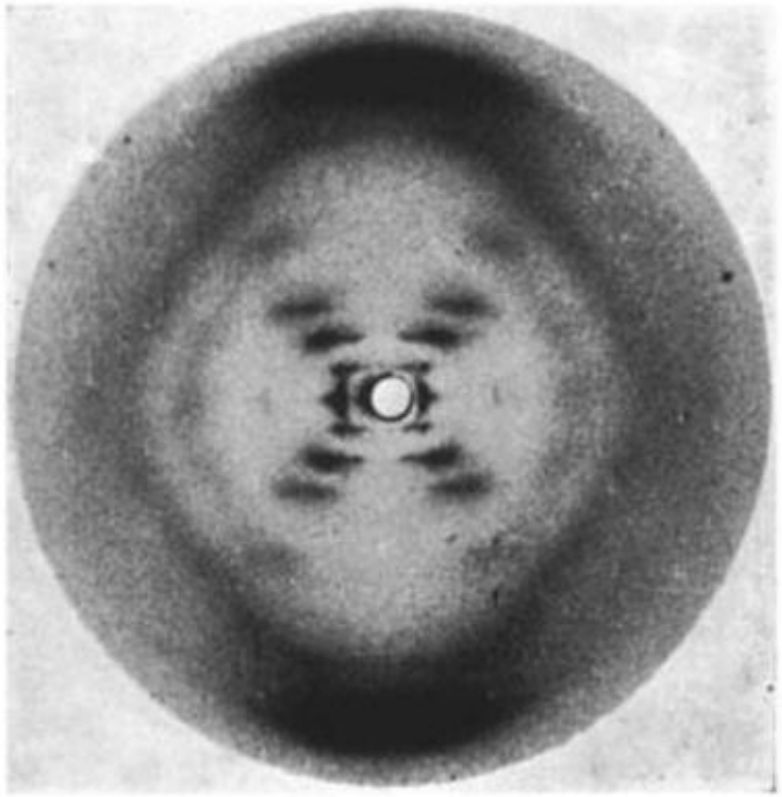
GENETICS
LECTURE 7

Katarzyna Osmańska-Załuska,PhD

Discovering the structure of DNA

- 28 February 1953 - discovering the structure of DNA: James Watson and Francis Crick announced in The Eagle Pub that they had solved the secret of life (as immortalised by a plaque in that pub)
- James Watson, Francis Crick, Maurice Wilkins, Rosalind Franklin

Discovering the structure of DNA

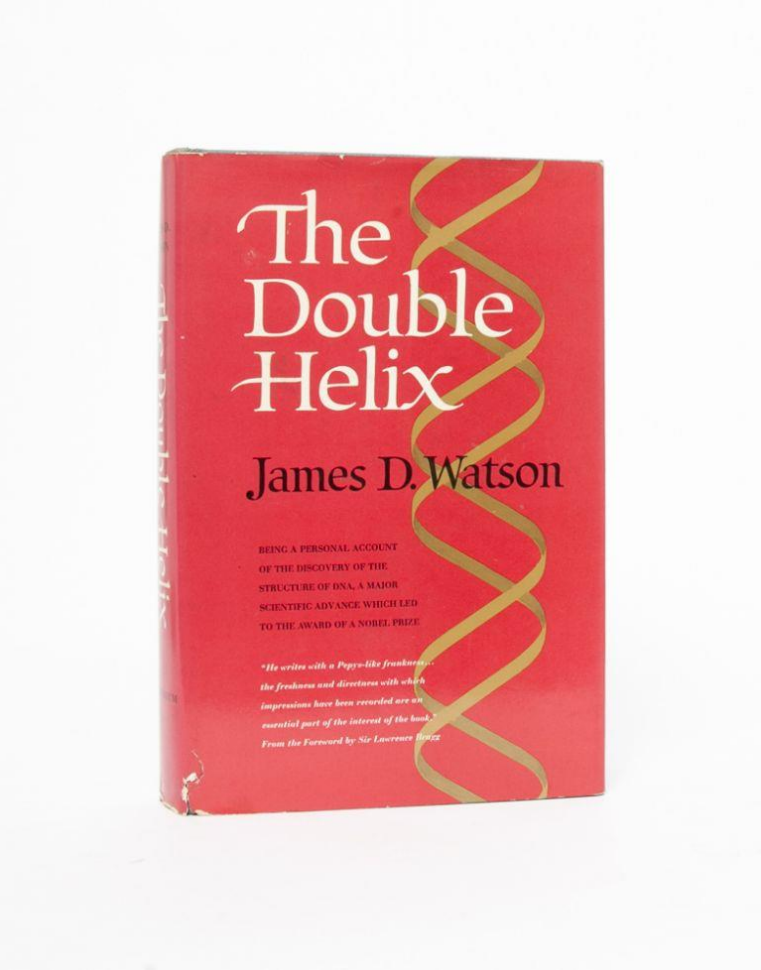


Photograph 51' showing a DNA sodium crystal made by Rosalind Franklin

Discovering the structure of DNA

- 1962 r. - award of the Nobel Prize: Francis Crick, James Watson and Maurice Wilkins
- Nobel Prize - Rosalind Franklin (died 4 years earlier) and Erwin Chargaff were omitted
- Chargaff's rule: guanine content equals cytosine content, while thymine content equals adenine content

Discovering the structure of DNA



"The Double Helix”

An autobiographical book by American geneticist James Watson describing the story of the discovery of the structure of DNA

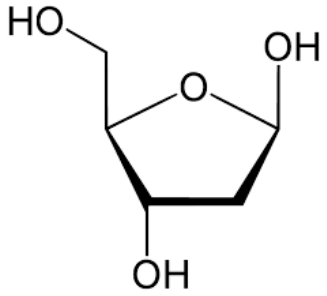
DNA - DEOXYRIBONUCLEIC ACID

- Polymer made up of chains of nucleotide monomers
- Each nucleotide contains a sugar, a nitrogenous base and a phosphate group
- Sugar: 2'-deoxyribose containing five carbon atoms labelled 1', 2' etc.
- Four nitrogenous bases: adenine and guanine (two carbon-nitrogen rings, purines) and thymine and cytosine (one ring, pyrimidines)

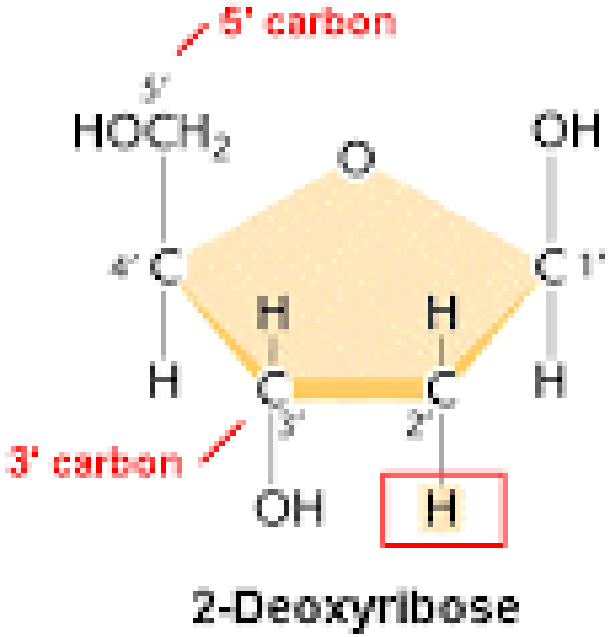
DNA - DEOXYRIBONUCLEIC ACID

- Nitrogenous bases linked to deoxyribose via 1' carbon
- Sugar + base = nucleoside
- Nucleotide - one, two or three phosphate groups linked to the 5' carbon of a sugar
- Nucleotides - individual molecules or in polymerised form in DNA and RNA structures

DNA - DEOXYRIBONUCLEIC ACID

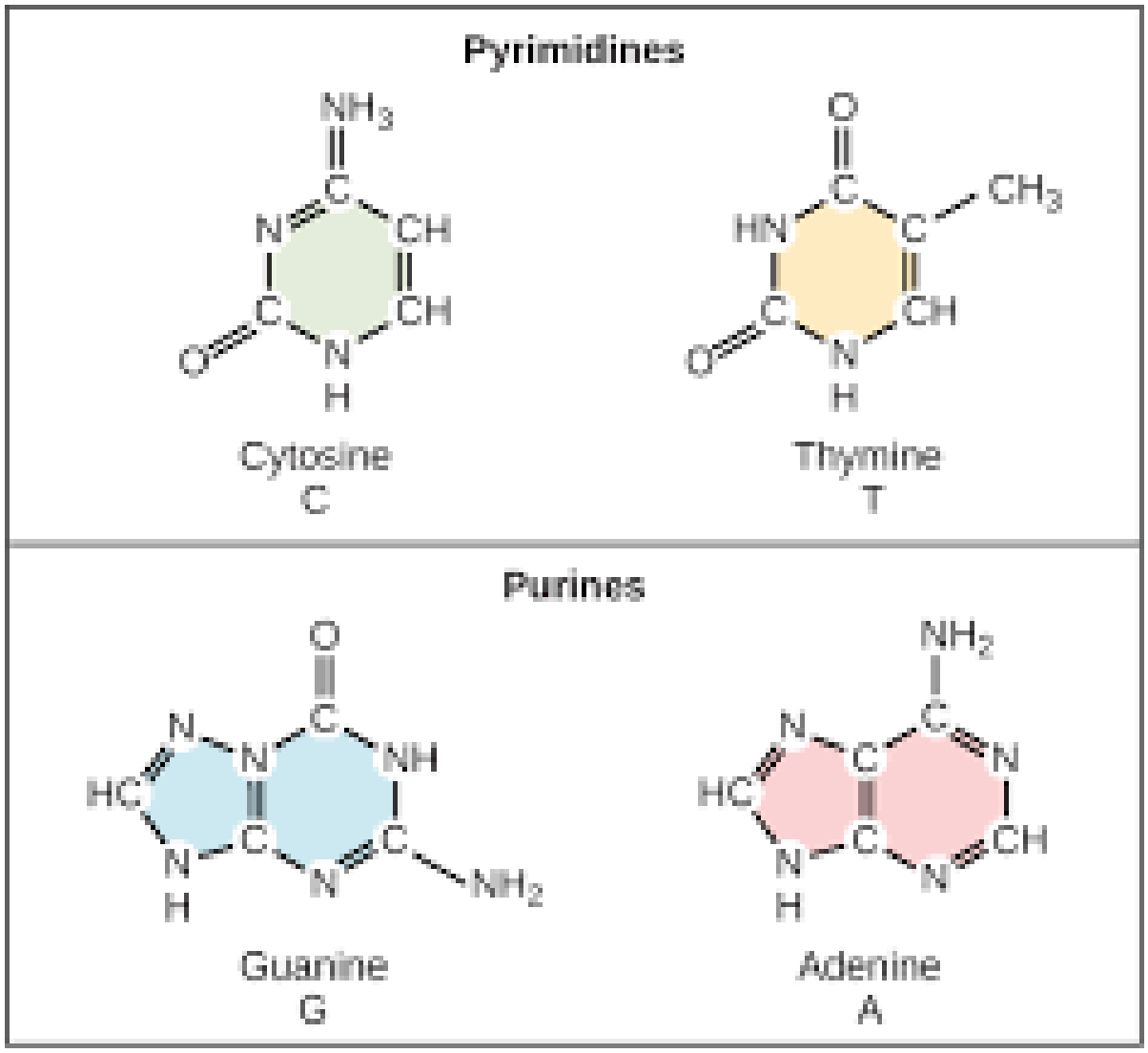


SUGAR - 2'-DEOXYRIBOSE

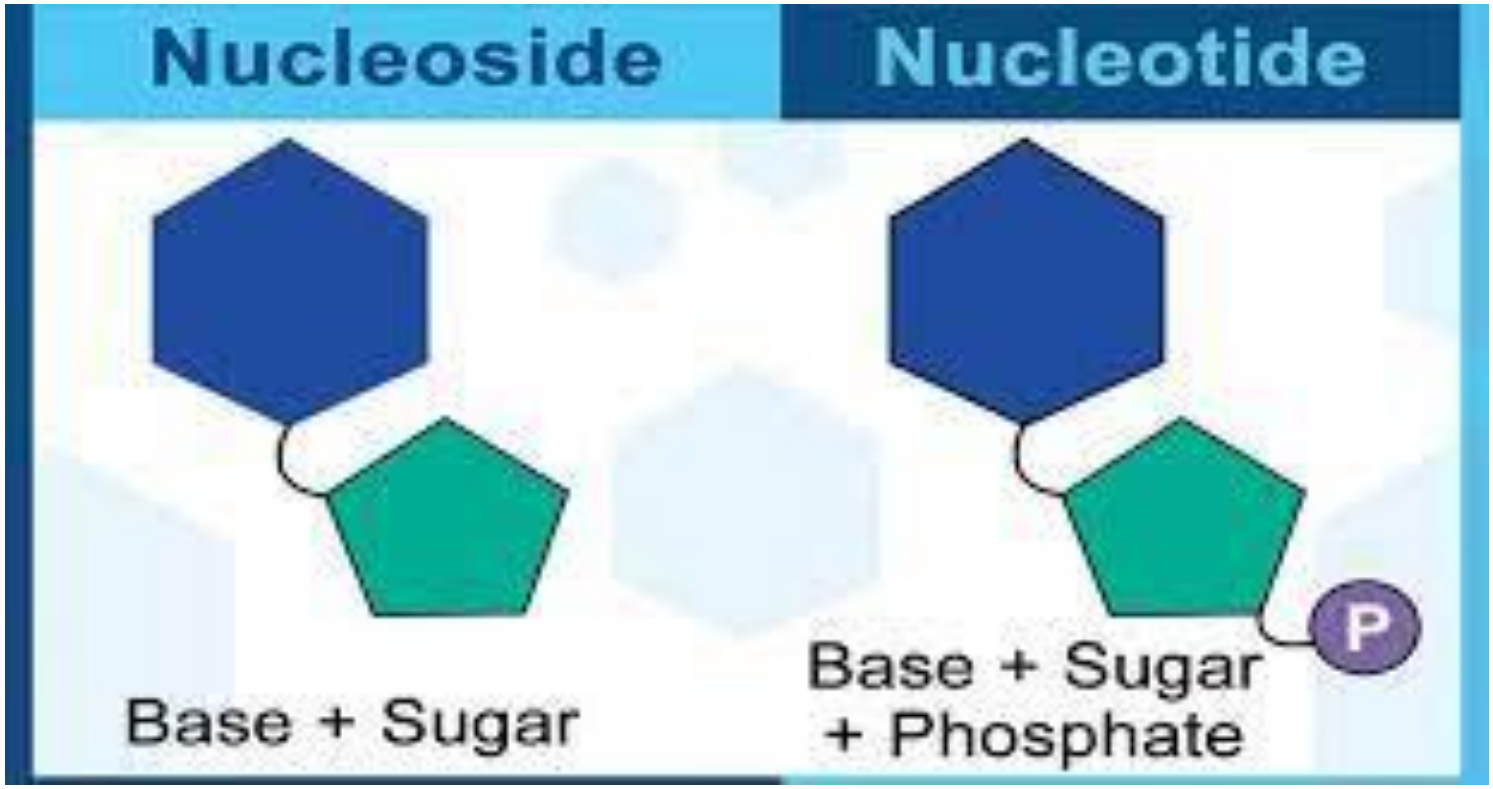


2-Deoxyribose

DNA - DEOXYRIBONUCLEIC ACID
NITROGENOUS BASES



DNA - DEOXYRIBONUCLEIC ACID
NUCLEOSIDE AND NUCLEOTIDE



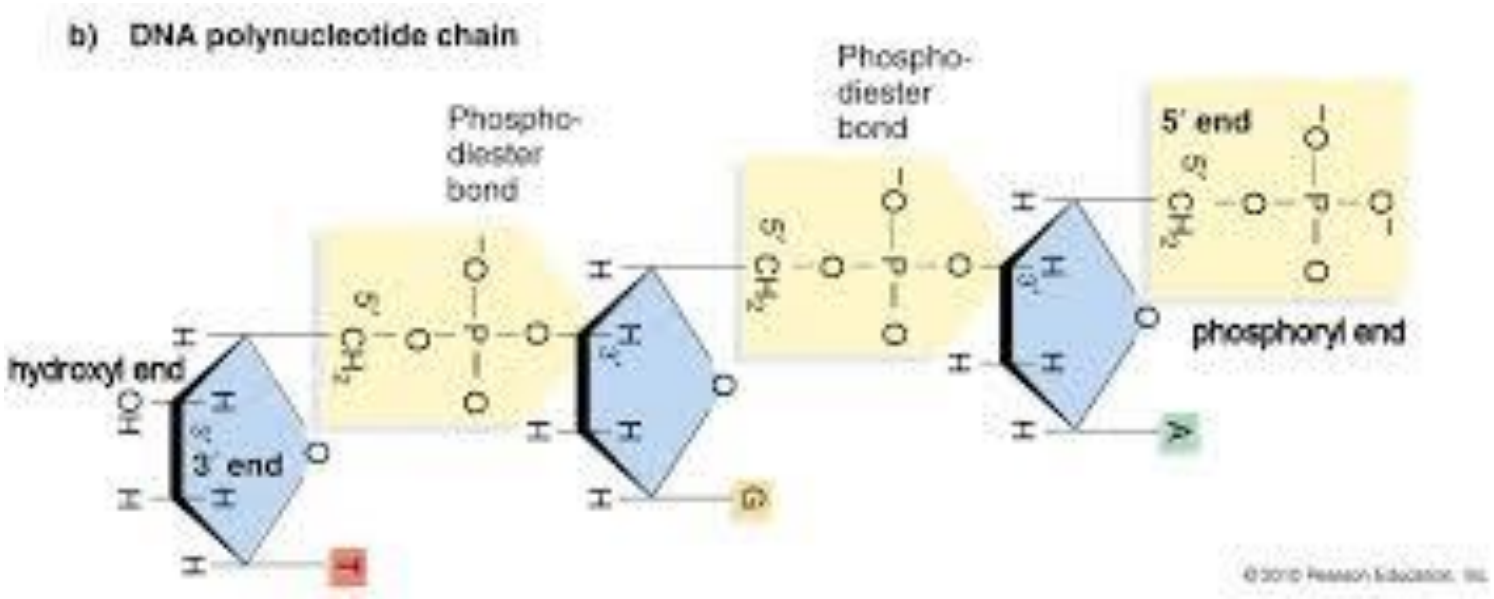
DNA POLYNUCLEOTIDES

- Linked triphosphate nucleotides with four bases form DNA polynucleotide chains
- Dropout of two phosphate groups during polymerisation
- Nucleotides are joined by the remaining groups
- Phosphodiester bond between the 5' phosphate group of one nucleotide and the 3' hydroxyl group of the next nucleotide

DNA POLYNUCLEOTIDES

- A polynucleotide has a free phosphate group 5' at one end (called the 5' end) and a free -OH group at the other end (called the 3' end)
- The sequence (order) of bases encodes the genetic information read in the 5' to 3' direction
- Polynucleotides are very long

DNA POLYNUCLEOTIDES



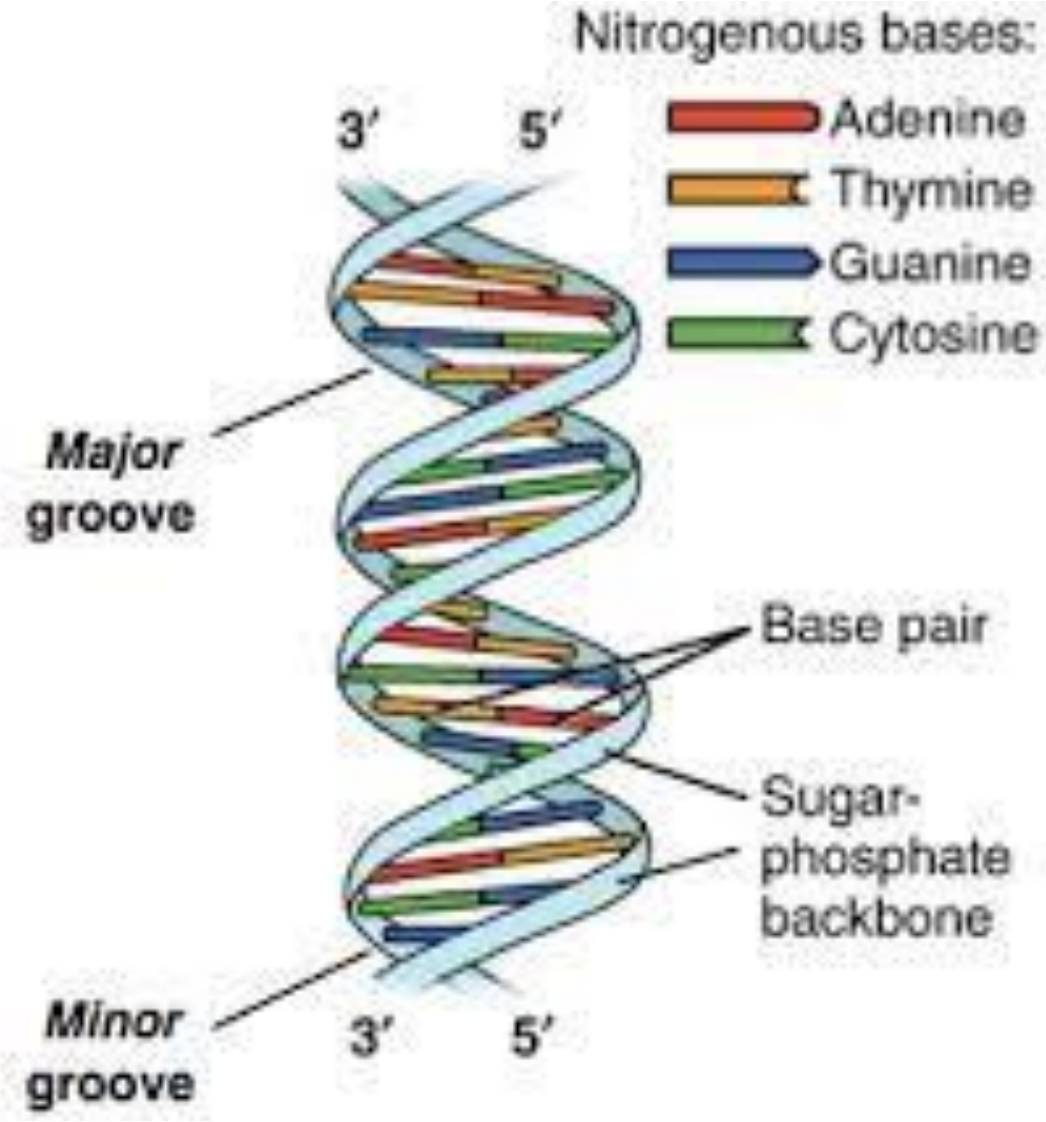
Double helix

- DNA molecules consist of two polynucleotide strands intertwined together in the form of a double helix (helix)
- The backbone of the molecule - the sugar connections to the phosphate groups
- Bases facing inwards, stacked one above the other
- Chains are counter-rotating - i.e. have opposite orientations

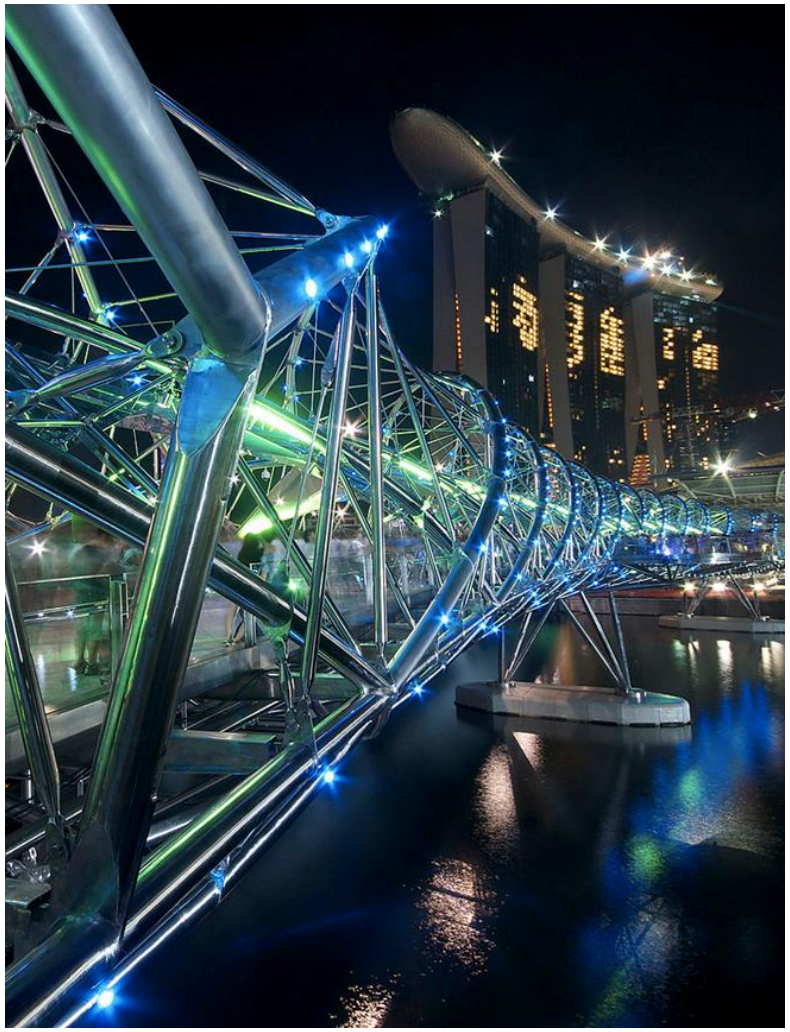
Double helix

- Double helix - right-handed with a full 360 degree twist every 10 bases - form B
- The helix contains a major groove (interacts with proteins) and a minor groove
- Other forms: form A (slightly more compact structure), forms C, D, E and Z (left-handed)

Double helix



Singapore walking bridge inspired by DNA structure

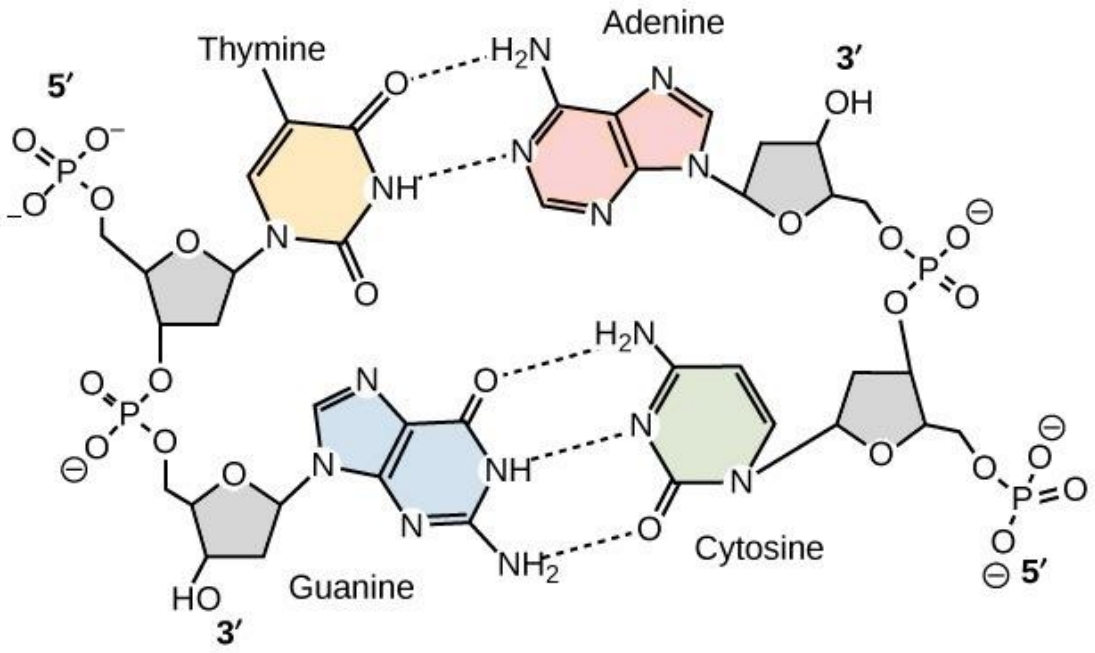


Complementary base pairing

- Hydrogen bonds between bases of two strands - helix stabilisation
- Pairing of purine with pyrimidine - always!
- Adenine with thymine, guanine with cytosine - complementary base pairing
- Limited base pairing - the sequences of the bases belonging to the two strands are interdependent and the sequence of one determines and predicts the sequence of the other

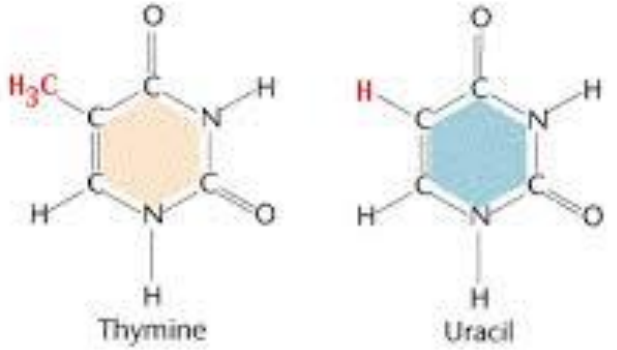
Complementary base pairing

- Preservation of genetic information during DNA replication and gene expression
- Disruption of hydrogen bonds: high temperature, chemical agents, enzyme action



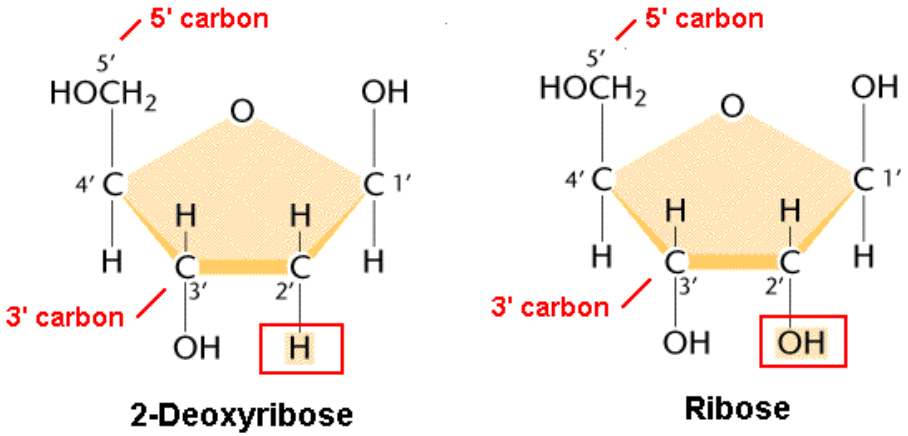
Structure of RNA (ribonucleic acid)

- Instead of thymine -> uracil



- Instead of 2'-deoxyribose -> ribose

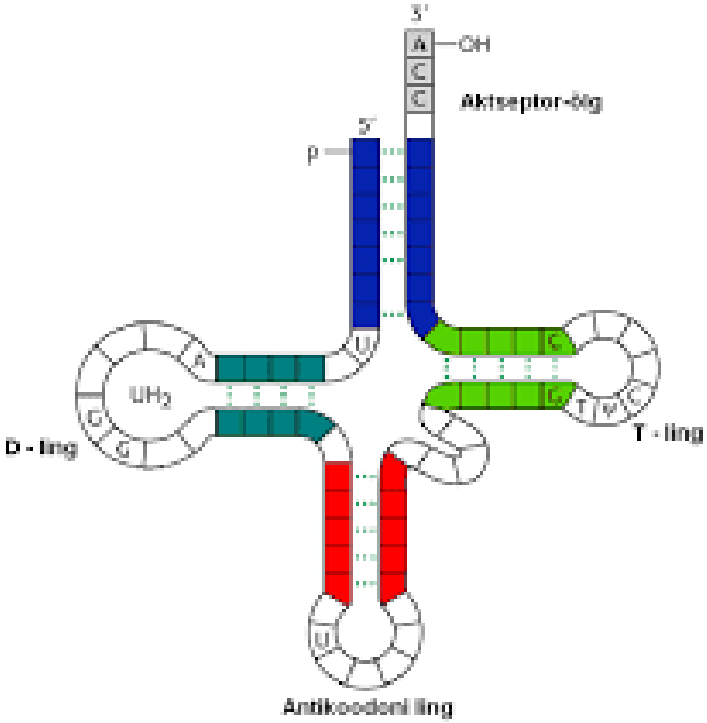
(Klug & Cummings 1997)

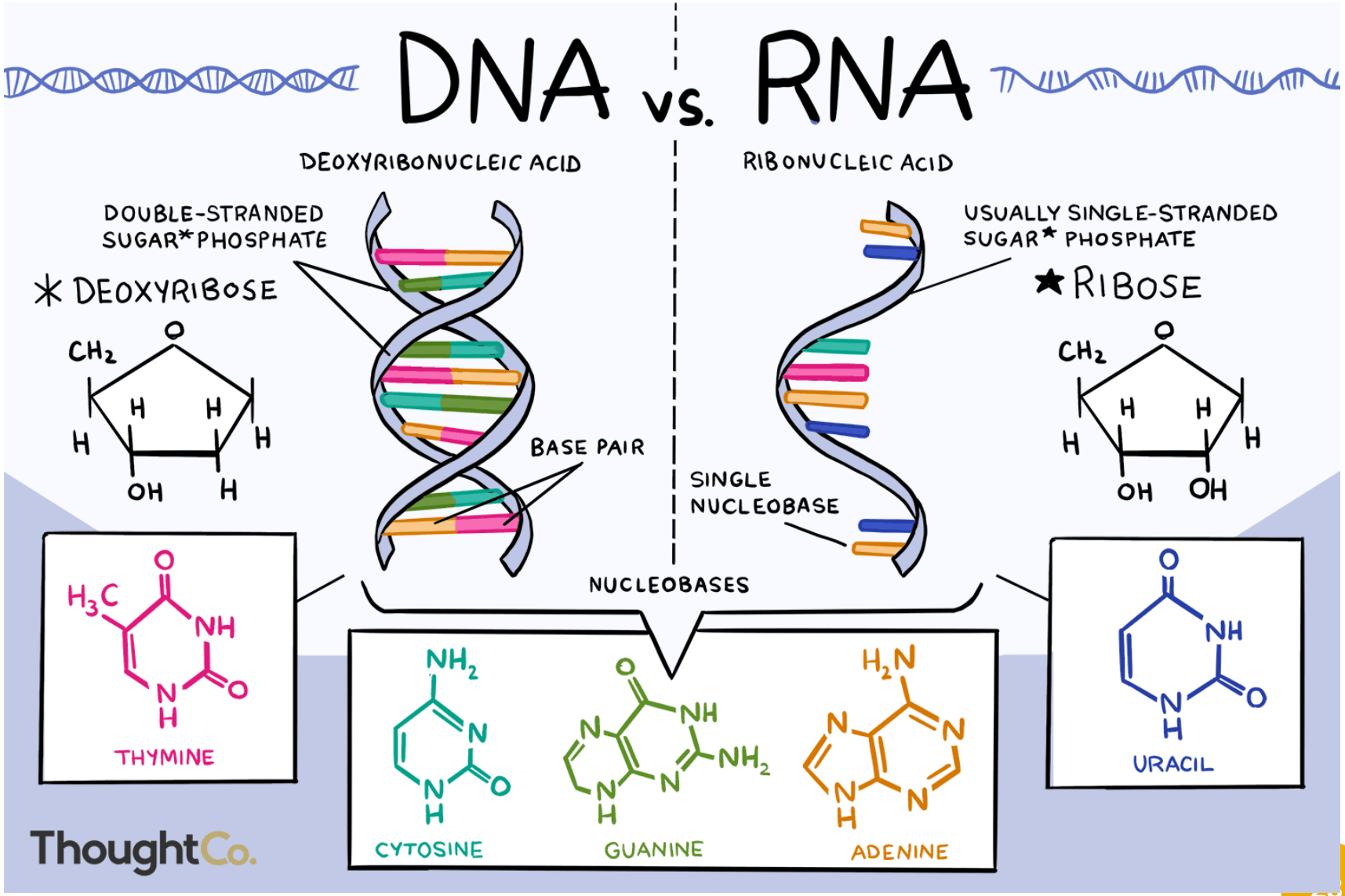


(Klug & Cummings 1997)

Structure of RNA (ribonucleic acid)

- RNA - single polynucleotide strand
- BUT short stretches of bases may occur between complementary sequences




DNA vs. RNA
DEOXYRIBONUCLEIC ACID
RIBONUCLEIC ACID
DOUBLE-STRANDED
SUGAR*PHOSPHATE
* DEOXYRIBOSE
BASE PAIR
USUALLY SINGLE-STRANDED
SUGAR* PHOSPHATE
★ RIBOSE
SINGLE
NUCLEOBASE
NUCLEOBASES
THYMINE
CYTOSINE
GUANINE
ADENINE
URACIL
ThoughtCo.

Genes

- Several definitions
- Broadest - gene as a segment of DNA encoding a functional product (protein or RNA)
- Unit of information; refers to a specific segment of DNA encoding a sequence of amino acids or polypeptides
- Human - 21 000 genes in 23 chromosomes

Genes

- Scattered and separated by non-coding intergenic DNA
- Information encoded on the matrix strand directing the synthesis of RNA molecules
- Each of the two DNA strands can act as a template strand

Gene expression

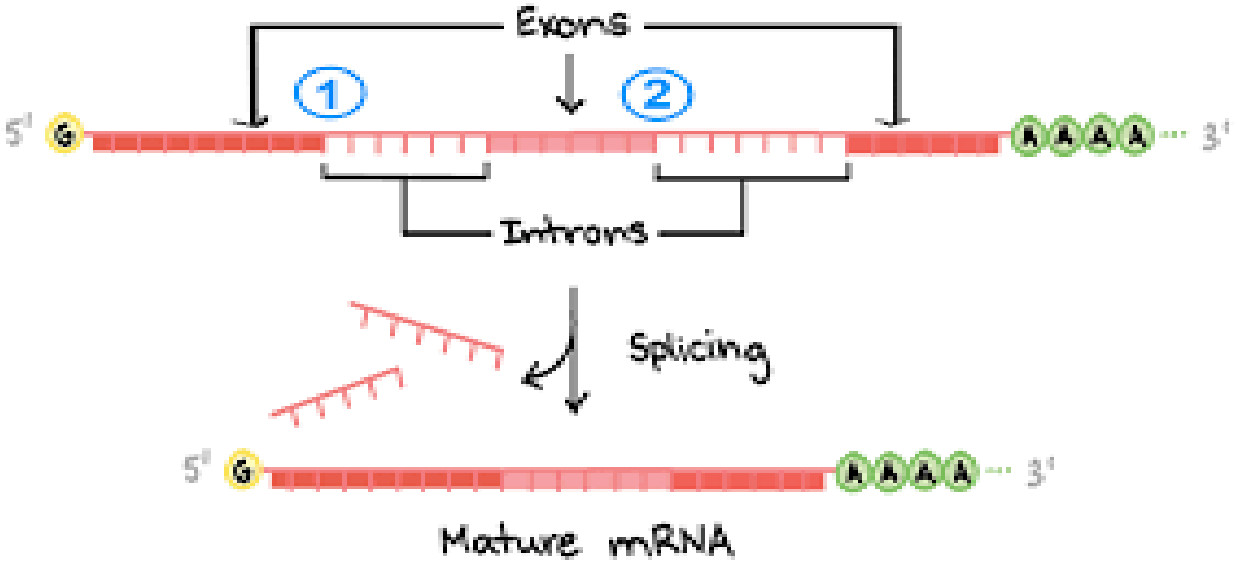
- The biological information encoded in the genes is made available in the gene expression mode
- This process synthesises RNA copies of genes, which then direct protein synthesis
- Transfer of information from DNA, before RNA, to proteins
- Coordination of the activity of many proteins influences cell function
- Gene expression ensures that genes are synthesised at the right place and time

Gene expression - promoters

- Gene expression is tightly regulated
- Not all genes present in a cell are active
- Different cell types benefit from the expression of different genes
- Promoter - DNA segment that regulates gene expression; located upstream of the coding sequence
- Promoter binds RNA polymerase and protein-associated transcription factors, initiating the synthesis of the RNA molecule

Introns and Exons

- Exons - coding sequence of a gene divided into a sequence of segments
- Introns - non-coding sequences; usually form the bulk of the gene sequence; number and size varies from gene to gene
- Introns are removed from RNA transcripts by splicing; this precedes protein synthesis



Pseudogenes

- Copies of genes containing sequence changes acquired during the evolutionary process and preventing the production of proteins

- These are evolutionary relics of the original genes

Gene expression

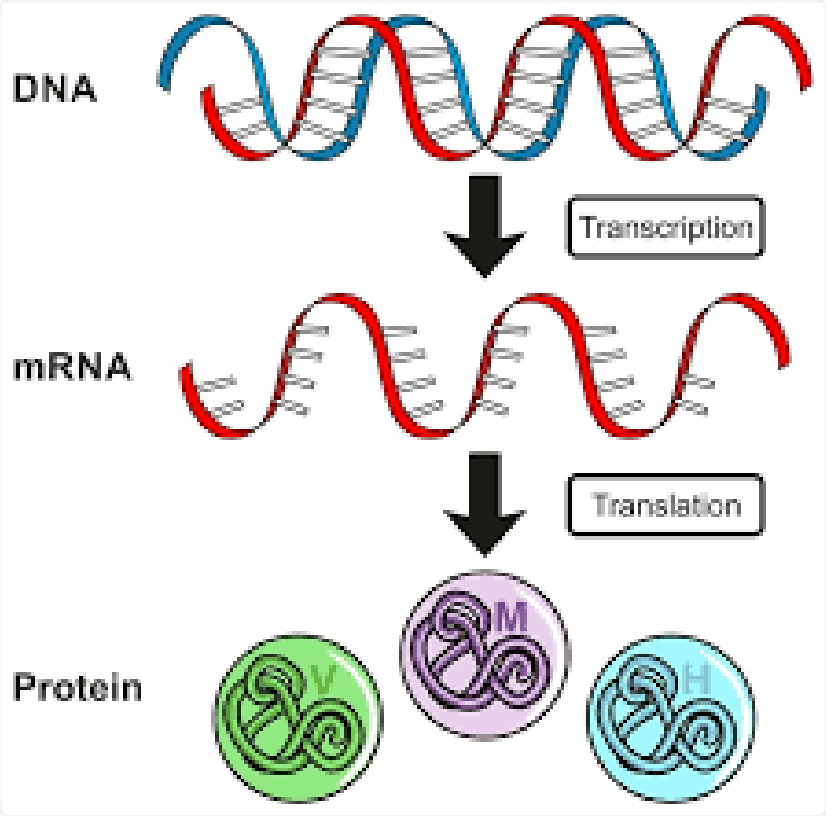
- Information necessary for the reproduction of an organism in DNA, encoded in sequences of bases and organised into strings of genes
- Gene expression - the process by which cells decode and use information to synthesise proteins that perform multiple functions in cells
- Expression process - copying information from DNA to RNA; sequence of bases complementary to the DNA template

Gene expression

- RNA surveillance of protein synthesis - the amino acid sequence determined by the order of bases in the RNA

- Colinearity of DNA and amino acid sequence of a polypeptide - the order of bases in the 5'-3' direction of the coding DNA strand determines the structural order of the encoded polypeptide (from the amino group to the hydroxyl end - OH)

Gene expression

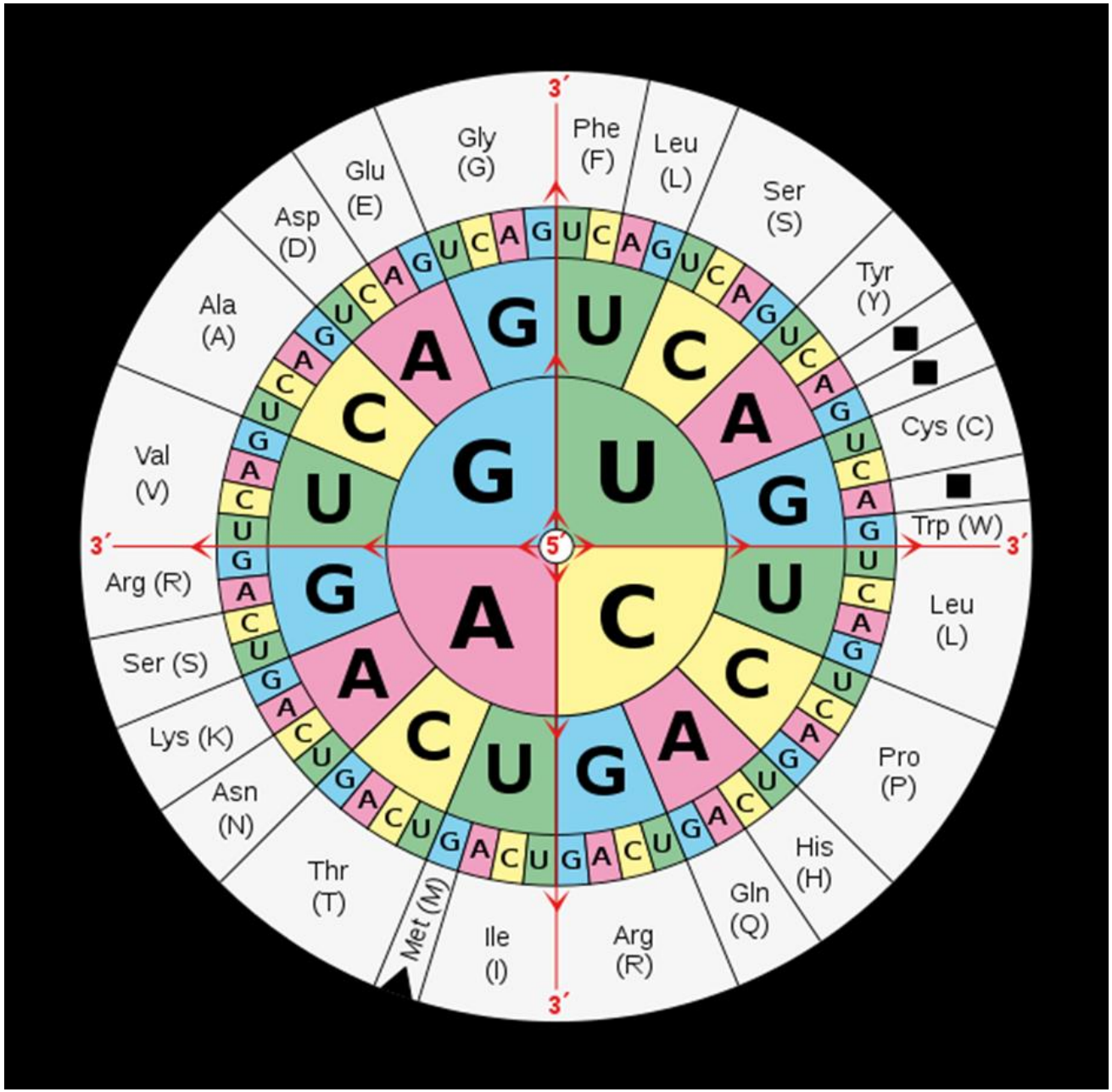


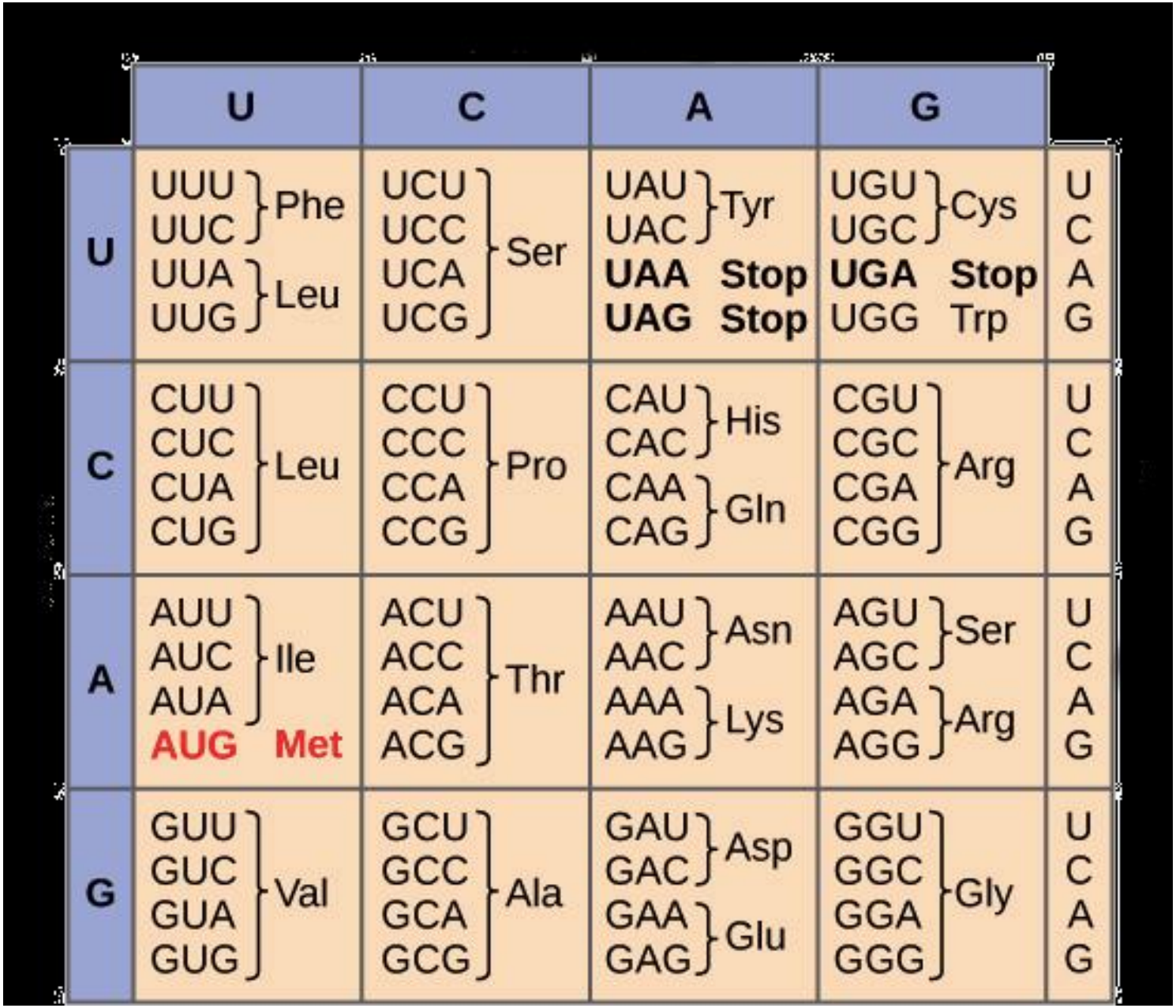
Genetic code

- Description of how to interpret the sequence of bases in DNA in terms of amino acid sequences during protein synthesis
- Gene divided into codons defining specific amino acids
- 64 different codons (combination of 4 bases in DNA and RNA: 4^3) encoding 20 amino acids
- All amino acids (exception: methionine and tryptophan) encoded by more than one codon - **degeneracy/redundancy phenomenon**

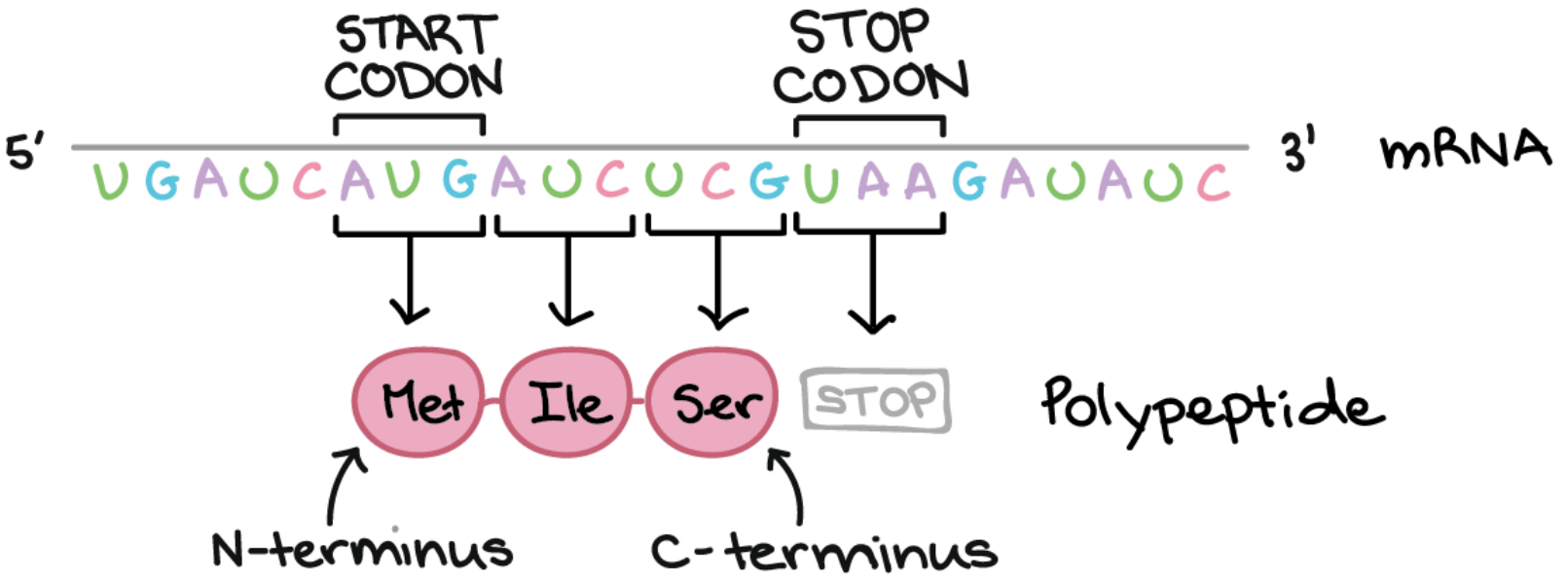
Genetic code

- Synonymous codons = codons that code for the same amino acid
- Difference most often in the third codon position - so-called wobble position
- Redundancy - minimising effect of mutation
- 61/64 codons - encoding an amino acid
- Three - **UAG**, **UGA** and **UAA** - STOP codons, or termination codons
- **AUG** - methionine codon, signalling the start of protein synthesis, initiator codon; all polypeptides start with methionine


3′
5′
3′
3′
3′
Phe (F)
Leu (L)
Ser (S)
Tyr (Y)
Cys (C)
Trp (W)
Leu (L)
Pro (P)
His (H)
Gln (Q)
Arg (R)
Ile (I)
Met (M)
Thr (T)
Asn (N)
Lys (K)
Ser (S)
Arg (R)
Val (V)
Ala (A)
Asp (D)
Glu (E)
Gly (G)

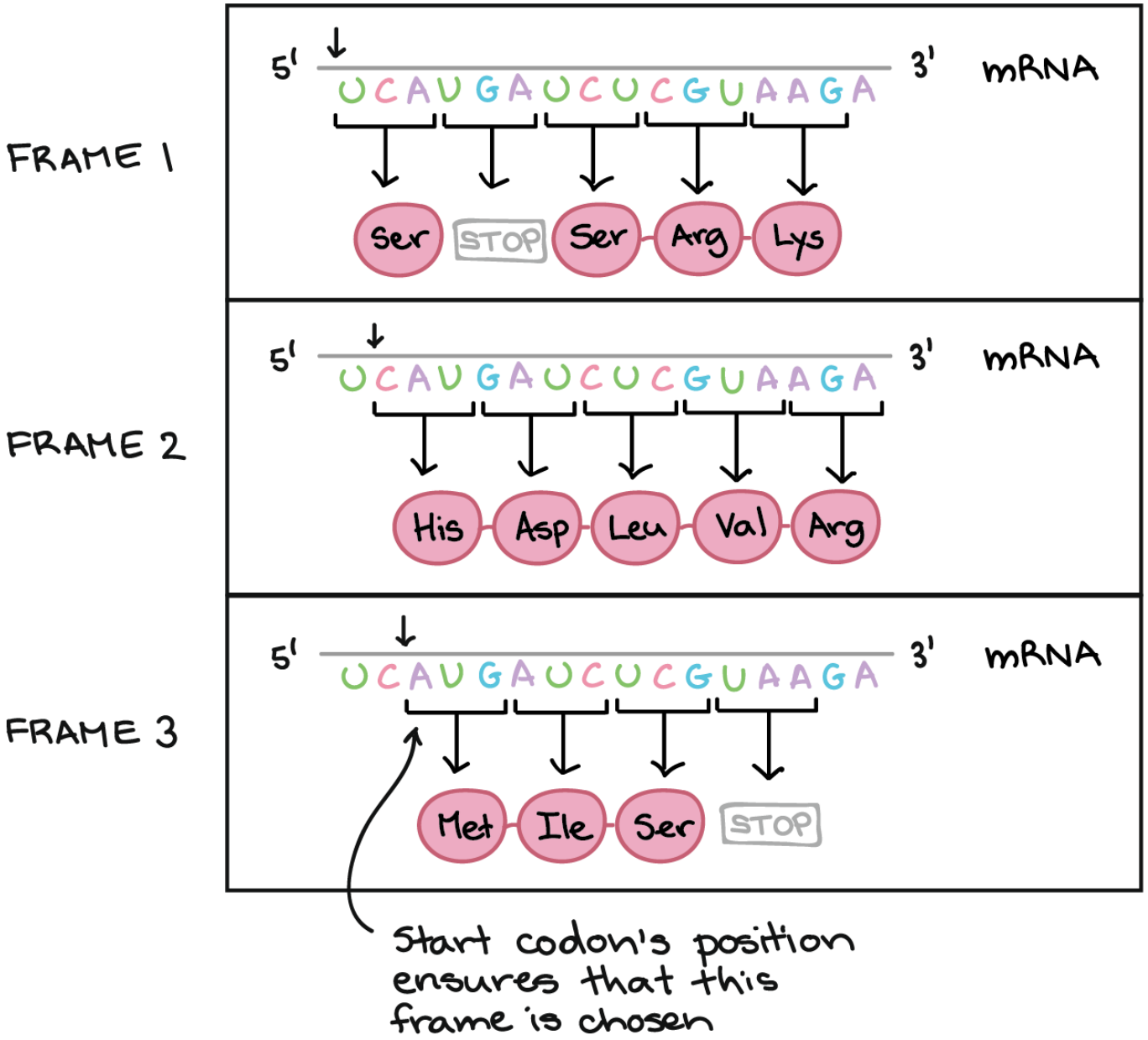


	U	C	A	G	
U	UUU, UUC } Phe UUA, UUG } Leu	UCU, UCC, UCA, UCG } Ser	UAU, UAC } Tyr **UAA Stop** **UAG Stop**	UGU, UGC } Cys **UGA Stop** UGG Trp	U C A G
C	CUU, CUC, CUA, CUG } Leu	CCU, CCC, CCA, CCG } Pro	CAU, CAC } His CAA, CAG } Gln	CGU, CGC, CGA, CGG } Arg	U C A G
A	AUU, AUC, AUA } Ile **AUG Met**	ACU, ACC, ACA, ACG } Thr	AAU, AAC } Asn AAA, AAG } Lys	AGU, AGC } Ser AGA, AGG } Arg	U C A G
G	GUU, GUC, GUA, GUG } Val	GCU, GCC, GCA, GCG } Ala	GAU, GAC } Asp GAA, GAG } Glu	GGU, GGC, GGA, GGG } Gly	U C A G


START CODON
STOP CODON
5'
3'
mRNA
U G A U C A U G A U C U C G U A A G A U A U C
Met
Ile
Ser
STOP
Polypeptide
N-terminus
C-terminus

Frameshift

- Initiating codon
- Identification of the start of protein synthesis
- Determination of reading frame for RNA-seq
- 3 different sets of codons depending on which rule is selected first
- Only one reading frame is useful information; the other two usually contain the STOP codon


FRAME 1
5'
3'
mRNA
UCAUGAUCUCGUAAGA
Ser
STOP
Ser
Arg
Lys
FRAME 2
5'
3'
mRNA
UCAUGAUCUCGUAAGA
His
Asp
Leu
Val
Arg
FRAME 3
5'
3'
mRNA
UCAUGAUCUCGUAAGA
Met
Ile
Ser
STOP
Start codon's position ensures that this frame is chosen

Open Reading frame (ORF)

- ORF - a set of codons that repeat continuously, bounded at the beginning by an initiator codon and at the end by a terminator codon

- Such sets are used to identify protein-coding DNA sequences in genome sequencing processes

Universal nature of the genetic code

- Occurs and functions in all organisms
- Uses the same codons for individual amino acids
- Exceptions: mitochondrial genomes (mitochondrion as a closed system), unicellular organisms

From DNA to proteins

3 stages of gene expression:

- transcription
- RNA processing
- translation

Transcription

- DNA as template - DNA master strand
- RNA synthesis using RNA polymerases
- RNA molecule ordered as in coding strand [sense strand (+)]
- Transcript - formed RNA molecule

Transcription

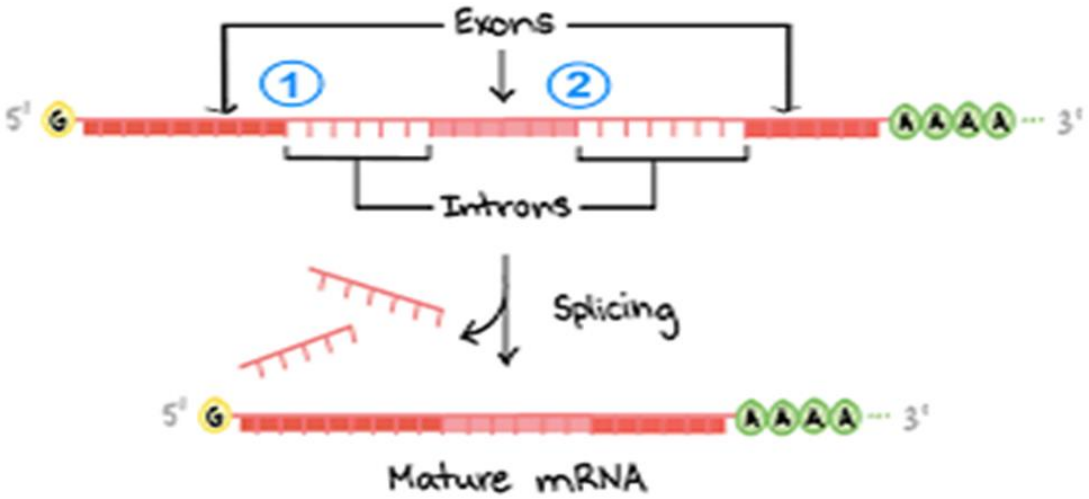
- RNA synthesis - polymerisation of ribonucleotide triphosphate subunits - ATP, UTP, GTP, CTP
- Phosphodiester bond - 3'-OH of one ribonucleotide reacts with 5'-phosphate of another
- Order of ribonucleotides determined by matrix DNA

Elongation

- Continuation of transcription
- Splicing of small length of double helix
- Splicing site - newly formed RNA bases conforming to DNA template strand, 12-17 bases
- Eukaryotes - pre-mRNA primary transcript: exons (coding) and introns (non-coding)
- Signalling of transcription termination by specific sequences

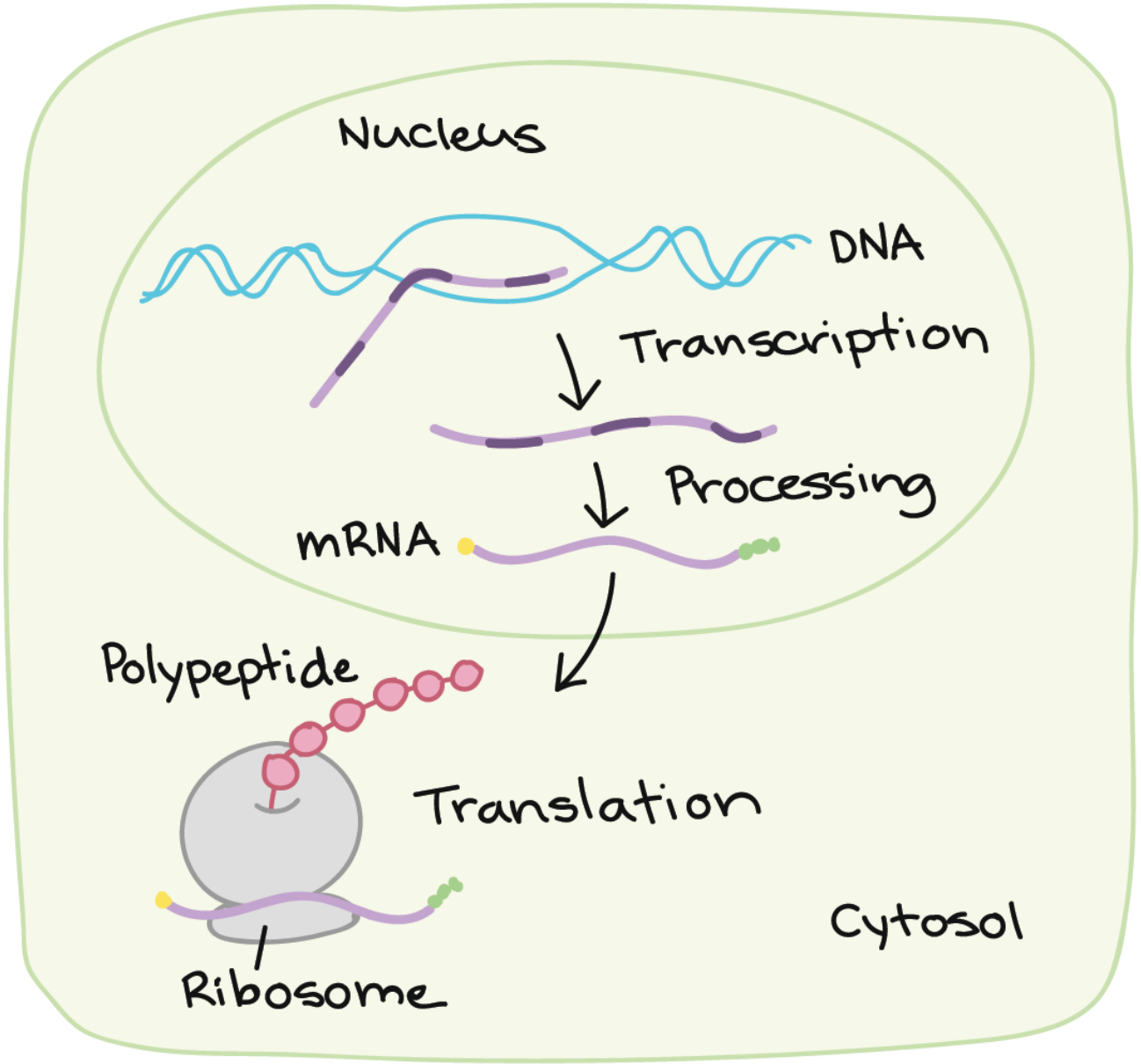
RNA PROCESSING

- Splicing
- Splicing patterns different in different tissues for the same mRNA - creation of multiple protein variants based on the same 'gene,
- Patterns can undergo pathological changes in cancerous tumours



RNA PROCESSING

- Modification of the 5' and 3' ends in eukaryotes
- End 5' - capping (cap), addition of modified 7-methylguanosine nucleotide
- 3' end - long adenine tails
- Mature mRNA molecules - translation process in ribosomes in the cytoplasm of the cell

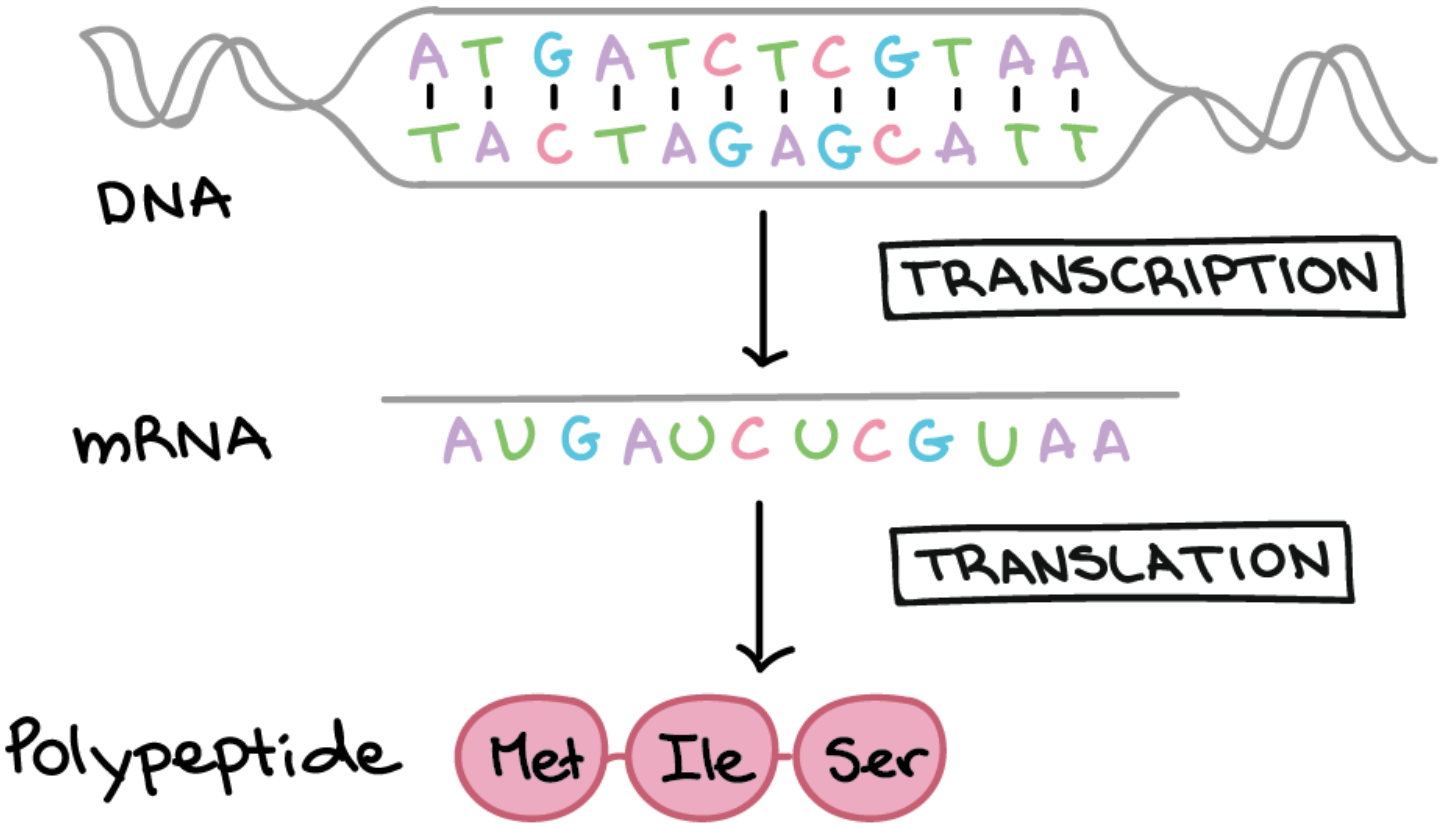

Nucleus
DNA
Transcription
Processing
mRNA
Polypeptide
Translation
Cytosol
Ribosome

Translation

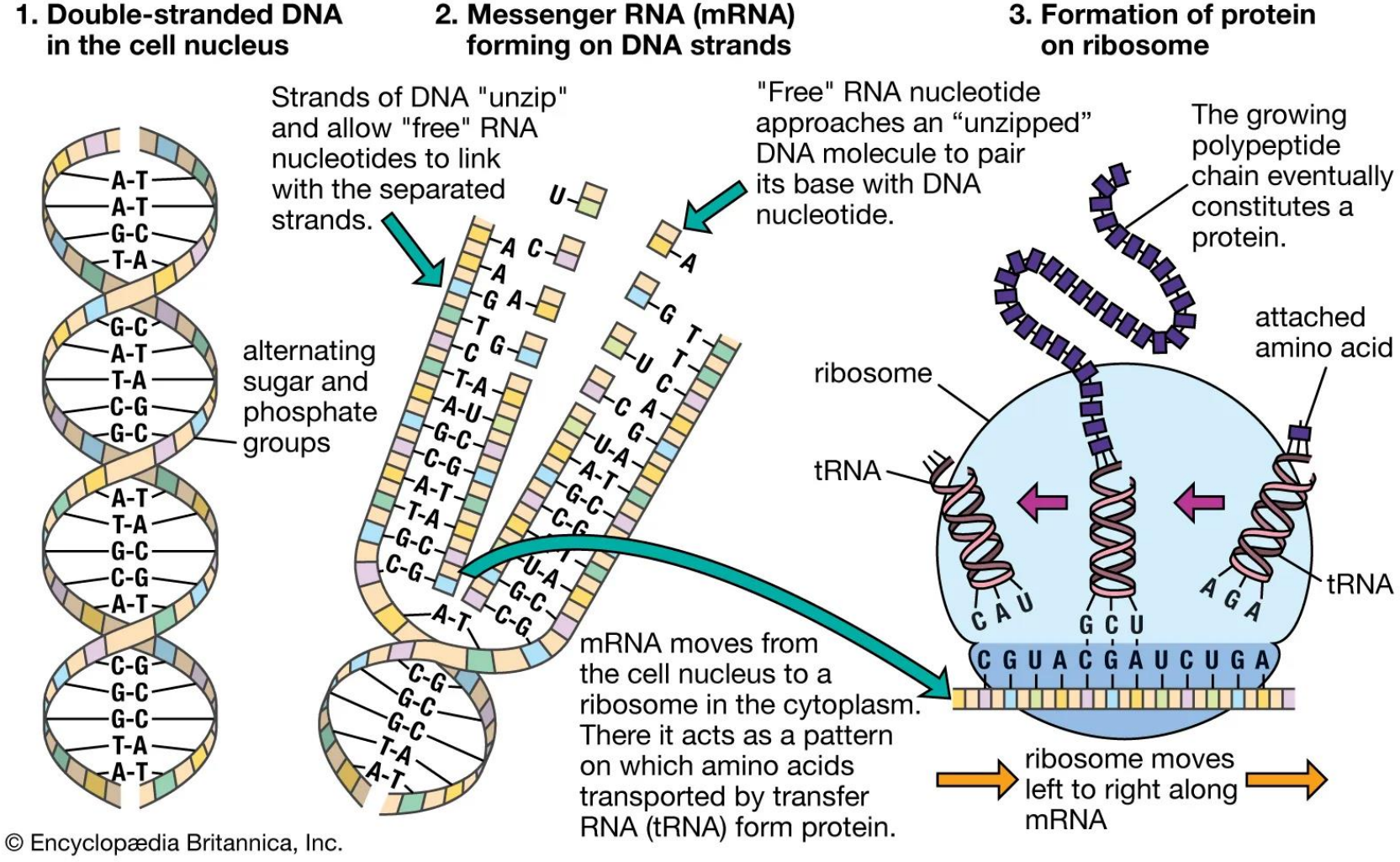
- Ribosomes
- Key role - tRNAs (transport RNAs) that deliver amino acids to ribosomes
- 31-40 types of tRNA in the cell - each binds one specific amino acid
- Binding of an amino acid to a tRNA by a covalent bond
- Recognition by the tRNA of the codon in the mRNA - placing the amino acid in the correct position
- Precise translation from mRNA to amino acid sequence

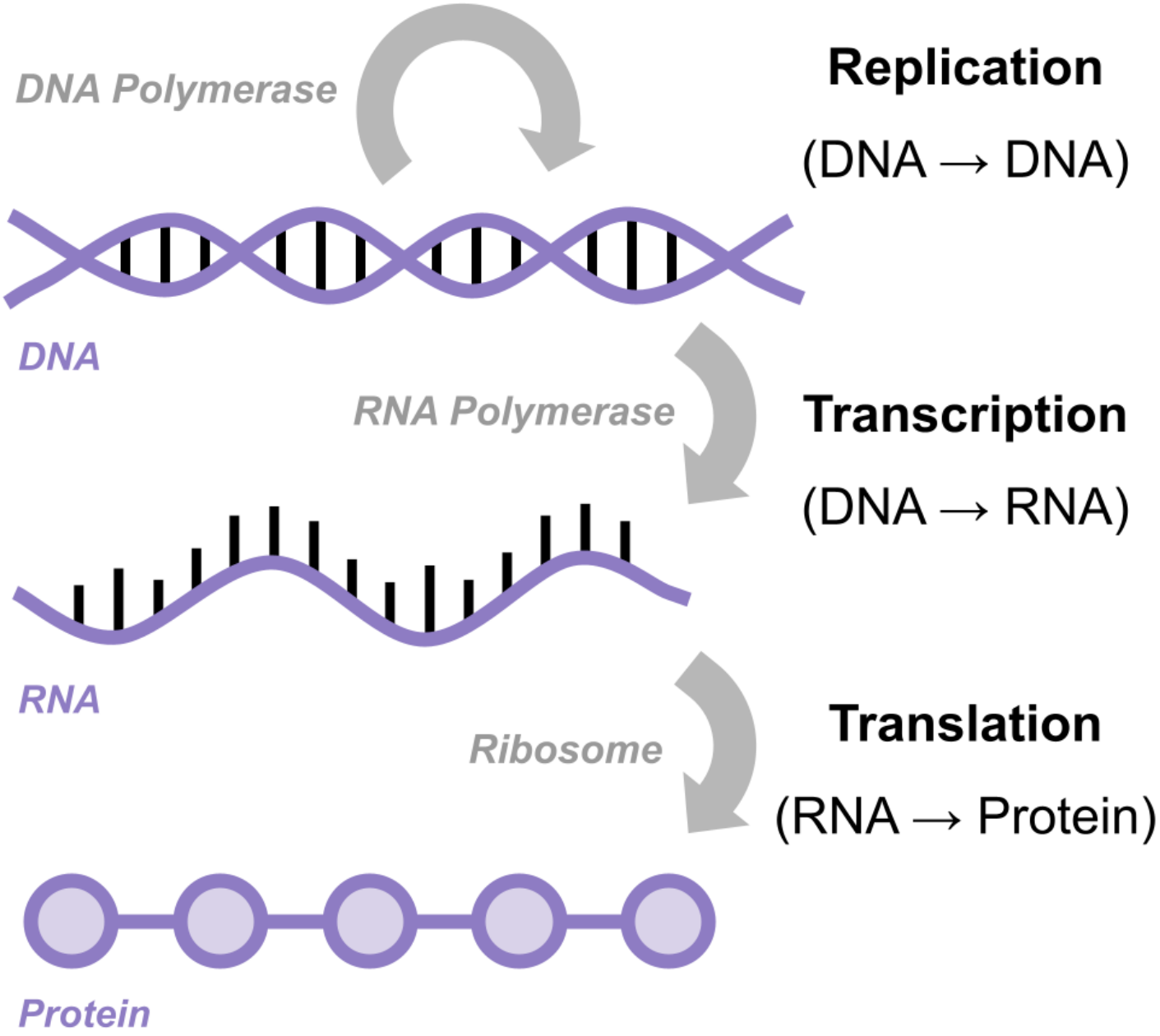
Translation

- Codon recognition via anticodon loop in tRNA (three nucleotides complementary to mRNA)
- Peptidyl transferase - peptide bond between similar two amino acids - amino acid chain elongation
- Polypeptide modification - addition of small chemical groups: methylation, hydroxylation, formylation, glycosylation


ATGATCTCGTAA
TACTAGAGCATT
DNA
TRANSCRIPTION
mRNA
AUGAUCUCGUAA
TRANSLATION
Polypeptide
Met
Ile
Ser

How DNA directs protein synthesis




DNA Polymerase
Replication
(DNA → DNA)
DNA
RNA Polymerase
Transcription
(DNA → RNA)
RNA
Ribosome
Translation
(RNA → Protein)
Protein

DNA MUTATIONS

- Maintenance of an unchanged amino acid sequence very important - effect on protein function and consequently on the whole organism
- Mutation - any change in DNA sequence resulting from chemical/physical agents or DNA replication errors.
- Fixation of mutations in the cell division process
- Location of the mutation in the gene is important
- Mutations in coding segments - effect on encoded proteins
- Mutations in non-coding segments - usually have no effect on proteins, but.....

DNA MUTATIONS

- TRANSITION - conversion of purine to purine or pyrimidine to pyrimidine (A to G, G to A, C to T, T to C)

- TRANSVERSION - conversion of purine to pyrimidine or pyrimidine to purine (A to C, A to T, G to C, G to T, C to A, C to G, T to A, T to G)

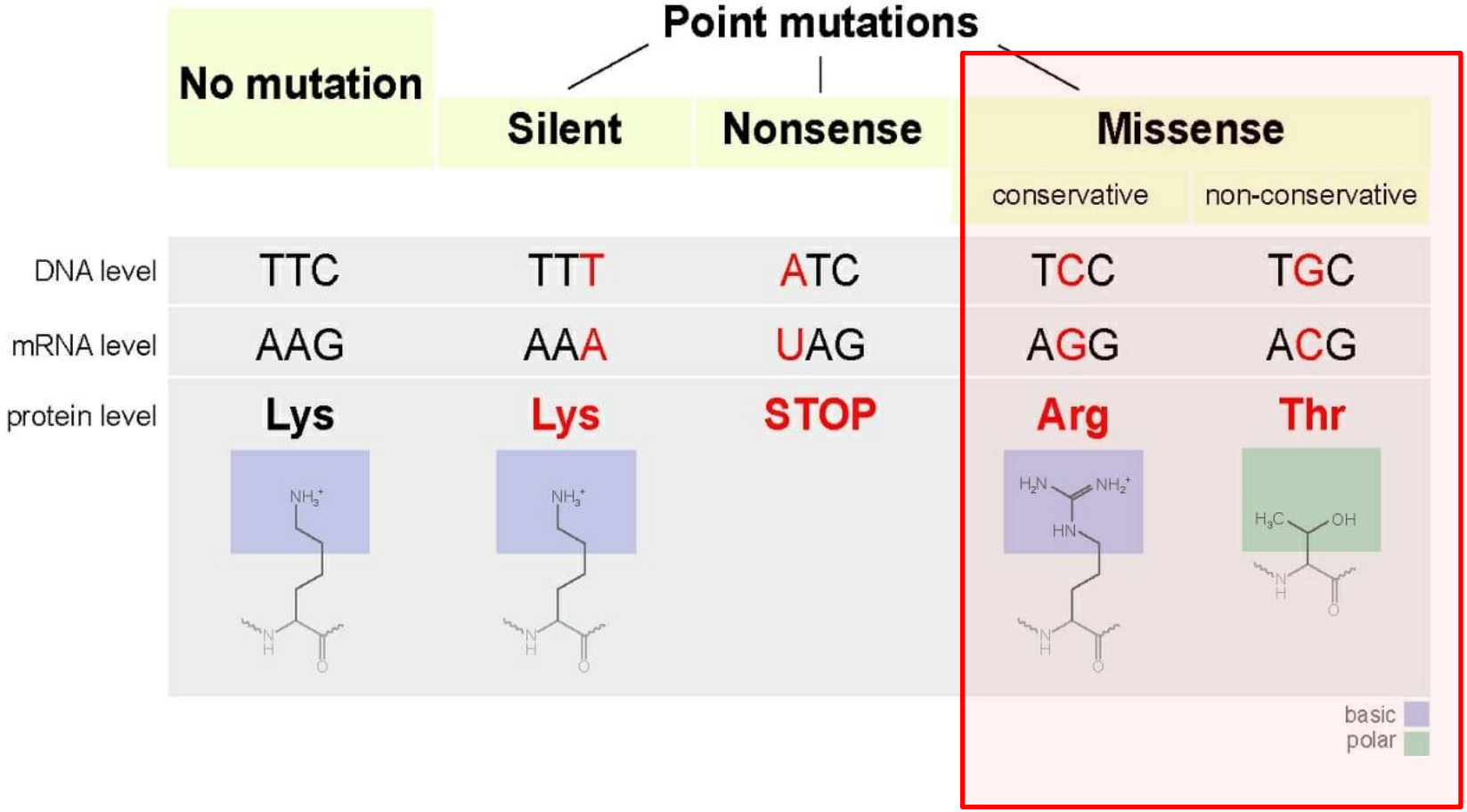
DNA MUTATIONS

1. Point mutations
   - Changing sense mutations
   - Nonsense mutations
   - Frameshift mutations
   - Silent mutations

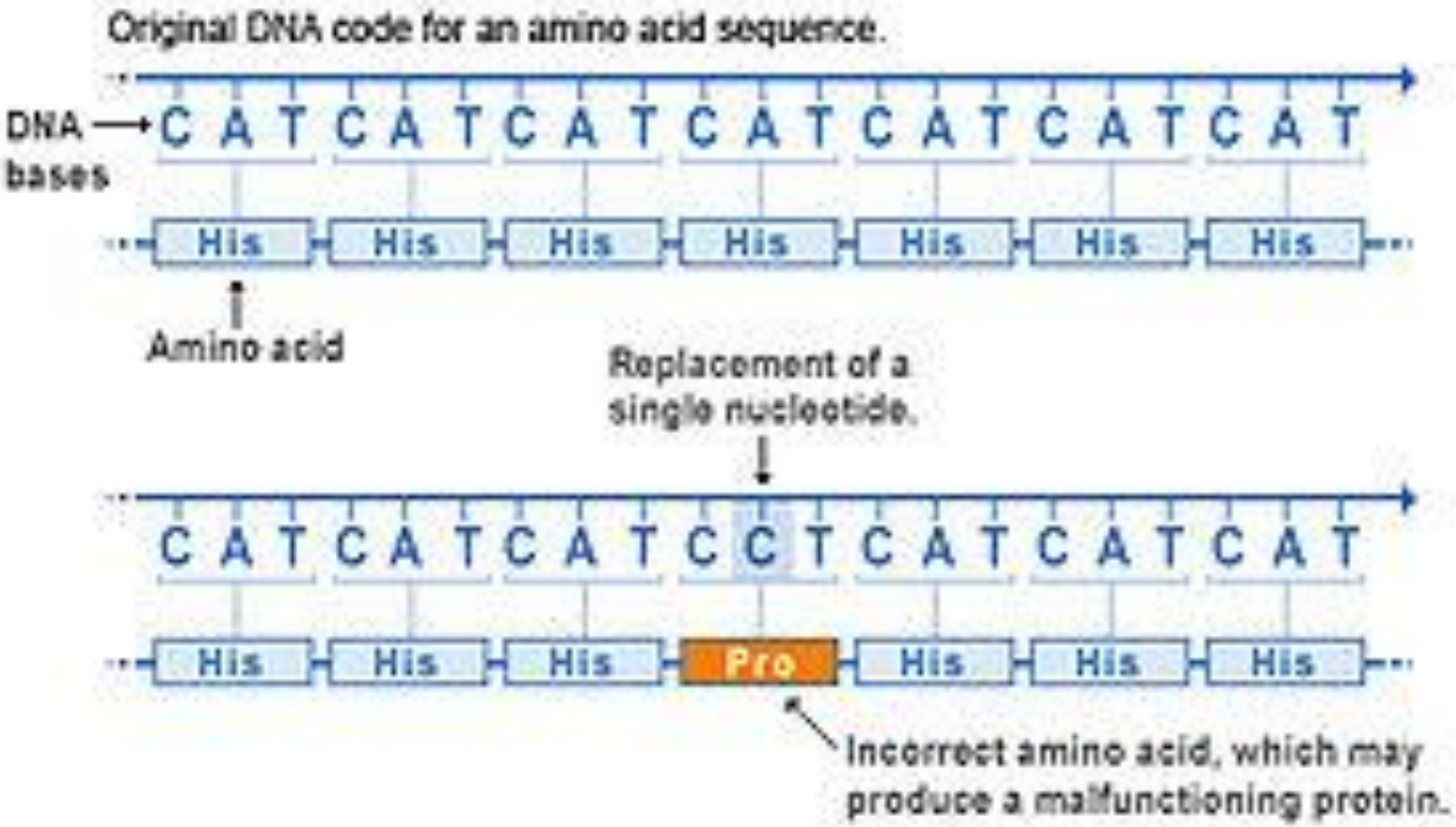
2. Broad mutations
   - Deletions
   - Insertions
   - Regroupings

Point mutations
Mutations of sense change (Missense)

- Single nucleotide changes
- Change of codon to another amino acid
- Change most often occurs in first or second codon rule
- Reduction of genetic code - change of third codon principle less likely to change amino acid
- Most proteins can tolerate minor changes in amino acid sequence
- But changes in structurally or functionally important parts of protein molecules can have a negative effect on the organism


Point mutations
No mutation
Silent
Nonsense
Missense
conservative
non-conservative
DNA level
TTC
TTT
ATC
TCC
TGC
mRNA level
AAG
AAA
UAG
AGG
ACG
protein level
Lys
Lys
STOP
Arg
Thr
basic
polar

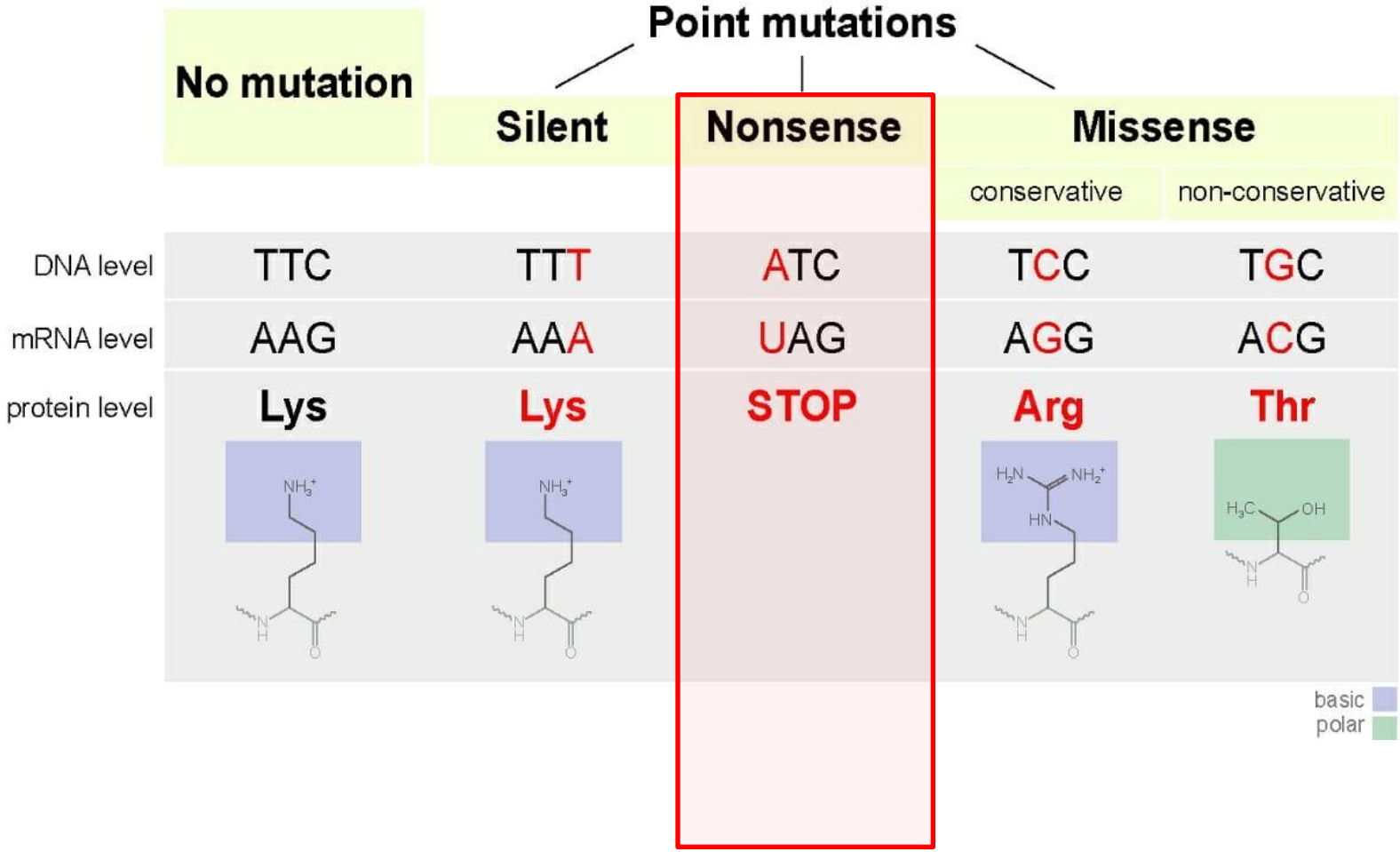
Missense mutation



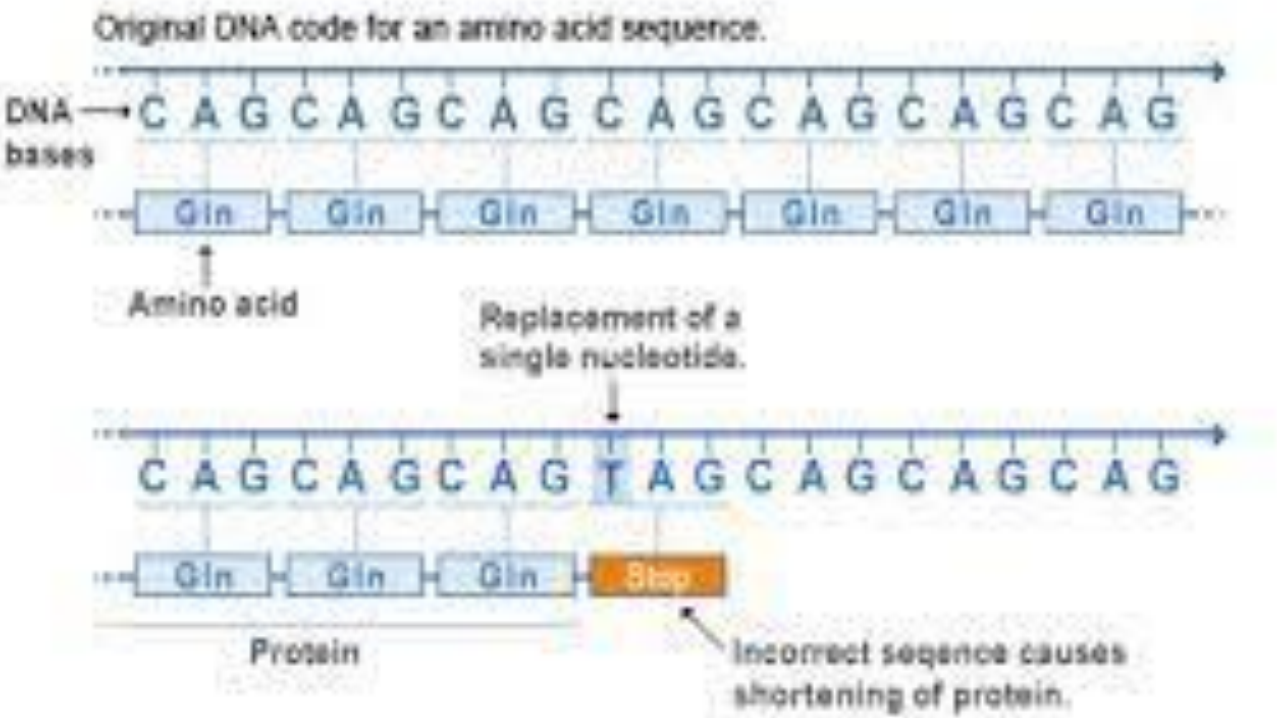
U.S. National Library of Medicine

Point mutations
Nonsense mutations

- Replacement of amino acid codon by STOP codon
- Premature termination of translation - protein of shortened length
- In eukaryotes - degradation of mRNA and complete absence of protein
- Significant effect on activity of encoded proteins


Point mutations
No mutation
Silent
Nonsense
Missense
conservative
non-conservative
DNA level
TTC
TTT
ATC
TCC
TGC
mRNA level
AAG
AAA
UAG
AGG
ACG
protein level
Lys
Lys
STOP
Arg
Thr
basic
polar

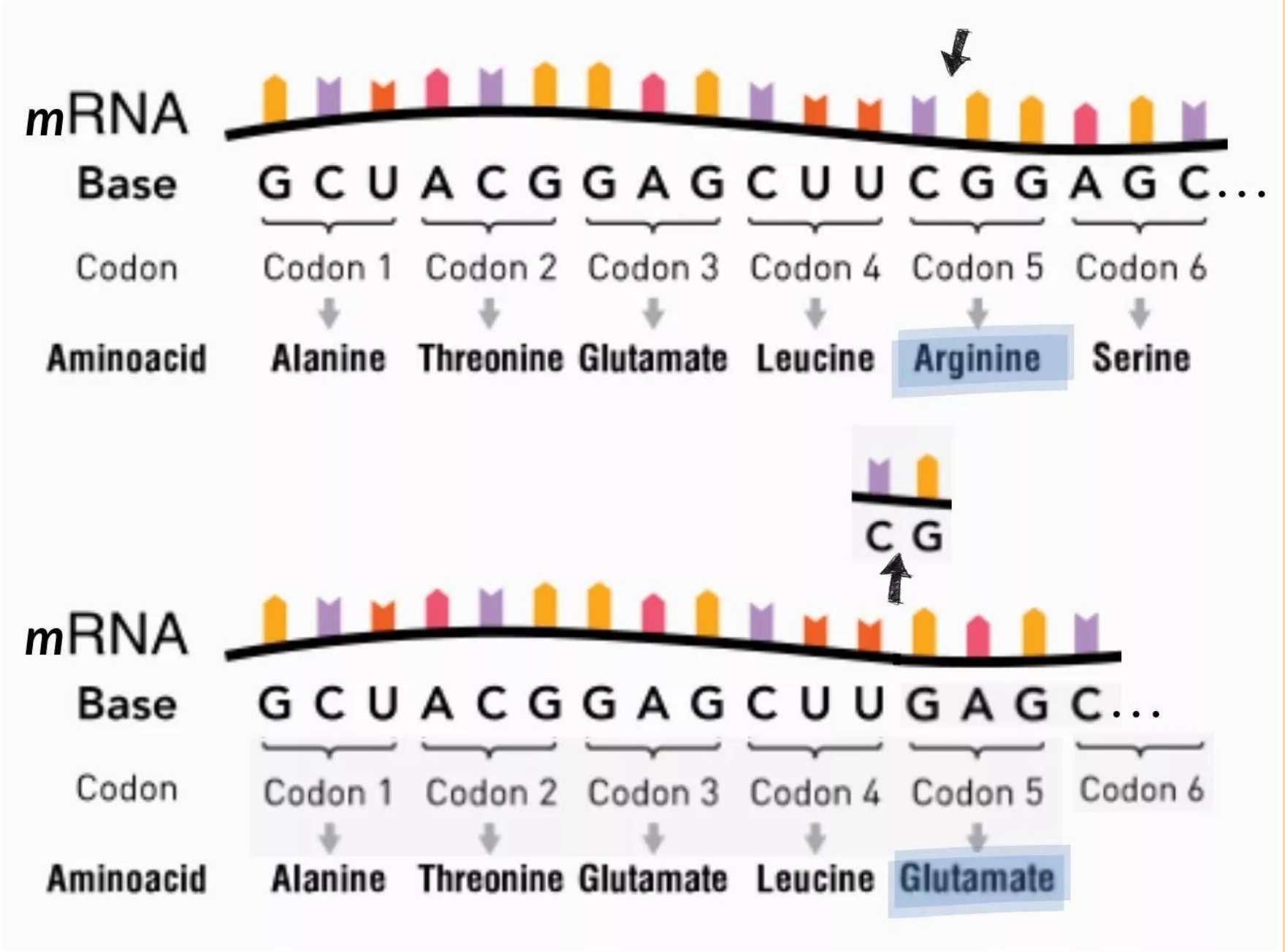
Nonsense mutation



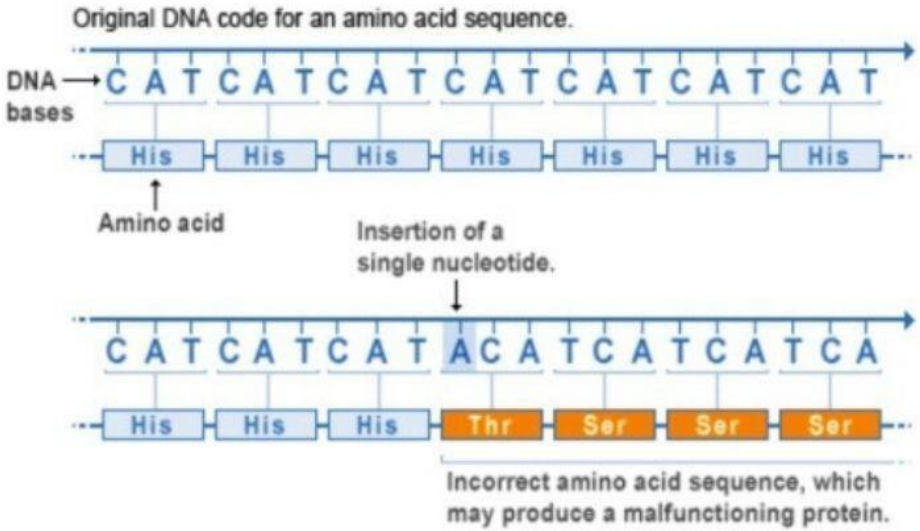
U.S. National Library of Medicine

Point mutations
Frameshift mutations

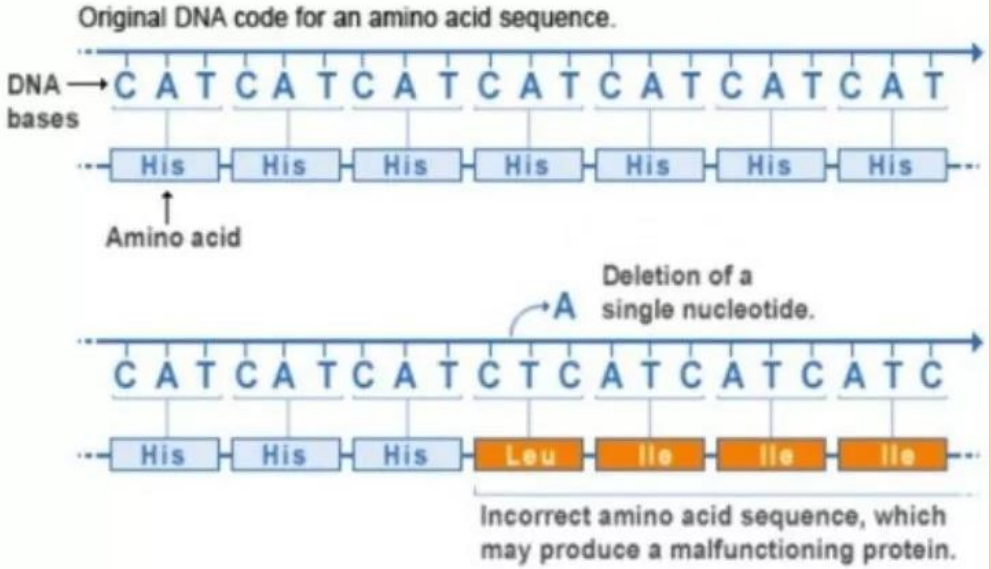
- Insertion of additional or deletion of bases in the DNA sequence
- Number other than 3 or multiples thereof - shift of reading frame, reading of other codons, change of protein sequence
- Serious effect on encoded proteins


mRNA
Base
G C U A C G G A G C U U C G G A G C...
Codon
Codon 1
Codon 2
Codon 3
Codon 4
Codon 5
Codon 6
Aminoacid
Alanine
Threonine
Glutamate
Leucine
Arginine
Serine
C G
mRNA
Base
G C U A C G G A G C U U G A G C...
Codon
Codon 1
Codon 2
Codon 3
Codon 4
Codon 5
Codon 6
Aminoacid
Alanine
Threonine
Glutamate
Leucine
Glutamate

Insertion mutation

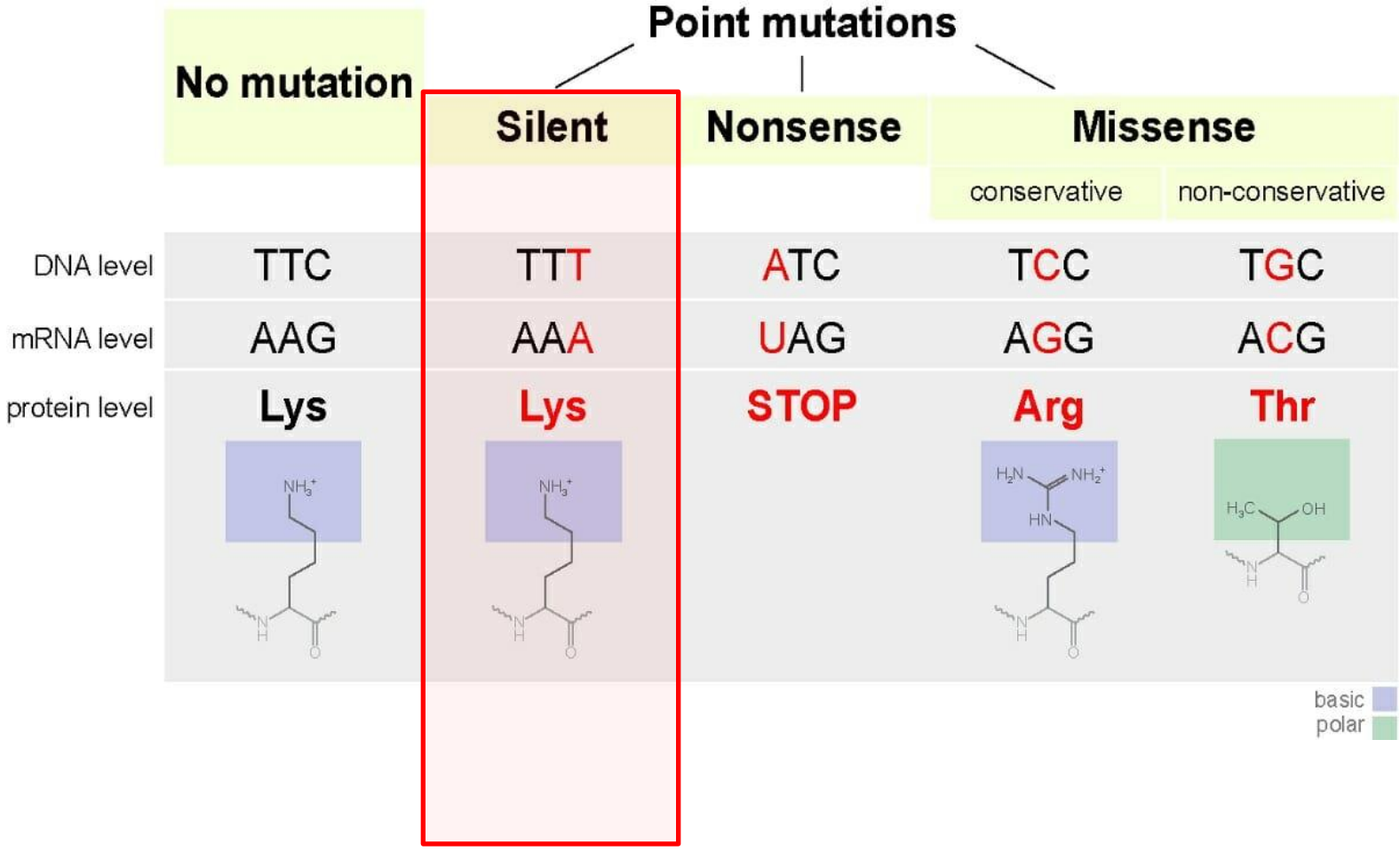


Deletion mutation



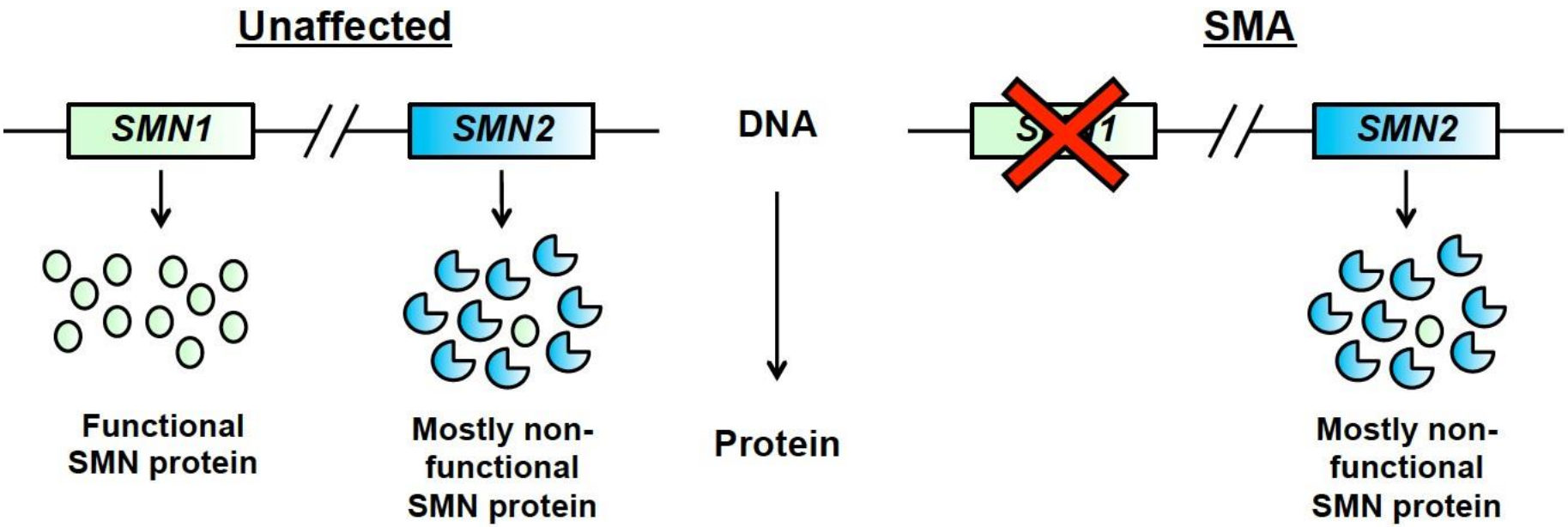
Point mutations
Silent mutations

- Mutation in the third codon rule - no amino acid substitution occurs
- No effect on encoded proteins
- Propensity to accumulate in DNA of organisms - polymorphisms


Point mutations
No mutation
Silent
Nonsense
Missense
conservative
non-conservative
DNA level
TTC
TTT
ATC
TCC
TGC
mRNA level
AAG
AAA
UAG
AGG
ACG
protein level
Lys
Lys
STOP
Arg
Thr
basic
polar

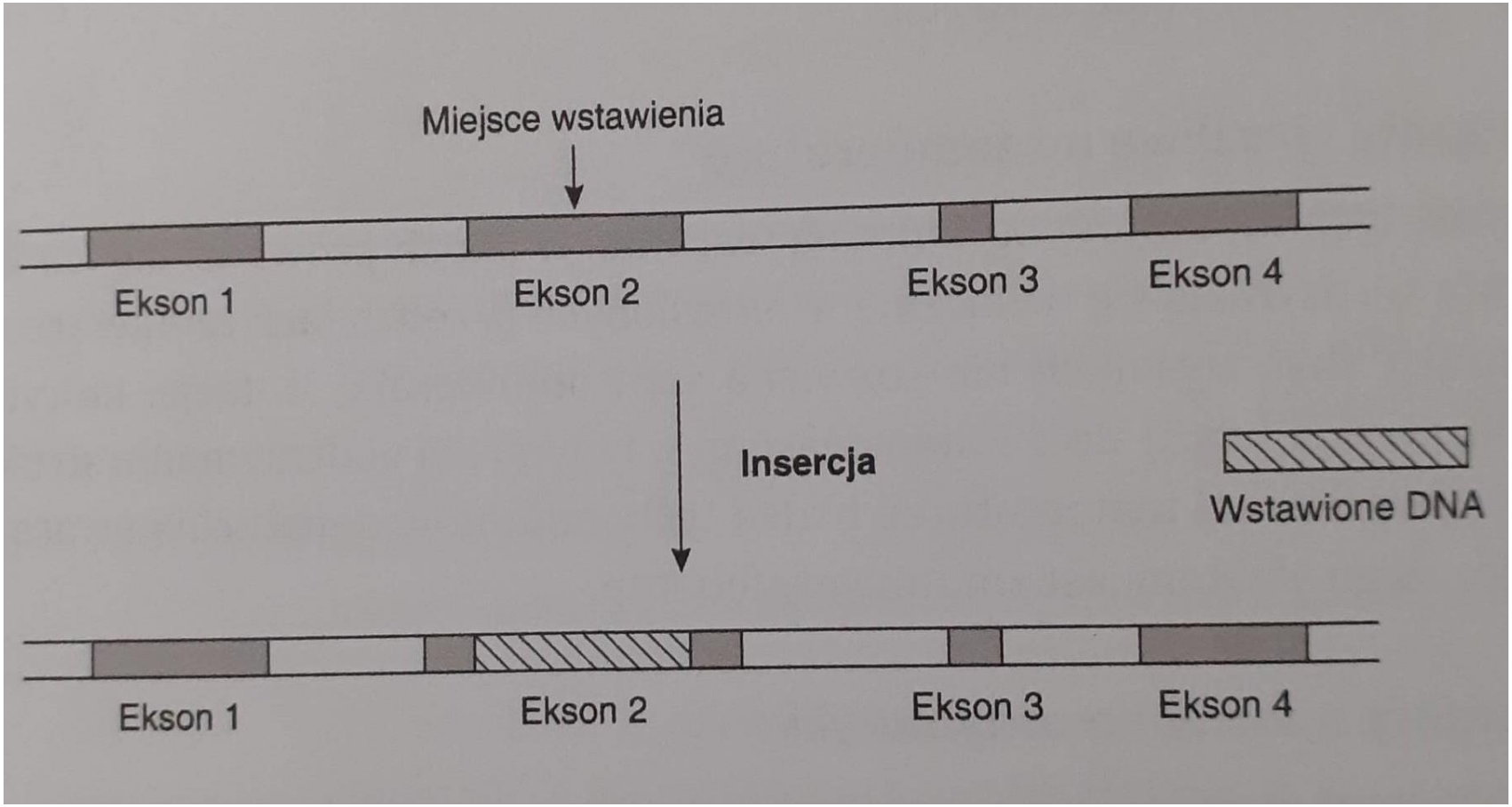
Broad mutations
Deletions

- Loss of part of DNA sequence
- Range - from single nucleotides to entire gene


Unaffected
SMN1
SMN2
DNA
SMA
SMN2
Functional SMN protein
Mostly non-functional SMN protein
Protein
Mostly non-functional SMN protein

Broad mutations
Insertions

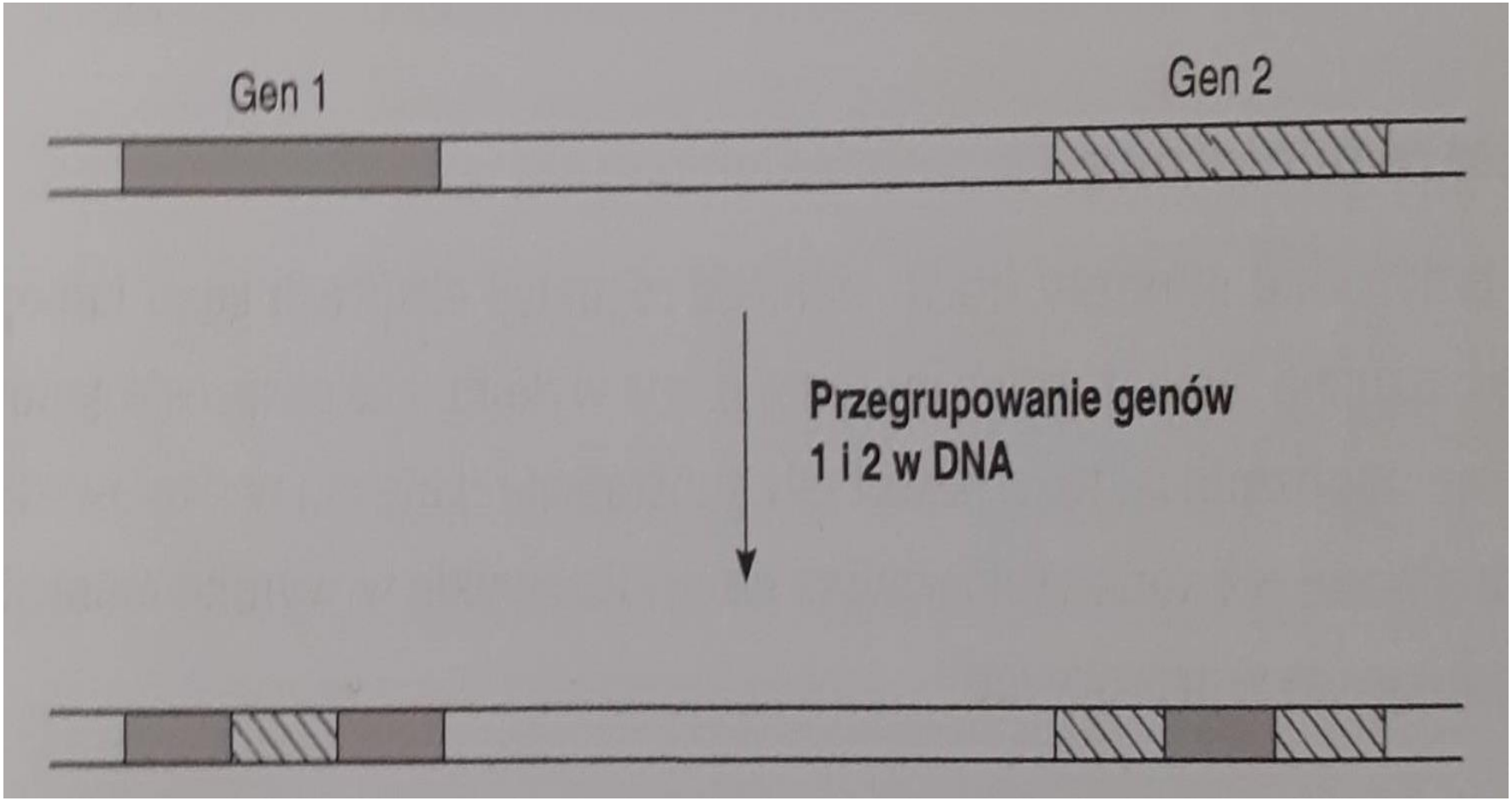
- Inclusion of additional nucleotides
- Usually from other parts of the chromosome
- Range - from single nucleotides


Miejsce wstawienia
Ekson 1
Ekson 2
Ekson 3
Ekson 4
Insercja
Wstawione DNA
Ekson 1
Ekson 2
Ekson 3
Ekson 4

Broad mutations
Regroupings

- Swap of places of DNA sequence segments belonging to one gene or coming from outside the gene

- Example - inversion


Gen 1
Gen 2
Przegrupowanie genów
1 i 2 w DNA