles, with each comprised of a large (50 S) and a small (30 S) subunit. Eukaryotic ribosomes are 80 S particles, comprised of a large (60 S) and a small (40 S) subunit. You will notice that the Svedberg units are not additive for the particles sizes; this is due to the effects of shape on sedimentation.

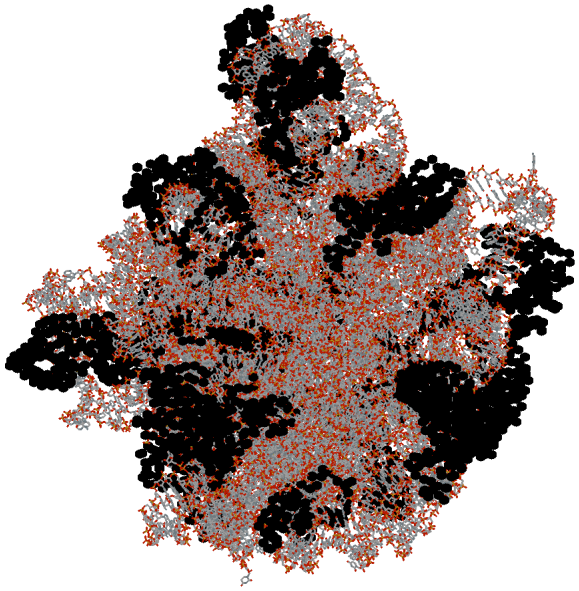
The eukaryotic 40S Ribosome contains 1 rRNA (18 S rRNA = 1900 bases) and about 35 different proteins. The 60S ribosome contains 3 rRNA (5 S = 120 bases, 5.8 S = 160 bases, and 28 S = 4700 bases), and about 50 proteins. The 5 S rRNA has its own gene; the others are synthesized as a single transcript that is then cleaved to release the mature RNA molecules that become part of the ribosome.

Copyright © 1999 – 2003 by Mark Brandt, Ph.D.

Until relatively recently, it was assumed that the ribosomal RNA performed a largely structural function. However, more recent data strongly suggests that the rRNA acts as the enzyme, with the protein acting as the structural scaffolding. These data include results from the recent high-resolution (2.4 Å) X-ray diffraction structure of the large subunit and low-resolution (5 Å) structure of the complete ribosome from the bacterium *Haloarcula marismortui*.



Examination of the high-resolution structure of the large subunit, and of the lower resolution structures of the entire particle (used to generate the cartoons above) strongly suggests that only RNA is present at the catalytic site. In the structure below, proteins are shown in black, and the orientation of the large subunit is similar to the center cartoon above. Examination of this structure suggests that no protein is present in the catalytic site.

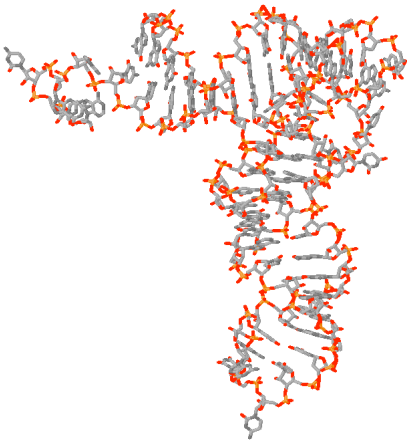


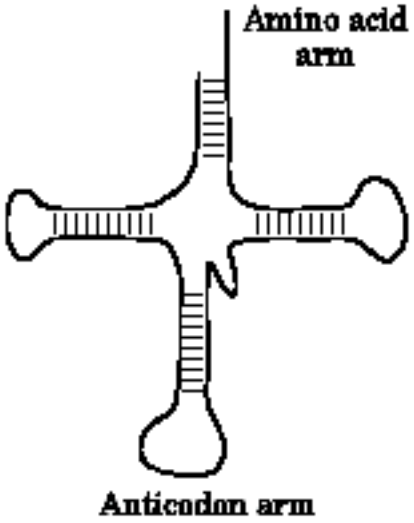
Transfer RNA (tRNA)

tRNA is a ~75 base molecule that carries the amino acids, and transfers them to the growing protein. tRNAs are thought to have a common tertiary structure (a

Copyright © 1999 – 2003 by Mark Brandt, Ph.D.

structure based on X-ray diffraction analysis is shown below. Analysis of the tRNA sequence suggests a cloverleaf secondary structure formed by regions of base pairing between the sections of the RNA strand, with this cloverleaf folding into the three-dimensional structure.





Messenger RNA (mRNA)

mRNA molecules contain the coding sequence for proteins. The mRNA molecules can vary considerably in size, with eukaryotic transcripts including the largest known ribonucleic acids. This is most obvious before splicing of introns, because many transcripts exceed 100 kb in length.

RNA bases

The bases used for RNA are attached to ribose. However, many are significantly modified from the typical four bases normally considered to be part of RNA. This is particularly true for tRNA. The modified bases include pseudouracil and methylated versions of cytosine and adenine.

Transcription

Transcription is the process of RNA synthesis using a DNA template.

Terminology:

Gene: a stretch of DNA containing both a template for RNA synthesis and sequences that allow the control of RNA production from the template region. When the mechanisms for protein synthesis were originally worked out, it was suggested that each gene corresponded to a single protein. This is clearly not true. Some genes (such as ribosomal RNA genes) do not code for proteins. In addition, alternate splicing appears to be a common phenomenon in higher eukaryotes, and current analyses suggest that the average human gene is the source of at least two different proteins.

Complementary: having the base sequence that allows base pairing to another sequence. The sequence 5´-TACTGGT is complementary to the sequence 5´-ACCAGTA, because each base in one sequence can base pair to the corresponding base in the other sequence.

Copyright © 1999 – 2003 by Mark Brandt, Ph.D.

Coding strand: the DNA sequence that corresponds to the RNA sequence of the transcript. The coding strand has the same sequence as the RNA (with T instead of U, and, in the case of tRNA molecules, standard bases instead of the modified bases found in the RNA).

Template strand: the DNA sequence that the polymerase actually uses to guide the synthesis of the growing RNA strand. The template strand is complementary to both the RNA strand and the DNA coding strand.

Upstream: on the 5´side of any given position on the coding strand.

Downstream: on the 3´side of any given position on the coding strand.

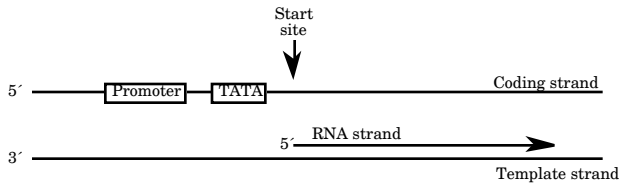
Consensus sequence: DNA sequences with regulatory functions usually have similar sequences (at least within a class – TATA boxes are similar to one another, but are quite different from promoter elements). However, the regulatory DNA sequence of one gene is rarely identical to the equivalent sequence in another gene. Studying a number of these sequences may make it possible to guess at sequence features that are most likely to occur. These are chosen by consensus (*i.e.* by their higher probability of being present).

Promoter: a DNA sequence recognized by the transcription initiation complex.

TATA box: one part of the promoter region. The TATA box contains a sequence with a high AT content. The high AT content allows easier strand separation. In addition, the DNA helix has a subtly different structure in AT-rich regions, which may make it easier for the RNA polymerase to find the correct starting place.

Initiation of transcription is a complex process

General concept:
1) Transcription factor proteins bind to the promoter element upstream of the coding sequence.
2) The DNA-dependent RNA polymerase binds to the promoter/protein complex.
3) The RNA polymerase complex separates the DNA strands.
4) The RNA polymerase begins synthesizing RNA.
Note that, unlike DNA polymerases, RNA polymerases do not need primers. Transcription only requires binding of the RNA polymerase to the promoter.



RNA Polymerase

E. coli has a single RNA polymerase responsible for all types of gene transcription. Humans have three transcription polymerases. RNA polymerase I transcribes

Copyright © 1999 – 2003 by Mark Brandt, Ph.D.

rRNA, RNA polymerase II transcribes hnRNA, which are the source of the mRNA used as templates for protein synthesis, and RNA polymerase III transcribes tRNA, and 5S rRNA, and a few other small RNA molecules. The RNA polymerases are large multiprotein complexes with about 10 subunits.

Regulation of transcription

Prokaryotes

Prokaryotes need only to decide whether or not to transcribe each gene based on the functions of the gene, and on the current environment.

Many genes are controlled by repressors.

The best understood prokaryotic gene control system is the *E. coli* ***lac*** **operon** (an operon is a series of genes [usually of related function] transcribed into a single RNA, from which multiple proteins can then be made). The genes of the *lac* operon are only needed if lactose is present in the environment. In the absence of lactose, the lac repressor binds to the promoter, and prevents transcription. When lactose is present, it binds to the repressor; the repressor then dissociates from the DNA, allowing transcription of the genes. However, if glucose levels are also high, the resulting low levels of cAMP do not allow cAMP receptor protein to bind to the CBP site, and transcription will not occur, because CBP/cAMP complex is necessary to recruit RNA polymerase.

Humans

RNA polymerase I and RNA polymerase III are not especially tightly regulated, since all cells need their products. (These two enzymes do have specialized transcription factors generally similar to those used by RNA polymerase II.) In contrast, RNA polymerase II, which generates the mRNA used for protein synthesis, is heavily regulated.

Multicellular organisms need an additional level of control – different cell types, which all contain the same genome, need to express different genes even under the same conditions. This is achieved by having multiple control elements, including different types of promoters and enhancers. Most of these control elements require specific transcription factors; during and after differentiation, cells control the transcription factors that are expressed and therefore limit the genes that are expressed.

DNA modification (possibly including methylation patterns) and structural organization into chromatin also assist in regulating gene expression.

Nucleosomes

Nucleosomes serve two functions: they assist in the packing of long molecules of DNA into a small volume (*i.e.* the 2 meters of DNA present in each cell must be packed into the small size of the nucleus of the cell), and they decrease the ability of the transcriptional machinery to interact with the DNA.

Nucleosomes consist of 146 bp of DNA wrapped around a protein complex. The proteins in the complex are called histones. **Five histones** are known; 1, 2A, 2B, 3,

Copyright © 1999 – 2003 by Mark Brandt, Ph.D.

and 4. Histone 1 is found associated with the DNA outside of the nucleosome, while the other four form an octameric complex that the DNA wraps around (the DNA encircles the octamer 1.7 times).

Genes that are actively transcribed have **fewer nucleosomes**, and more exposed DNA.

Basal transcription

Basal transcription has some similarities to prokaryotic transcription.

1) Transcription factors bind to the promoter. TFIID (TF = transcription factor, II = Pol II, D for the protein designation) binds to the TATA box
2) Other transcription factors bind and stabilize the complex.
3) RNA Polymerase II binds to the complex, and initiates transcription.

For some genes, basal transcription levels are significant (these are typically genes required by all cells), usually because the gene has a strong promoter.

Enhancers

For most genes, however, other control elements are required. Enhancers are DNA sequences present within about 10 kb of the transcription start sites. Enhancers can greatly alter transcription rates. In most cases, enhancer sequences interact with regulated transcription factors (such as steroid receptors, among *many* others). The enhancer-transcription factor complex then stabilizes the binding of transcription factors to the promoter region, and stimulates binding of the RNA polymerase. (In some cases, enhancer binding turns off transcription.)

Enhancers can function at a significant distance from the promoter; this is probably due to the formation of loops in the DNA, which allow direct contact between the protein complexes present at the enhancer(s) and the RNA polymerase II recruitment complex present at the promoter.

Copyright © 1999 – 2003 by Mark Brandt, Ph.D.