

The structure and function of DNA

BY ALLEN F RYAN

DNA structure and replication

Genetic information within multicellular organisms, including man, is stored in molecules of deoxyribonucleic acid (DNA), which reside within the chromosomes of each cell nucleus. A DNA molecule consists of two very long chains, or strands, of modified sugar molecules known as nucleotides. The two strands of the DNA molecule are connected to each other by chemical bonds between nucleotides, and form a spiral: the linked strands of the double helix. It is the nature of the bonds between strands that endow DNA with its unique ability to replicate itself.

The four nucleotides of DNA are adenine, cytosine, guanine and thymine (A, C, G & T). The sequence of these four nucleotides within a DNA strand can occur in any order, and that sequence makes up the genetic code. However, the nucleotides of opposite strands can only bind with one partner: A with T, or G with C. Thus

the two strands of DNA are different, but complementary to each other, in sequence. The two DNA strands are separated during cell division, and each strand then serves as a template to which free nucleotides bind. Each nucleotide finds its matching partner, so that the missing strand is replicated. DNA synthesis stitches the free nucleotides together, resulting in the formation of two new strands. The final result is two identical copies of the double helix, one for each new cell.

This process of replication is not flawless. Approximately one time in every billion nucleotide replication, an error occurs: an inappropriate nucleotide may be inserted, an extra one is inserted, or one may be skipped. This results in a change to the genetic code: a mutation. If this happens during cell divisions that result in an egg or sperm, that mutation can be passed on to the next generation (Figure 1).

The genetic code

The genetic code of DNA determines what proteins can be produced by a cell, as well as the rate and time of protein production. Proteins consist of strings of amino acids, and the identity of each amino acid is encoded by a sequence of three DNA nucleotides known as a codon. A protein is produced when another type of molecular chain, ribonucleic acid (RNA), is copied from DNA. In this process, the two DNA strands separate temporarily, and a different type of nucleotide binds to the DNA strand, with each RNA nucleotide recognising a specific DNA nucleotide. Copying of DNA into RNA is known as transcription. The resultant messenger RNA (mRNA) strand does not stay bound to the DNA, however. It is released, and another form of RNA, transfer RNA (tRNA) binds to each three-nucleotide codons of the mRNA. tRNAs have a specific amino acid bound to one

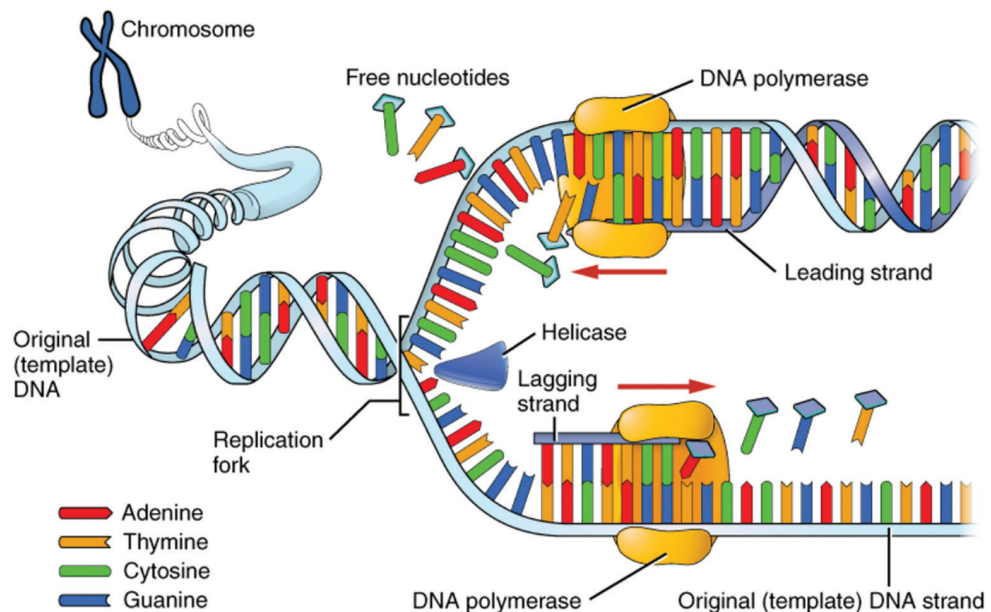


Figure 1: DNA replication. During the process of DNA replication, the original DNA molecule is uncoiled by DNA helicase, separating the two strands. Each then forms the template upon which DNA polymerase synthesises a new double helix, with complementary nucleotide pairs ensuring faithful replication of the original DNA sequence. (Figure courtesy of OpenStax College, The nucleus and DNA replication, <http://cnx.org/content/m46073/1.4/>)

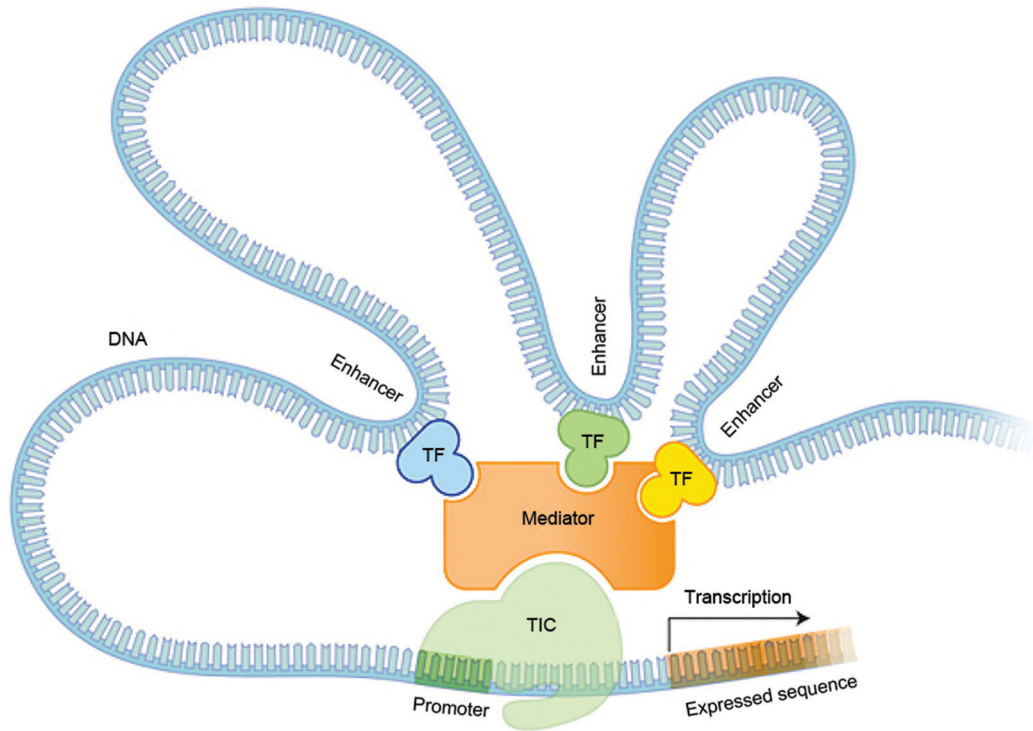


Figure 2. Gene transcription and regulation. During transcription of a gene into mRNA, the transcription initiation complex (TIC) consisting of Pol II and associated DNA binding proteins, binds to the DNA of a gene at the promoter, adjacent to the expressed (mRNA coding) sequence. The mediator complex connects the TIC to regulatory transcription factors (TFs) that are bound to enhancer or repressor DNA elements, which can be distant from the promoter. The particular TFs that bind to the mediator help to regulate in which cells genes are transcribed, as well as the time and rate of transcription. (Figure adapted from Nature Education, Gene expression.)

end, and a complementary codon for that amino acid at the other. They bind sequentially to the mRNA, the amino acids are linked by a protein-synthesizing enzyme, and the resulting protein is formed from the correct sequence of amino acids. There are also stop codons that tell tRNAs to cease translating when the protein is complete. Thus the DNA code is faithfully transcribed into mRNA and then further translated into protein, mediated by the recognition of nucleotides from DNA to mRNA to tRNA. Taken together, these processes make up gene expression.

The structure of genes

Each protein that can be produced from DNA is encoded in a gene. The parts of the gene that can produce protein are known as exons. There are usually several or even many exons per gene, separated by noncoding DNA sequences known as introns. When mRNA is produced, the introns are removed to produce a single, long strand of codons. By using different combinations of exons, more than one form of a protein can be produced from the same gene. Other parts of the gene are not used as a template for

mRNA. They function by controlling the production of mRNA. This control is also mediated by the sequence of the DNA in these noncoding regions.

The human genome encodes approximately 29,000 genes. However, the great majority of our 3-billion nucleotide long genome does not consist of protein coding genes, which make up only 1-2% of our DNA. Noncoding DNA may be regulatory in nature, nonfunctional, or have functions that we do not yet understand.

The human genome is broken up into sections, known as chromosomes. Moreover, each of our 23 chromosomes comes in pairs: one inherited from our mother and one from our father. Exceptions are the X and Y chromosomes of males, who have only one copy of each, and therefore only one copy of each X or Y chromosome gene. (Females have two copies of the X chromosome, and no Y chromosome.) Thus, almost all of our genes exist as a maternal and a paternal copy, which can be different from one another. For example, in the case of eye colour, a child can inherit a gene for blue eyes from one parent, and brown eyes from the other. Because the brown eye copy is dominant, the child

will have brown eyes.

Gene regulation

DNA is copied into mRNA by a molecule called RNA polymerase II (Pol II), which binds to the DNA of a gene adjacent to the sequence encoding the mRNA. Pol II cannot operate alone, but requires additional protein partners. These include a number of DNA binding proteins known as transcription factors that recognise and bind to specific DNA sequences in the gene. These include the proximal promoter, which lies immediately adjacent to the initial site of mRNA transcription and to which RNA polymerase binds. Several transcription factors bind to the proximal promoter with Pol II to form a transcription initiation complex. Another important component of transcription is a molecular complex called mediator, which binds directly to Pol II and is required for the transcription initiation complex (TIC) to operate. Additional non coding areas of the gene can act to increase or decrease the activity of the proximal promoter. They are known as enhancer or repressor elements. The mediator complex has binding sites for many transcription

factors that bind specifically to enhancer and repressor elements of the gene. Binding to these regulatory transcription factors appears to alter the shape of the mediator, which can in turn radically alter the activity of the TIC. The mediator thus acts as a bridge between enhancer / repressor regions, regulatory transcription factors and the TIC. The regulatory sequences present in the gene, and the specific transcription factors present in the cell, comprise yet another code that allows different cells to produce different proteins, and to produce different proteins at different times.

Another form of gene regulation is controlled by DNA modifications that can either expose DNA for transcription or make it unavailable. Addition of too many methyl groups to DNA can render it incapable of binding transcription factors, or can attract DNA silencing proteins. Acetylation or methylation of histones, proteins around which DNA can coil, can similarly influence the accessibility of genes for transcription. Silencing of genes by these epigenetic mechanisms is an important form of long-term gene regulation. The mediator complex can also alter histone architecture, providing an additional regulatory mechanism (Figure 2).

Protein production can also be regulated by genes that do not themselves encode proteins. Rather they encode short RNAs that are complementary to sequences in mRNAs. If they bind to an mRNA, the resultant double-stranded RNA will be degraded, halting translation of protein. Production of these microRNAs is another important mechanism of gene control.

Inheritance and mutations

DNA replication is required so that each cell within an organism possesses the same genetic code. Starting from a single fertilised egg, billions of cell divisions are required to form an individual. The genetic code is also the basis for reproduction. Passing genetic information from one generation to the next ensures that the traits of the parents are passed faithfully to their offspring. In the sex organs, cell division occurs in such a manner that only one chromosome from each pair is passed on to a gamete, which is either an egg or sperm. During fertilisation, these gametes combine to produce a single cell with the normal number of chromosomes. Thus the genetic code of the offspring gets equal contributions from each parent. Which of the two chromosomes from each pair that a parent passes on to an

individual gamete is random, so there are many possible combinations of genes that can come from any set of parents.

As noted above, replication of DNA can result in mutations. If this occurs in divisions leading to a gamete, the defect in DNA can be passed on to the offspring. Mutations in our DNA can have many consequences, from no effect whatsoever to devastating changes in health, or even death. This depends largely upon where in the genome the mutation occurs. If it occurs in a protein-coding sequence or an important regulatory sequence, the likelihood of a health effect is greatest. A mutation in nonfunctional DNA would have little or no effect. In rare cases, a mutation is beneficial, and will be passed on to future generations, which forms the basis for evolution.



Allen F Ryan,

PhD,
Departments of Surgery/
Otolaryngology,
UCSD School of Medicine
and VA Medical Center,
La Jolla, CA 92093-0666,
USA

E: afryan@ucsd.edu

**Declaration of Competing
Interests**
None declared