**Topic:** Stages of gene expression

Objectives of the class:

To get acquainted with gene expression - the transfer of genetic information from DNA to a trait;

tо learn what the genetic code is, how genetic information is encoded on DNA;

to study the processes of RNA transcription and protein translation.

Why should a doctor know about genes?

The importance of genetics for medicine is huge. There are more than 4,000 forms of inherited diseases in human populations. About 5% of children are born with hereditary or congenital diseases.

Knowledge of the molecular basis of genetics will allow:

to study the hereditary mechanisms that ensure human health;

to provide diagnostics, treatment and prevention of hereditary diseases, etc.

*Take your notebooks. Let's start compiling a biological dictionary for today's lesson. Write it down.*

What is a gene?

Gene is a part of a chromosome. Genes are made up of nucleotides, units of DNA. A gene is the basic functional unit of heredity, they are inherited from parents. They control the development of all living organisms.

Some genes act as instructions to make molecules called proteins. However, many genes do not code for proteins.

In eukaryotes (such as animals, plants, and fungi), genes are contained within the cell nucleus.

Every person has two copies of each gene, one inherited from each parent. Most genes are the same in all people, but a small number of genes (less than 1 percent of the total) are slightly different between people. Alleles are forms of the same gene with small differences in their sequence of DNA bases. These small differences contribute to each person’s unique physical features.

*Write it down.*

Genome

Genome is the totality of nuclear and cytoplasmic (the mitochondrial or the chloroplast) DNA in the cell.

The genome includes both the genes and the non-coding sequences of DNA.

Also, the total genetic sequence in the single chromosome of a bacteria, or the sequence of RNA in an RNA virus.

In humans genes are more than 2 million bases. The Human Genome Project estimated that humans have between 20,000 and 25,000 genes.

*Write it down.*

Genotype

Genotype is the genetic set of genes of an individual organism or cell.

The genotype determines the hereditary potentials and limitations of an individual from embryonic formation through adulthood. Each individual have a unique genotype.

The genotype under the environment determines individual’s appearance and behavior.

Molecular basis of heredity and variability.

Nucleic acid is a high molecular weight organic compound, a biopolymer (polynucleotide) formed by nucleotide residues. Nucleic acids DNA and RNA are present in the cells of all living organisms and perform the most important functions of storing, transmitting and realizing hereditary information.

Nucleotide - a nucleic acid monomer

Each nucleotide is built from a nitrogenous base (purine or pyrimidine), a carbohydrate pentose (ribose or deoxyribose), which has five carbon atoms, and a phosphoric acid residue.

Phosphodiester bond

Phosphodiester bonds are central to all life on Earth as they form the backbone of the nucleic acid strands.

Phosphodiester bonds is formed between the phosphoric acid residue of one nucleotide and the sugar residue of another nucleotide.

Deoxyribonucleic acid, or DNA, is a biological macromolecule that carries hereditary information in many organisms.

DNA is a double-helix polymer of nucleotides. The nitrogenous bases in DNA are of four types – adenine, guanine, thymine and cytosine.   
They make up the genetic code.

Two DNA strands are linked by hydrogen bonds formed between complementary nitrogenous bases. Adenine is complementary to thymine, and guanine is complementary to cytosine. Two hydrogen bonds are formed between adenine and thymine, and three are formed between guanine and cytosine.

RNA, ribonucleic acid, complex compound of high molecular weight that functions in cellular protein synthesis and replaces DNA as a carrier of genetic codes in some viruses. RNA consists of nucleotides, forming strands of varying lengths. The nitrogenous bases in RNA are adenine, guanine, cytosine, and uracil, which replaces thymine in DNA.

*Write down a table comparing DNA and RNA in a notebook.*

Of the many types of RNA, three are the most well-known. These are information RNA (mRNA), transport RNA (tRNA), and ribosomal RNA (rRNA). They are present in all organisms.

In protein synthesis, mRNA carries genetic codes from the DNA in the nucleus to ribosomes, the sites of protein translation in the cytoplasm. Ribosomes are composed of rRNA and protein.

Molecules of tRNA, which contain fewer than 100 nucleotides, bring the specified amino acids to the ribosomes, where they are linked to form proteins.

The central dogma of molecular biology is the rule observed in nature for the implementation of genetic information: information is transmitted from nucleic acids to a protein, but not in the opposite direction. The rule was formulated by Francis Crick in 1958. The transfer of genetic information sequentially from DNA to RNA and then from RNA to protein is universal for all cellular organisms without exception, and underlies the biosynthesis of macromolecules. Genome replication corresponds to the informational transition DNA → DNA.

The genetic code and his properties

Information on DNA is organized in a special way. This is called the genetic code.

The genetic code is made up of codons, which are three-letter chains of nucleotides. Each codon codes for one specific amino acid. The code defines the order in which amino acids are added to the polypeptide chain during protein synthesis. Hence, the genetic code dictates the sequence of amino acids in a protein.

The genetic code has properties that prove the unity of the origin of life on Earth.

*Write it down.*

1. Triplet:

The coding units or codons of amino acids are composed of three consecutive nucleotides

2. Degeneracy

most amino acids have more than one codon.

3. Non-overlap

The same base is used only when one codon is formed.

4. Continuity:

There is no gap between two codons

5. Unambiguity

One codon always codes for a specific amino acid

6. Versatility

The genetic code is universal for all types of living organisms: viruses, prokaryotes and eukaryotes.

7. Collinearity

DNA is a linear polynucleotide chain and protein is a linear polypeptide chain. The amino acid sequence in a polypeptide chain corresponds to the sequence of nucleotide bases in the gene (DNA) that encodes it.

8. Parity

A specific gene transcribes a specific mRNA that produces a specific polypeptide. Based on this, a cell can have as many types of polypeptides as it has types of genes.

DNA replication

Replication is a self-copying mechanism and the main property of hereditary material, which is DNA molecules.

A feature of DNA is that usually its molecule consists of two complementary strands that form a double helix. In the process of replication, the chains of the parent DNA molecule diverge, and a new complementary chain is built on each. As a result, from one double helix, two are formed, identical to the original. That is, from one DNA molecule, two are formed, identical to the matrix and to each other.

Thus, DNA replication occurs in a semi-conservative way, when each daughter molecule contains one mother chain and one newly synthesized one.

The divergence of the chains of the original DNA molecule is provided by the helicase enzyme, which in certain places on the chromosomes breaks hydrogen bonds between the nitrogenous bases of DNA. Helicases move along DNA with the expenditure of ATP energy.

To prevent the chains from joining again, they are held at a distance from each other by destabilizing proteins. The result is replication zones called replication forks.

Replication forks are not formed anywhere in DNA, but only at the origin of replication, which consists of a specific sequence of nucleotides. A replication eye is formed, in which two DNA strands diverge.

Each individual DNA strand of the old molecule is used as a template for the synthesis of a new complementary strand. The addition of nucleotides to the growing daughter chain is provided by the enzyme DNA polymerase.

In the replication fork, free nucleotides in the nucleoplasm are attached to the liberated hydrogen bonds of the chains according to the principle of complementarity.

Polymerase not only lengthens the growing chain, but is also capable of detaching erroneous nucleotides, i.e., it has a corrective ability.

The synthesis of a new DNA strand proceeds in the direction from the 5 'to the 3' end.

Since DNA strands are antiparallel, and the synthesis of a new strand is possible only in the 5´ → 3´ direction, the daughter chains in the replication fork will be synthesized in different directions.

On the 3´ → 5´ matrix, the assembly of a new polynucleotide sequence occurs continuously, in the 5´ → 3´ direction. The antiparallel matrix is ​​characterized by a 5´ → 3´ direction, so the synthesis of the daughter chain along the fork is not possible here.

Therefore, synthesis on the 5´ → 3´ matrix is ​​performed in small sections - Okazaki fragments in the opposite direction of the replication fork.

Continuous build is faster than fragmentary build. Therefore, one of the daughter DNA strands is called the leading, the second - lagging.

**Gene expression**

This is the implementation of hereditary information from gene to trait.

In this case, a trait should be understood as the result of biochemical reactions involving proteins synthesized from the DNA (gene) matrix.

Expression stages differ between prokaryotes and eukaryotes.

In prokaryotes, in the cytoplasm of the cell, transcription first takes place - the transfer of information from DNA to RNA, then translation - protein synthesis. Simultaneously with these processes, constantly, there is an activation (attachment of transport RNA) of amino acids and their transport to ribosomes - the site of protein synthesis.

In eukaryotes, transcription takes place in the cell nucleus. Since their gene has an exon-intron structure, processing also takes place in the nucleus - the stage of maturation of messenger RNA.

Then the messenger RNA leaves the nucleus, and the stages of amino acid activation and translation take place in the cell cytoplasm.

**Transcription**

Transcription is the first step in gene expression that involves the formation of an RNA molecule from DNA.

(17) The transcription consists of three stages

initiation

elongation

termination

Transcription begins when DNA dependent RNA polymerase binds to a promoter. The attachment site is the TATA box

Elongation

After RNA polymerase is attached, the double strand of DNA unwinds and the strands are detached from each other. Matrix, that is, the one on which RNA is synthesized, the chain becomes the one that has the direction 3` - 5`. RNA is always synthesized in the 5`-3` direction and becomes an exact copy of the semantic (codogenic) DNA strand, which also has a 5`-3` direction. With the exception that instead of the thymine nucleotide, uracil is included in the RNA.

Termination - end of transcription

Processing

Since the structural gene of eukaryotes contains uninformative regions, the result of transcription is an mRNA precursor.

Its complete transformation into mRNA is called processing.

Processing includes:

End splitting protection

Capping. A methylated nucleotide, CAP, is attached to the 5 'end of the messenger RNA precursor.

Polyadenylation. About 200 adenyl nucleotides are attached to the 3 'end.

As a result, mRNA is protected from the action of endonucleases that break down nucleic acids.

Splicing

During splicing, introns (non-coding regions) are removed and exons (coding regions) are spliced.

Translation - Protein Synthesis

Translation is the process in which ribosomes in the cytoplasm or ER synthesize proteins after the process of transcription of DNA to RNA in the cell's nucleus.

(27) Translation consists of three phases

initiation

elongation

termination

(28) Translation takes place with the help of special organioids – ribosomes.

The ribosome consists of two subunits, small and large, each of which is a molecule of ribosomal RNA, to which enzyme proteins are attached.

Initiation

The small subunit of the ribosome attaches to the 5 'end of the mRNA and moves along the RNA to the start codon of AUG. Transport RNA with the amino acid methionine binds to mRNA. The large subunit of the ribosome completes the complex in such a way that the first tRNA would be in the P-site.

Elongation.

A second tRNA is placed in the A site, whose anticodon is complementary to the second codon on the mRNA. A peptide bond occurs between the two amino acids. The ribosome shifts along the RNA strand with a step of 3 nucleotides (or 1 codon). From the A site, the tRNA moves to the P site, and the next tRNA, complementary to the third codon on the mRNA, gets into the A site. Etc.

(31) Termination

The genius code consists of 64 codons. Of these, 61 code for amino acids. Three codons (UAG-amber, UGA-opal or umber, UAA-ocher) do not encode amino acids, but denote the end of translation. These are stop codons or nonsense codons. When the stop codon reaches the A site, translation is complete, the synthesized protein leaves the ribosome.

Chargaff's rule is a nucleotide distribution in DNA strings, discovered by Austrian chemist Erwin Chargaff in early 1950s in Columbia University.

Chargaff's rule (or parity rule) holds that in double-stranded DNA molecule observed percentage base pair equality: %A equally= %T and %G = %C. This finding, with the results of x-ray diffraction analysis by Rosalind Franklin, served as one of the grounds for the Watson-Crick double helix model.

Правила Чаргаффа - основных правила распределения нуклеотидов в цепочках ДНК, открытые австрийским химиком Эрвином Чаргаффом в начале 1950-х годов в Колумбийском университете.  
  
Первое правило Чаргаффа (или первое правило четности) гласит, что в двухниточной молекуле ДНК наблюдается процентное равенство пар оснований: %A = %T и %G = %C. Это открытие, с результатами рентгеновского дифракционного анализа Розалинды Франклин, послужило одним из оснований для модели двойной спирали Уотсона-Крика.  
  
If the amount of adenin is 15%, then the amount of thymine is also 15%. In total, 30%. The total nucleotide content of the two DNA chains is 100%. Then 100 minus 30 equals 70. This is the amount of guanine and cytosine. Since the amount of guanine is equal to the amount of cytosine 70, it is divided by 2. This results in an amount of 35% guanine.

Если количество аденина равно 15%, то количество тимина тоже равно 15%. В сумме 30%. Полное содержание нуклеотидов в двух цепочках ДНК соответствует 100%. Тогда 100 минус 30 равно 70. Это количество гуанина и цитозина. Так как количество гуанина равно количеству цитозина 70 делим на 2. В результате количество гуанина равно 35%.

How to solve a problem on protein biosynthesis?

A chain of DNA, which has a direction from 5' to 3', is called kodogeneous. It contains information about a protein. The second DNA chain is the transcription chain from which mRNA is built. This chain has a direction from 3' to 5'. In other words, it is antiparallel to the codogenic chain.   
We are given the following chain  
5' ATG ACA GAG 3'  
On the principle of complementarity, we are building a second chain of DNA:  
3` TAC TGT CTC 5`  
Based on this chain, we build the mRNA on the principle of complementarity.  
Any RNA is always synthesized in a direction from 5' to 3'.  
5' AUG ACA GAG 3'.  
mRNA is a copy of a codogenic chain with one exception. Instead of thymine, uracil is involved in mRNA synthesis.  
In the genetic code table, we find an amino acid that corresponds to a triplet of nucleotides on mRNA.

Цепочка ДНК, которое имеет направление от 5` до 3`называется кодогенной. На ней записана информация о белке. Вторая цепочка ДНК, это транскрибируемая цепочка, на основании которой строится мРНК. Эта цепочка имеет направление от 3 до 5. То есть она антипараллельна кодогенной цепочке.

Нам дана следующая цепочка

5` ATG ACA GAG 3`

По принципу комплементарности строим вторую цепь ДНК:

3` TAC TGT CTC 5`

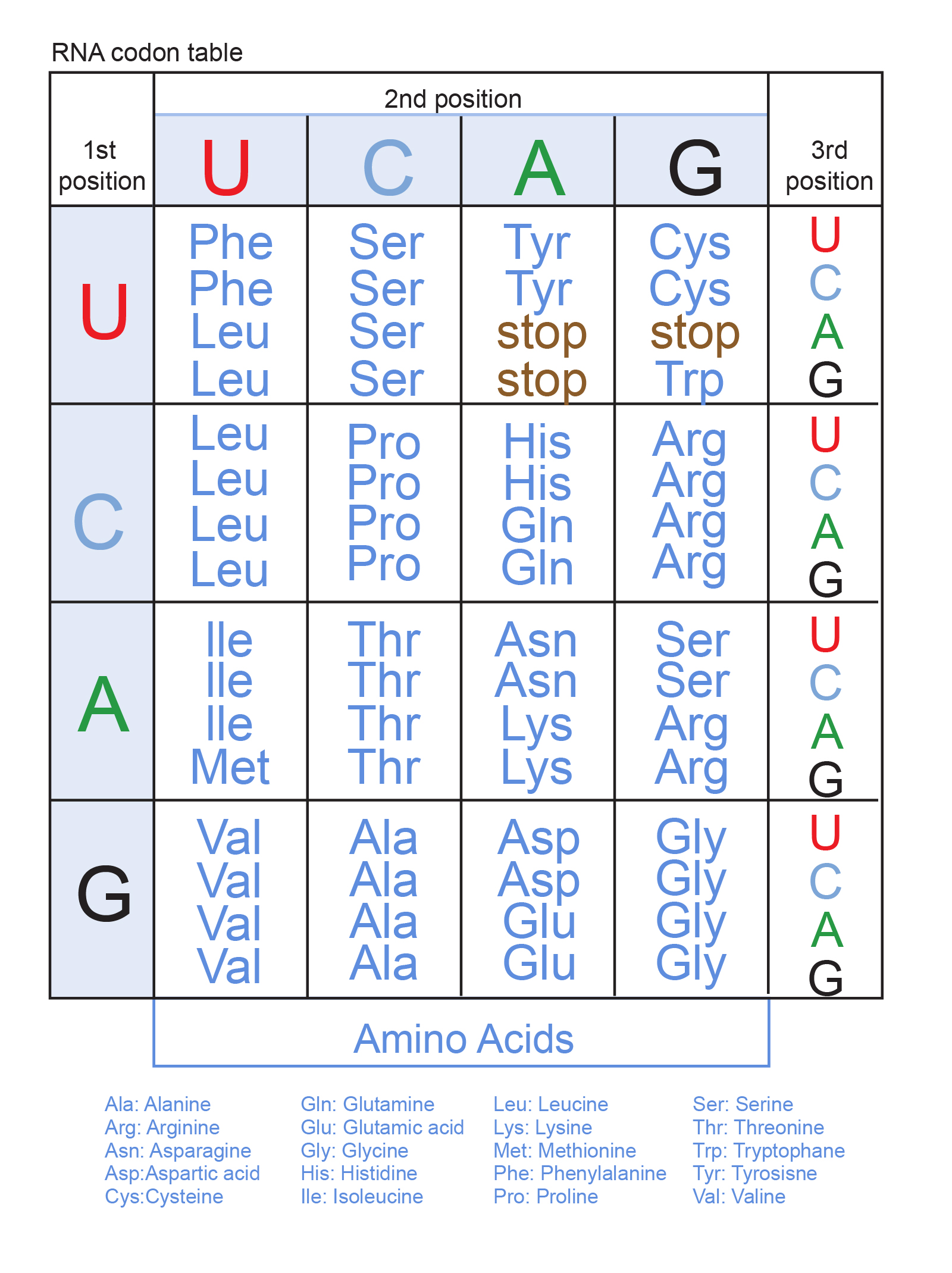
На основании этой цепи по принципу комплементарности строим мРНК.

Любая РНК всегда синтезируется в направлении от 5` до 3`.

5` AUG ACA GAG 3`

мРНК является копией кодогенной цепи с одним исключением. Вместо тимина в синтезе мРНК участвует урацил

В таблице генетических кодов находим аминокислоту, которая соответствует триплету нуклеотидов на мРНК.

In a patient with cystinuria, a group of amino acids was found in the urine, which correspond to the following triplets: mRNA: UCU UGU GCU GGU CAG CGU AAA. It is known that in a healthy person, amino acids are found in the urine: alanine, serine, glutamic acid and glycine. Build the gene for a healthy and sick person. Find the differences.

Determine which amino acids are found in the patient's urine.

There are serine, cesteine, alanine, glicine, glutamine, arginine, lysine

Let's compare it with the composition of amino acids in the urine of a healthy person. alanine, serine, glutamic acid and glycine

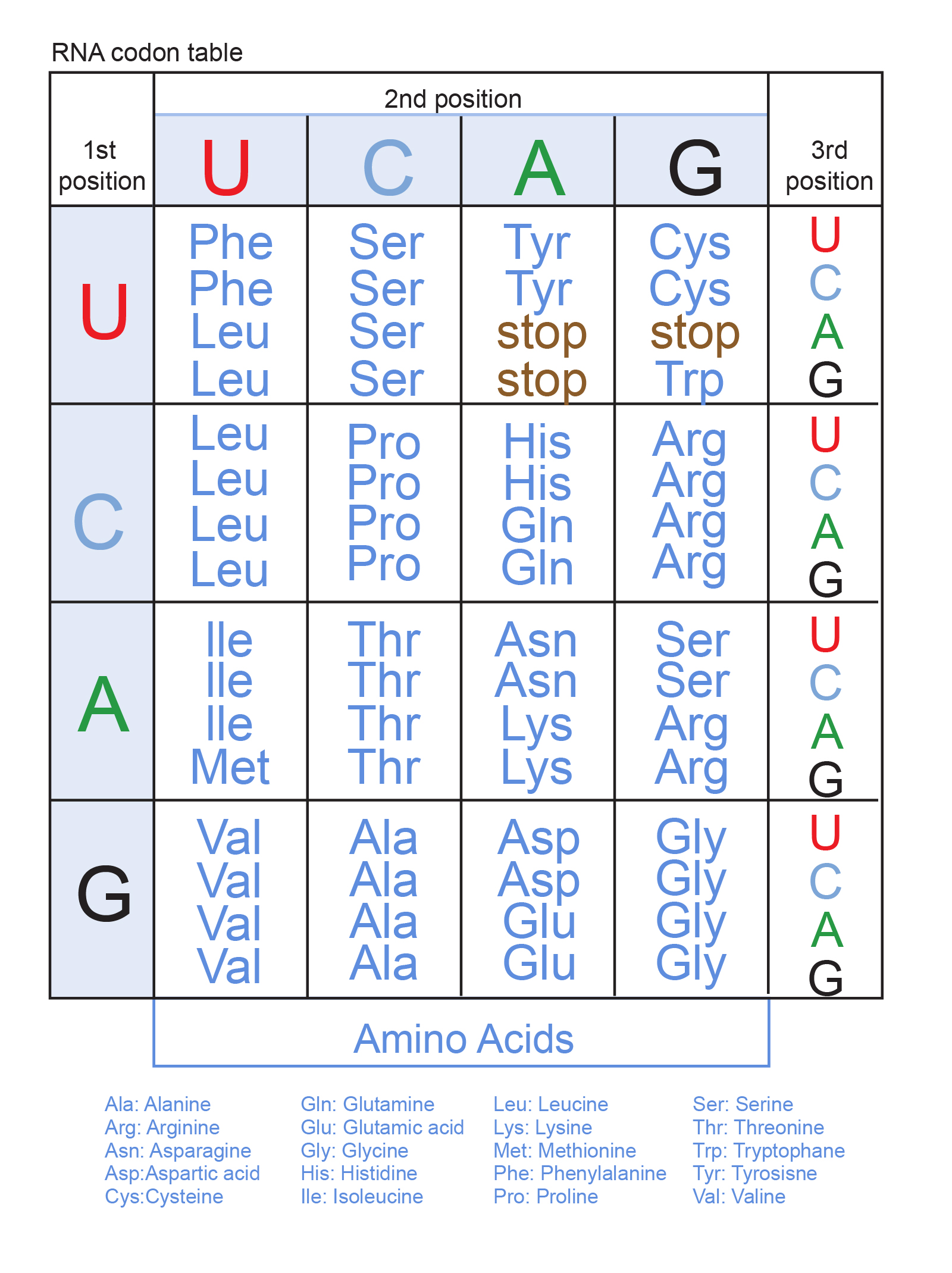
Conclusion. A person is sick with cystinuria if amino acids are found in his urine. There are cesteine, glutamine, arginine, lysine

Determine the possible number of information triplets in the region of the DNA molecule, consisting of 360 base pairs, and in the mRNA molecule, containing 300 nucleotides.

One triplet consists of three nucleotides

360 split 3 amount 120 triplets or codons amount 120 proteins

300 split 3 amount 100 triplets or codons amount 100 proteins

  
What amino acids can be transported to the ribosome by t-RNA with anticodons: AUG, AAA, GUC, GCU, CGA, CUC, UAA, UUC?

Amino acids are encoded by codons on the mRNA. TRNA anticodons are complementary to mRNA codons. To determine amino acids, you first need to determine the mRNA codons.

5`UAC UUU CAG CGA GCU GAG AUU AAG 3`

Proteine tyr phe gln arg ala glu ile lis

The section of the B chain of insulin is represented by the following amino acids: phenylalanine - valine - glutamine - histidine - leucine - cysteine - glycine - serine - histidine. Determine the section of the DNA molecule encoding this polypeptide, as well as its length and mass, if it is known that the mass of one nucleotide is 360 daltons, and the length is 0.34 nm.

In this task, we have to do a reverse translation. That is, determine the structure of DNA by protein.

The genetic code is degenerate. And an amino acid can be encoded by multiple triplets. When solving a problem, use only one of them.

MRNA 5` UUU GUU CAA CAU UUA UGU GGU AGU UCU CAC3`

DNA 1 strend 3`AAA CAA GTT GTA AAT ACA CCA TCA AGA GTG 5`

DNA 2 strend 5` TTT GTT CAA TTA TGT GGT AGT TCT CAC 3`

Two DNA strands contain 60 nucleotides

60 multiply 360 equally 21600 daltons

Two DNA strands are parallel. The length of the DNA molecule is 30 nucleotides

30 multiply 0.34 equally 10.2 nm