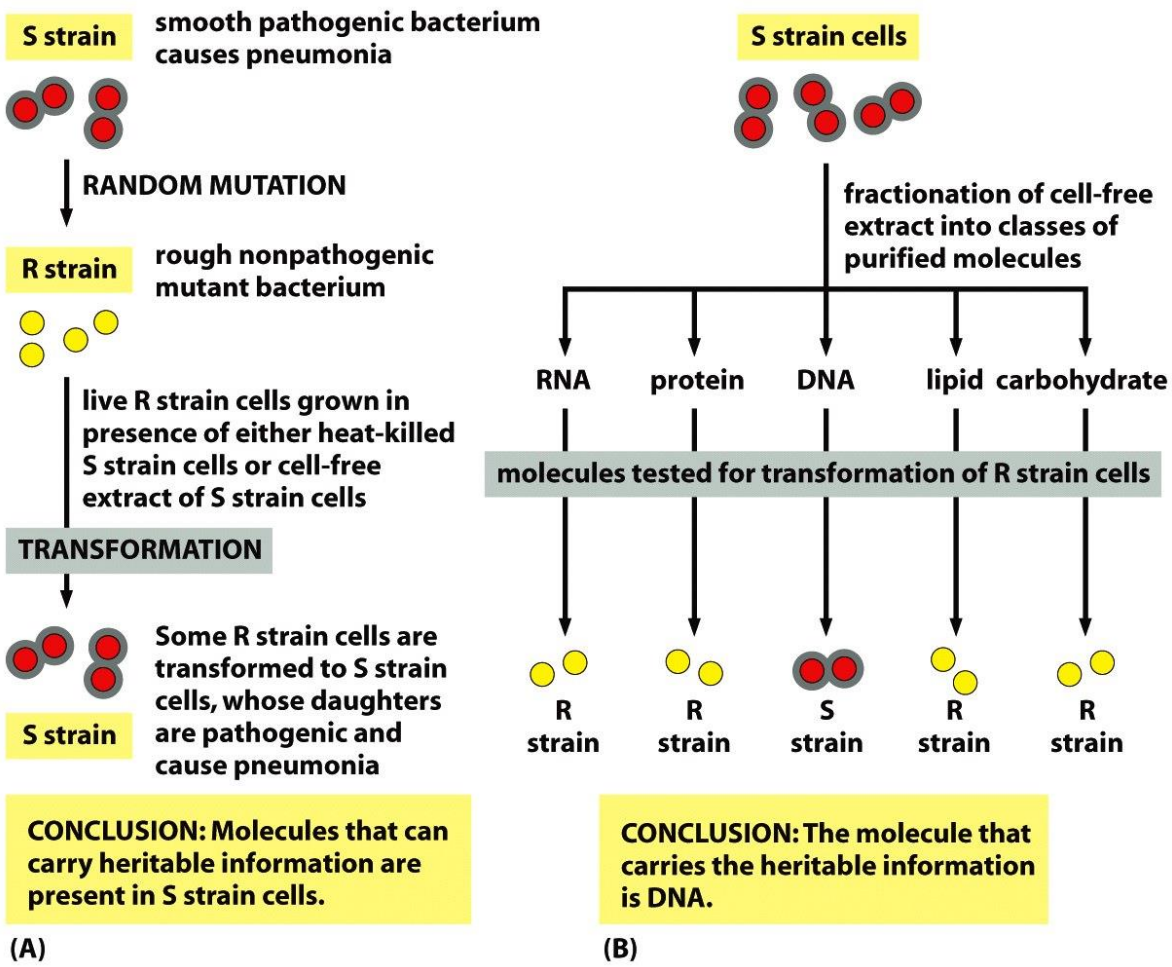
The structure and function of biopolymers during the transitions of genetic information

Daniel Renčiuk

VFU, Brno 05.10.2016

Genetic information is coded by DNA

The experiment combining two strains of Streptococcus pneumoniae bacteria.



O. Avery
C. MacLeod
M.McCarty

STUDIES ON THE CHEMICAL NATURE OF THE SUBSTANCE INDUCING TRANSFORMATION OF PNEUMOCOCCAL TYPES

INDUCTION OF TRANSFORMATION BY A DESOXYRIBONUCLEIC ACID FRACTION ISOLATED FROM PNEUMOCOCCUS TYPE III

BY OSWALD T. AVERY, M.D., COLIN M. MACLEOD, M.D., AND MACLYN MCCARTY,* M.D.

(*From the Hospital of The Rockefeller Institute for Medical Research*)

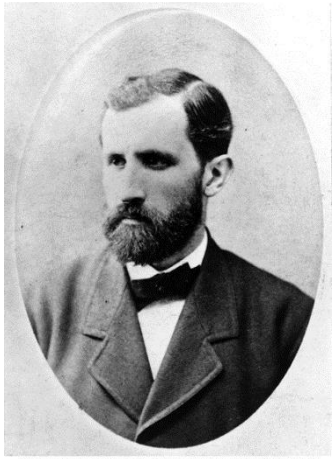
PLATE 1

(Received for publication, November 1, 1943)

Biologists have long attempted by chemical means to induce in higher organisms predictable and specific changes which thereafter could be transmitted in series as hereditary characters. Among microörganisms the most striking example of inheritable and specific alterations in cell structure and function that can be experimentally induced and are reproducible under well defined and adequately controlled conditions is the transformation of specific types of Pneumococcus. This phenomenon was first described by Griffith (1) who succeeded in transforming an attenuated and non-encapsulated (R) variant derived from one specific type into fully encapsulated and virulent (S) cells of a heterologous specific type. A typical instance will suffice to illustrate the techniques originally used and serve to indicate the wide variety of transformations that are possible within the limits of this bacterial species.

Nuclein

- Nuclein – acidic substance rich in nitrogen and phosphorus
- J. F. Miescher - 1864
- Isolated from blood of wounded patients and cleaved by pepsin (proteolytic enzyme)



Roles of genetic material

Genotype role – storage of genetic information and its transition to the offspring

Fenotype role – expression of genetic information to a particular properties of an individual

Evolutionary role – adaptation of an organism/species to the environment through the changes in genetic information

Terminology

Gene – several "definitions" depending on the point of view:

- classic genetics (Mendel) – elementary unit of hereditary genetic information
- molecular genetics – part of DNA coding for RNA (and as a consequence coding for some property of the individual)
  - structural genes coding for mRNA/protein (+ regulatory regions)
  - genes coding for functional RNA (miRNA, …)
- strict – structural gene – part of DNA that codes for protein sequence

Allele – particular variant of the respective gene

Genome – complete DNA of organism (molecular) x complete genetic information of organism x sum of genes (classic)

Genotype – the combination of particular alleles of all genes in individual

Phenotype – the sum of actual individual properties (as a result of expression of particular genotype in the respective environment)

Genophor – the carrier of genetic information, usually a molecule of DNA (often used for bacteria)

Information biopolymers

Deoxyribonucleic acid (DNA)

• linear heteropolymer composed from 2-deoxyribonucleotides connected by fosfodiester bonds
• usually as a stable and resistant double helix
• serves as a storage of genetic information, as a template for its reproduction (replication) and as a template for the expression of genetic information to the fenotype (transcription)

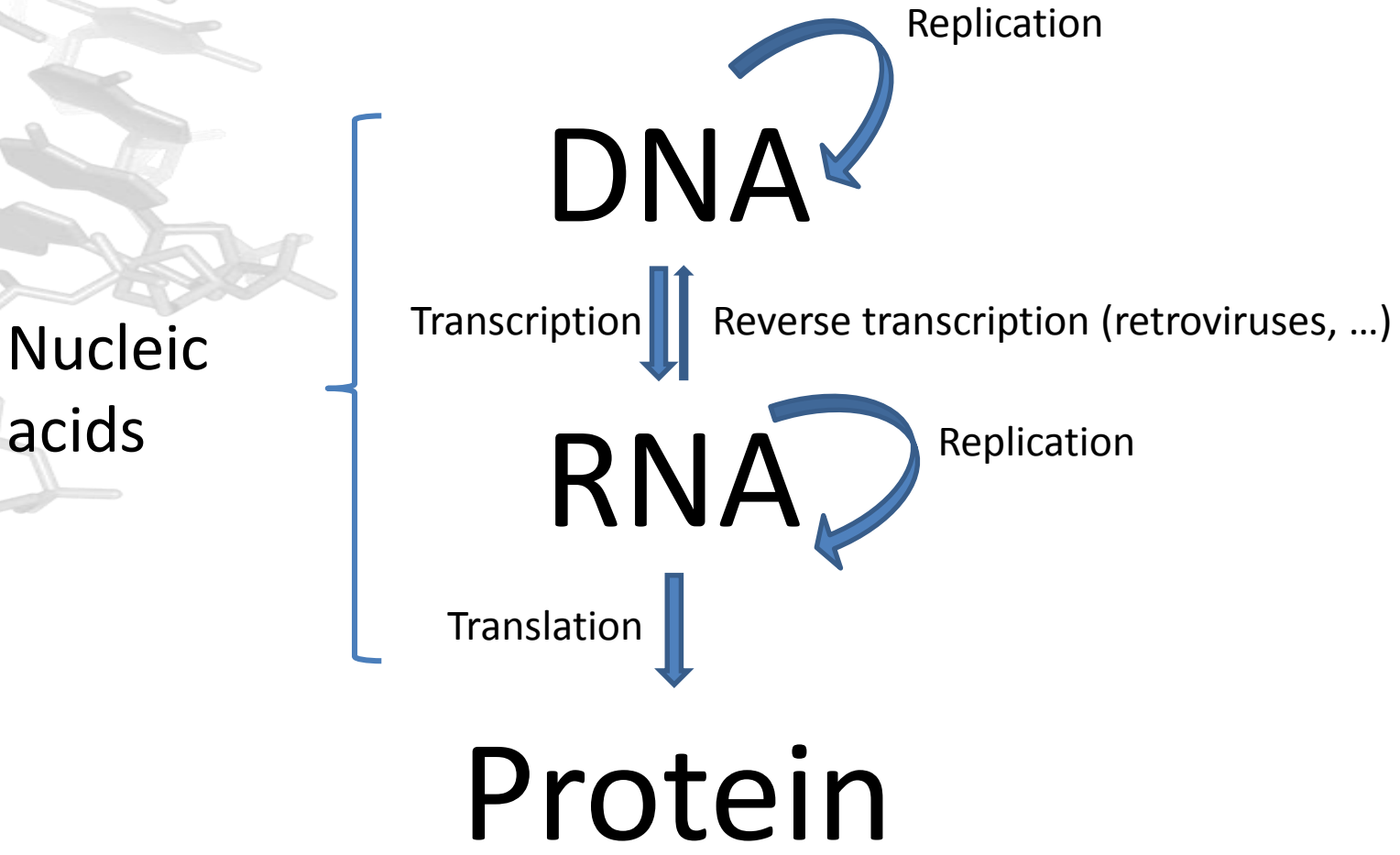
Ribonucleic acid (RNA)

• linear heteropolymer composed from ribonucleotides connected by fosfodiester bonds
• usually as a single-stranded structure of variable length, structure and reactivity
• many functions depending on type of RNA (see later)

Protein

• linear heteropolymer composed from 20 (21) amino acids connected by peptidic bonds
• highly variable structures, properties and functions

Central dogma of molecular biology



Gene

POU5F1 POU class 5 homeobox 1 [*Homo sapiens* (human)]

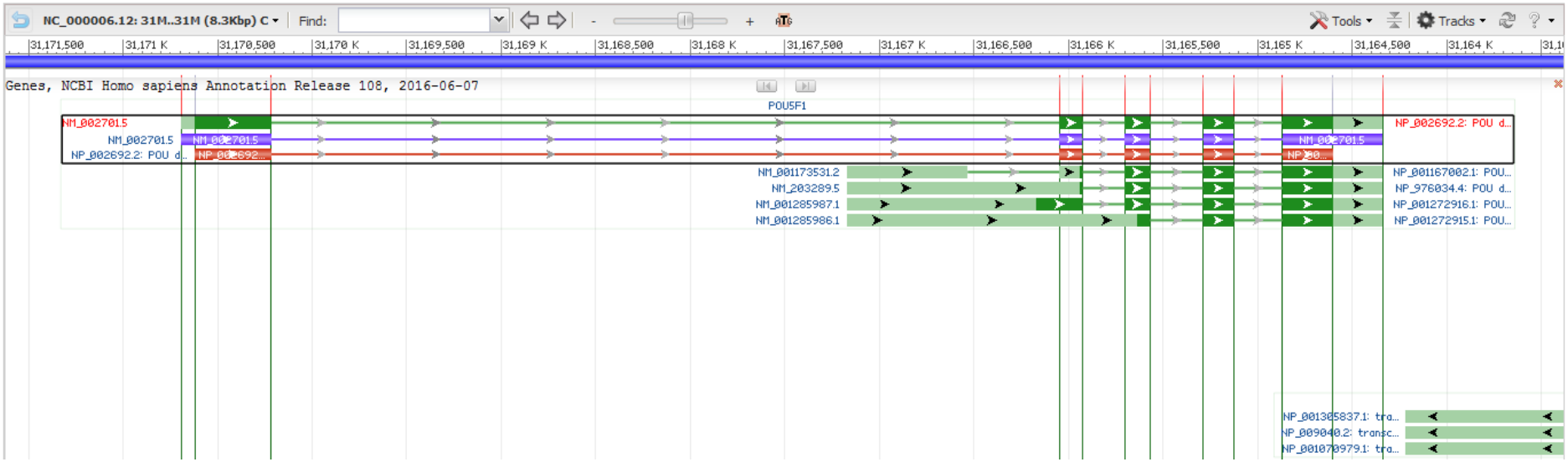
Gene ID: 5460, updated on 17-Sep-2016

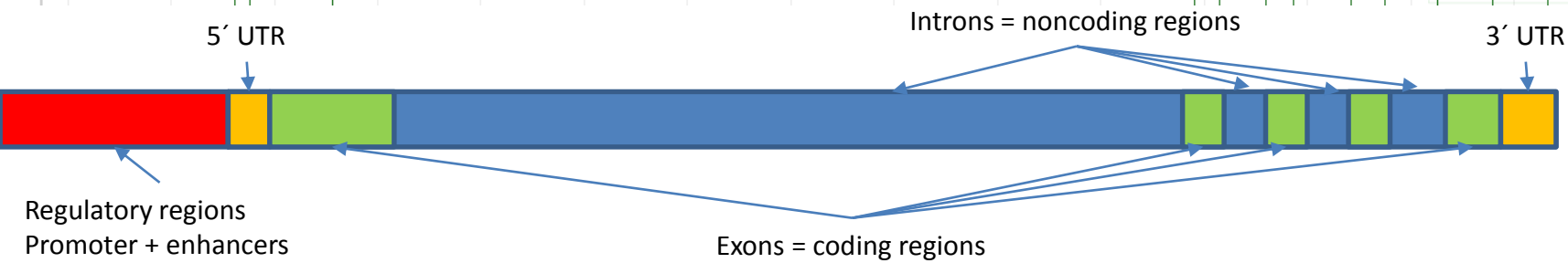
Summary

Official Symbol	POU5F1 provided by HGNC
Official Full Name	POU class 5 homeobox 1 provided by HGNC
Primary source	HGNC:HGNC:9221
See related	Ensembl:ENSG00000204531 HPRD:01252; MIM:164177; Vega:OTTHUMG00000031206
Gene type	protein coding
RefSeq status	REVIEWED
Organism	Homo sapiens
Lineage	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo
Also known as	OCT3; OCT4; OTF3; OTF4; OTF-3; Oct-3; Oct-4
Summary	This gene encodes a transcription factor containing a POU homeodomain that plays a key role in embryonic development and stem cell pluripotency. Aberrant expression of this gene in adult tissues is associated with tumorigenesis. This gene can participate in a translocation with the Ewing's sarcoma gene on chromosome 21, which also leads to tumor formation. Alternative splicing, as well as usage of alternative AUG and non-AUG translation initiation codons, results in multiple isoforms. One of the AUG start codons is polymorphic in human populations. Related pseudogenes have been identified on chromosomes 1, 3, 8, 10, and 12. [provided by RefSeq, Oct 2013]
Orthologs	mouse all

https://www.ncbi.nlm.nih.gov/gene/5460

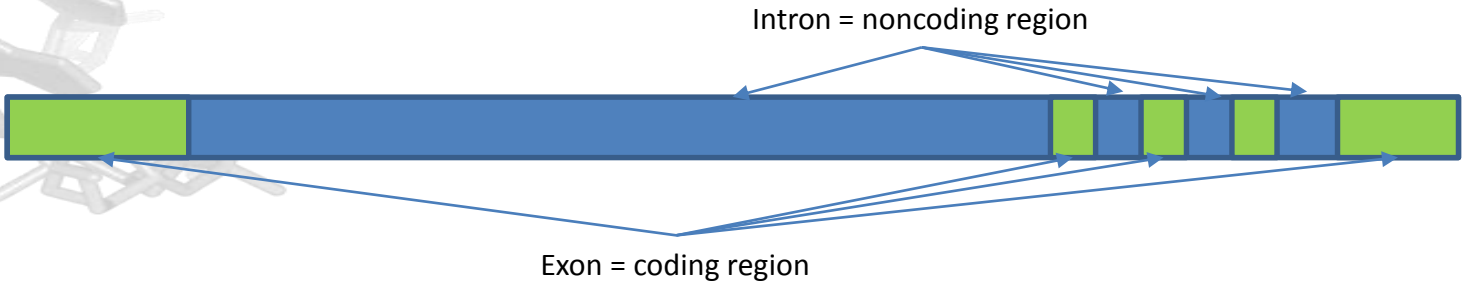
Go to nucleotide: Graphics FASTA GenBank





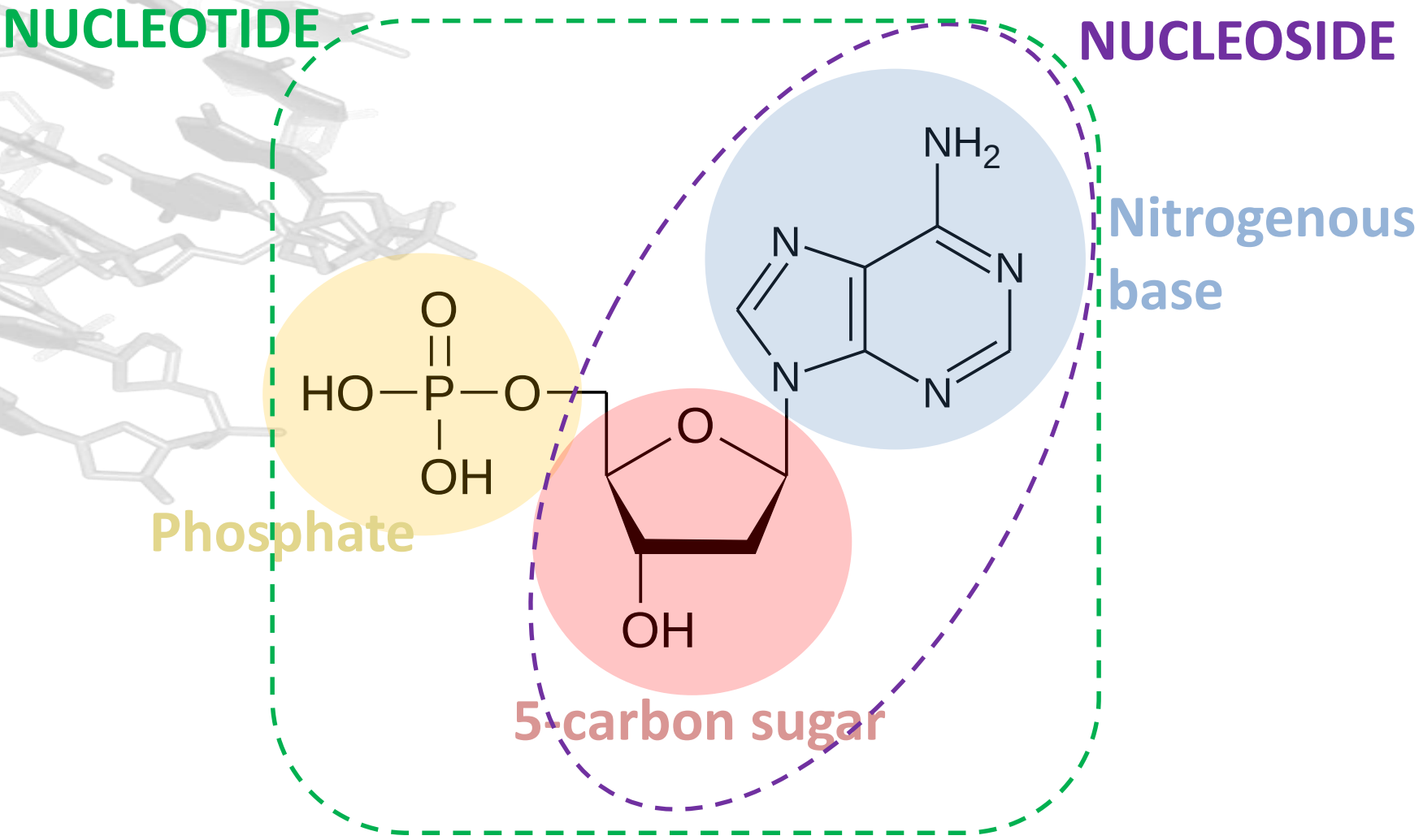
Genes

Most **Eucaryotic** genes contain **introns**, that are transcribed into primary RNA transcript and introns are consecutively removed by **splicing** process on **spliceosome** to form final **mRNA**.



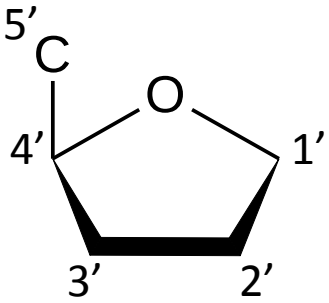
Procaryotic genes do not contain **introns** and they are directly transcribed into mRNA that serves as a template for translation into protein.

Nucleic acids

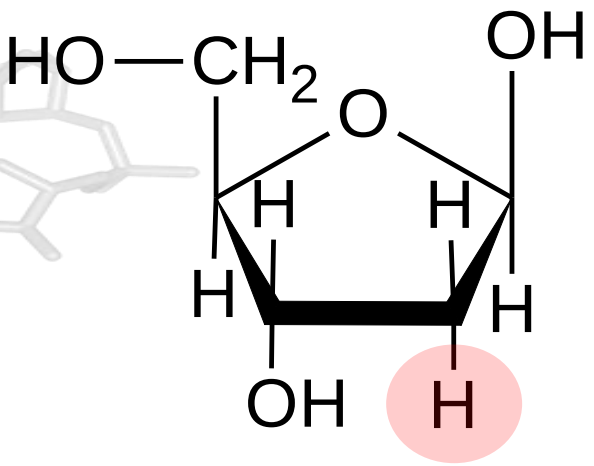


5-carbon sugar - pentose

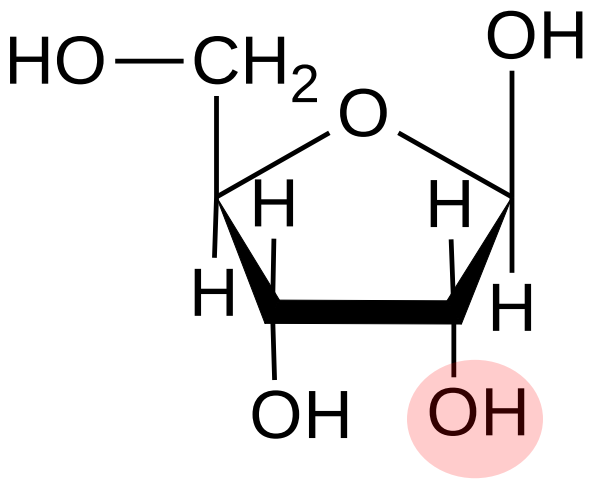
DNA



RNA

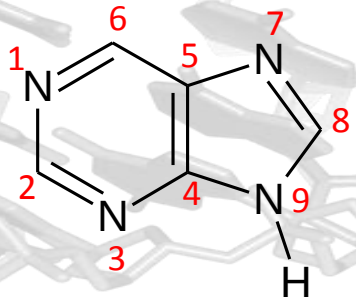


β-D-2-deoxyribose

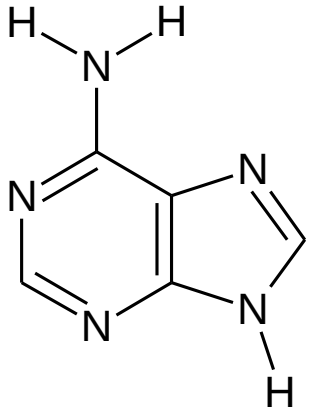


β-D-ribose

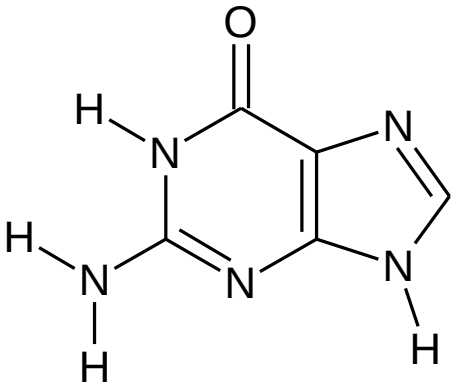
Base



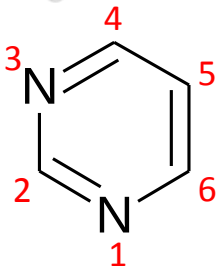
Purine



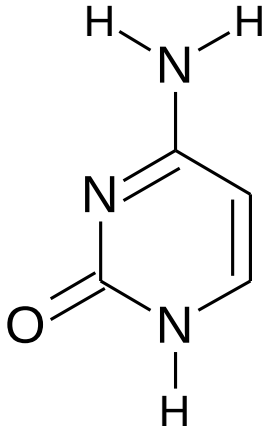
Adenine



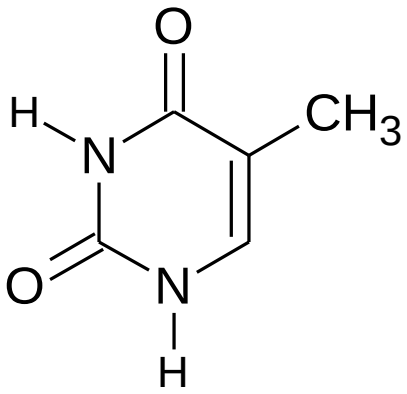
Guanine



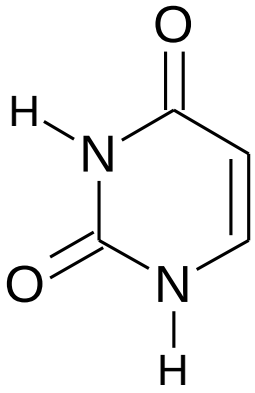
Pyrimidine



Cytosine



Tymine



Uracil

Nucleic acid nomenclature

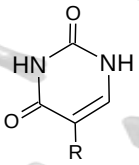
	Báze	Nukleosid	Nukleotid NTP
G	Guanine	Guanosine	Guanosine triphosphate
A	Adenine	Adenosine	Adenosine triphosphate
T	Tymine	Tymidine	Tymidine triphosphate
C	Cytosine	Cytidine	Cytidine triphosphate
U	Uracil	uridine	Uridine triphosphate

DNA – prefix deoxy-

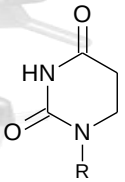
NMP = monophosphate
NDP = diphosphate

Modified bases

tRNA + mRNA



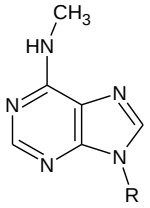
pseudouracil



dihydrouracil

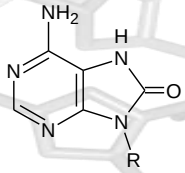
metylguanine

metylinosine

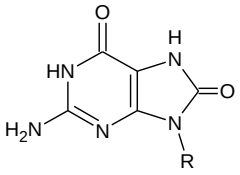


metyladenine

Oxidative damage



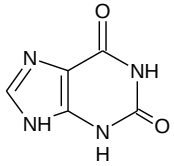
8-oxo adenine



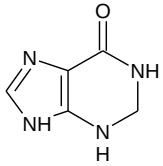
8-oxo guanine

5-hydroxymethyl uracil

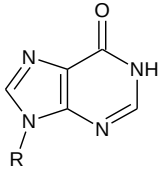
Metabolism



xanthine

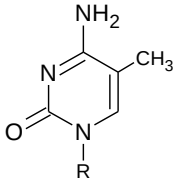


hypoxanthine

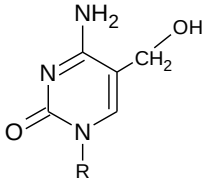


inosine

Epigenetics

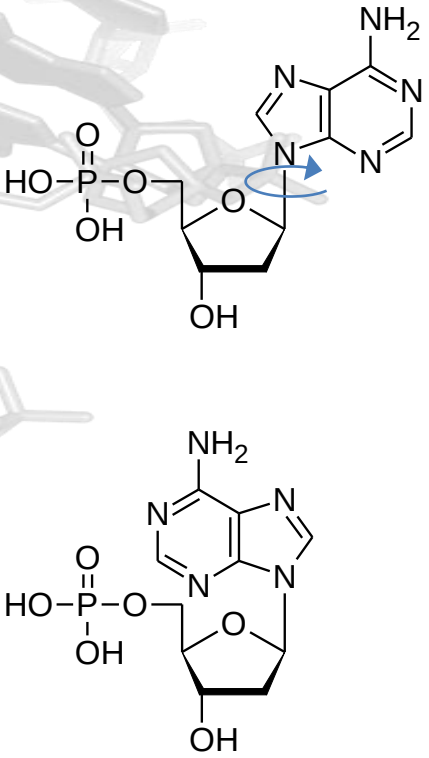


5-methyl cytosine



5-hydroxymethyl cytosine

Conformation of N-glykosidic bond



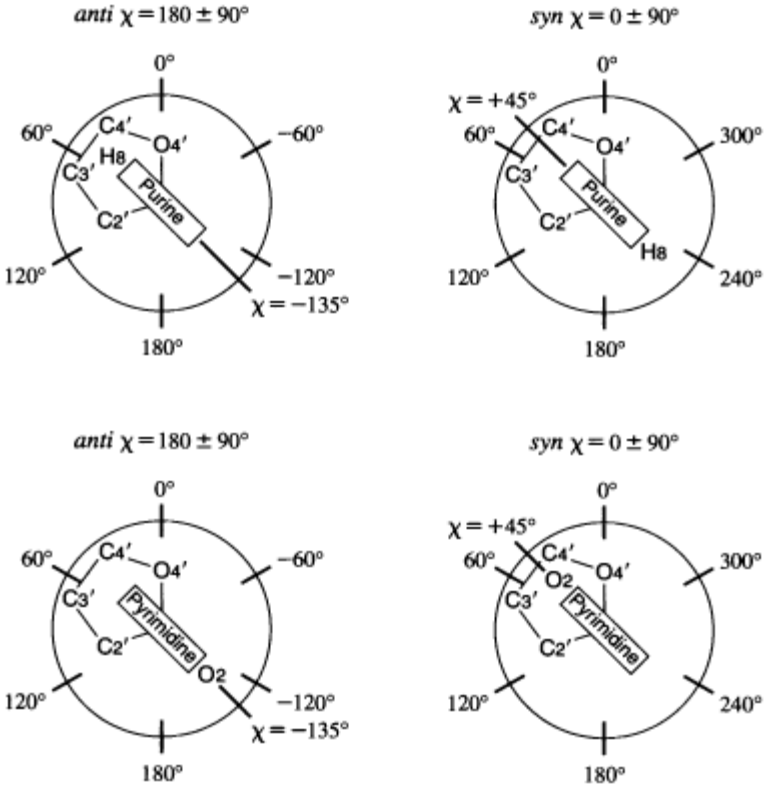
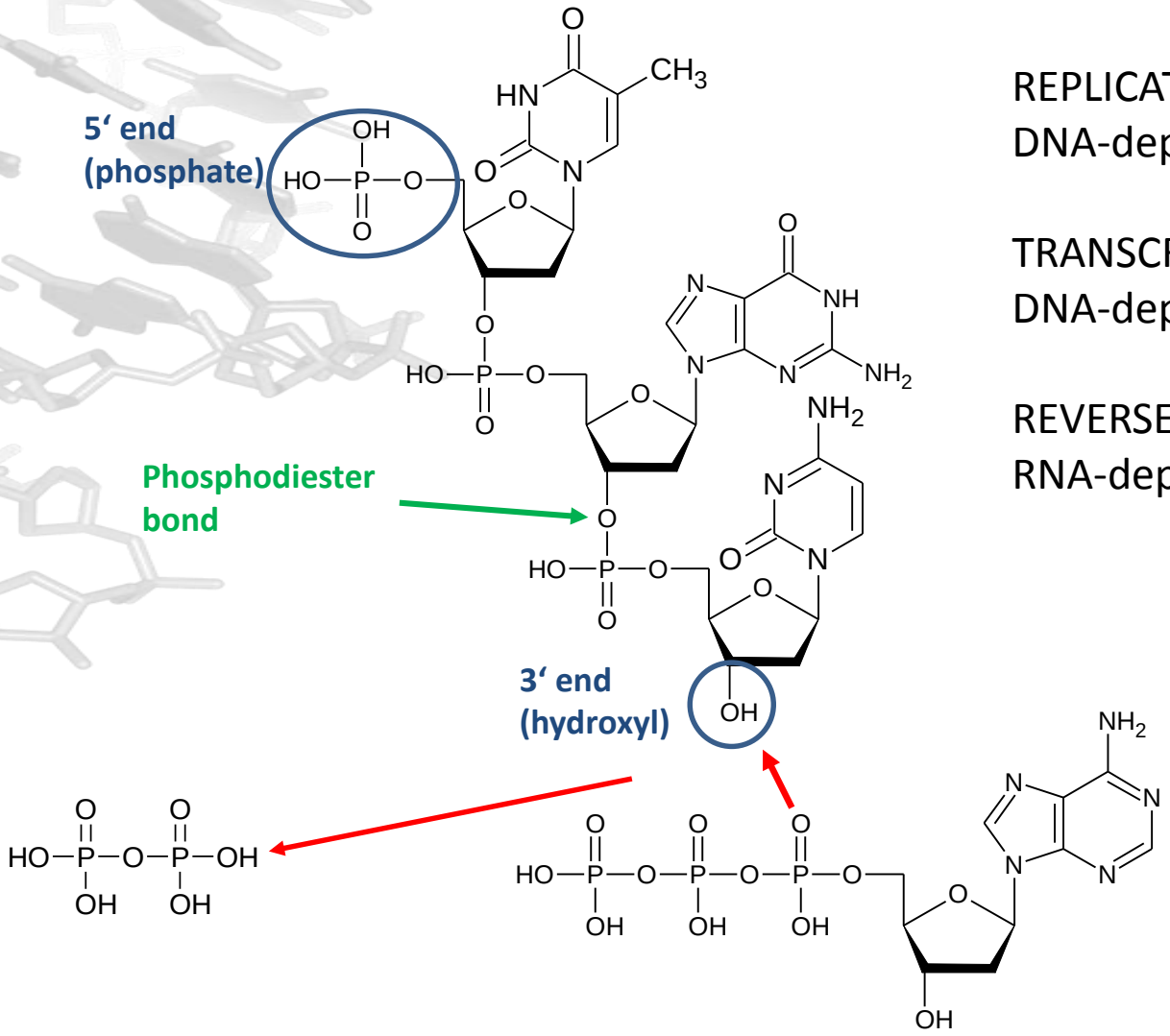


Figure 2-8
The glycosidic torsion angle χ is defined by O4′– C1′–N9–C4 for purines and O4′–C1′–N1–C2 for pyrimidines. When $\chi = 0°$ the O4′–C1′ is eclipsed by the N9–C4 bond in purines and by the N1–C2 bond in pyrimidines. The syn conformations correspond to $0° \pm 90°$; anti conformations correspond to $180° \pm 90°$. In nucleotides steric hindrance limits the conformations actually found to a much narrower range of angles that depend on sugar pucker and base. The syn conformations are usually found with $\chi = 45° \pm 45°$; anti conformations are usually found with $\chi = -135° \pm 45°$.

Formation of sugar-phosphate backbone



REPLICATION
DNA-dependent DNA polymerase

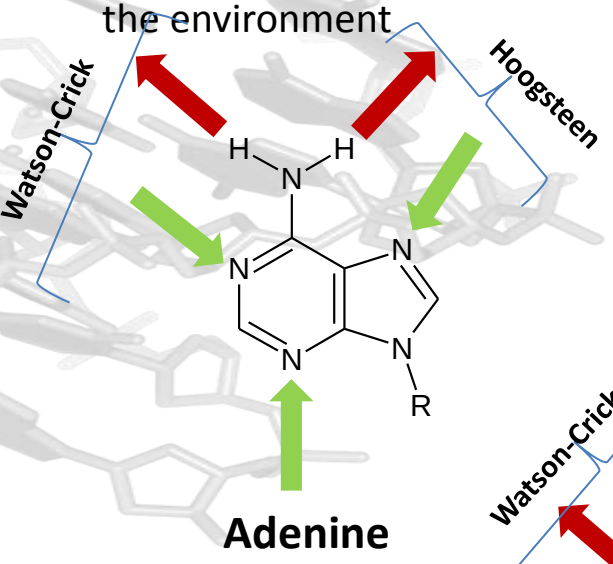
TRANSCRIPTION
DNA-dependent RNA polymerase

REVERSE TRANSCRIPTION
RNA-dependent DNA polymerase

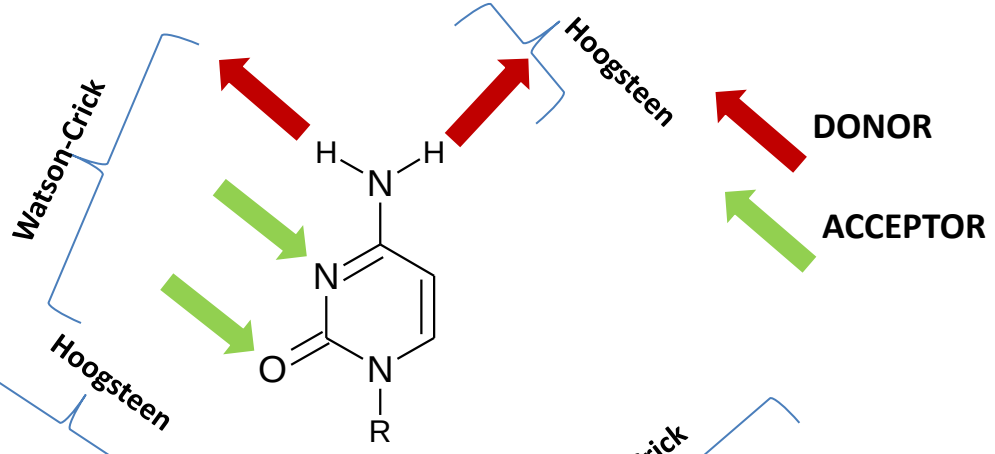
Base reactivity – hydrogen bonds

Hydrogen bond – weak electrostatic interaction of two polar groups – one covalently bonds hydrogen (DONOR – usually -N-H a -O-H); the second (ACCEPTOR) is usually N or O

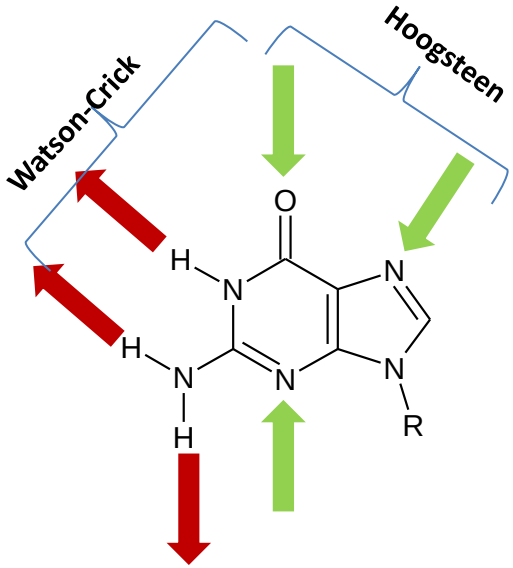
Length: ~ 2.8 A (2 - 3,4 A) Energy: < 1 kcal/mol both depends particular atoms and on the environment



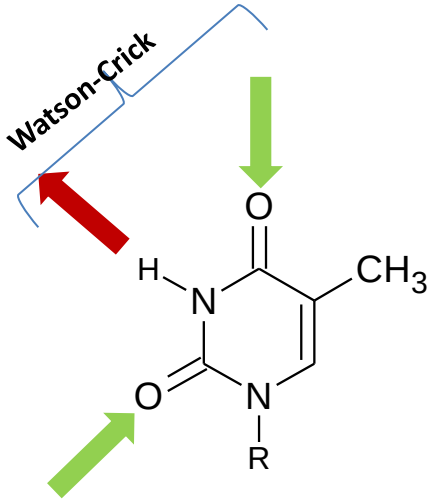
Adenine



Cytosine

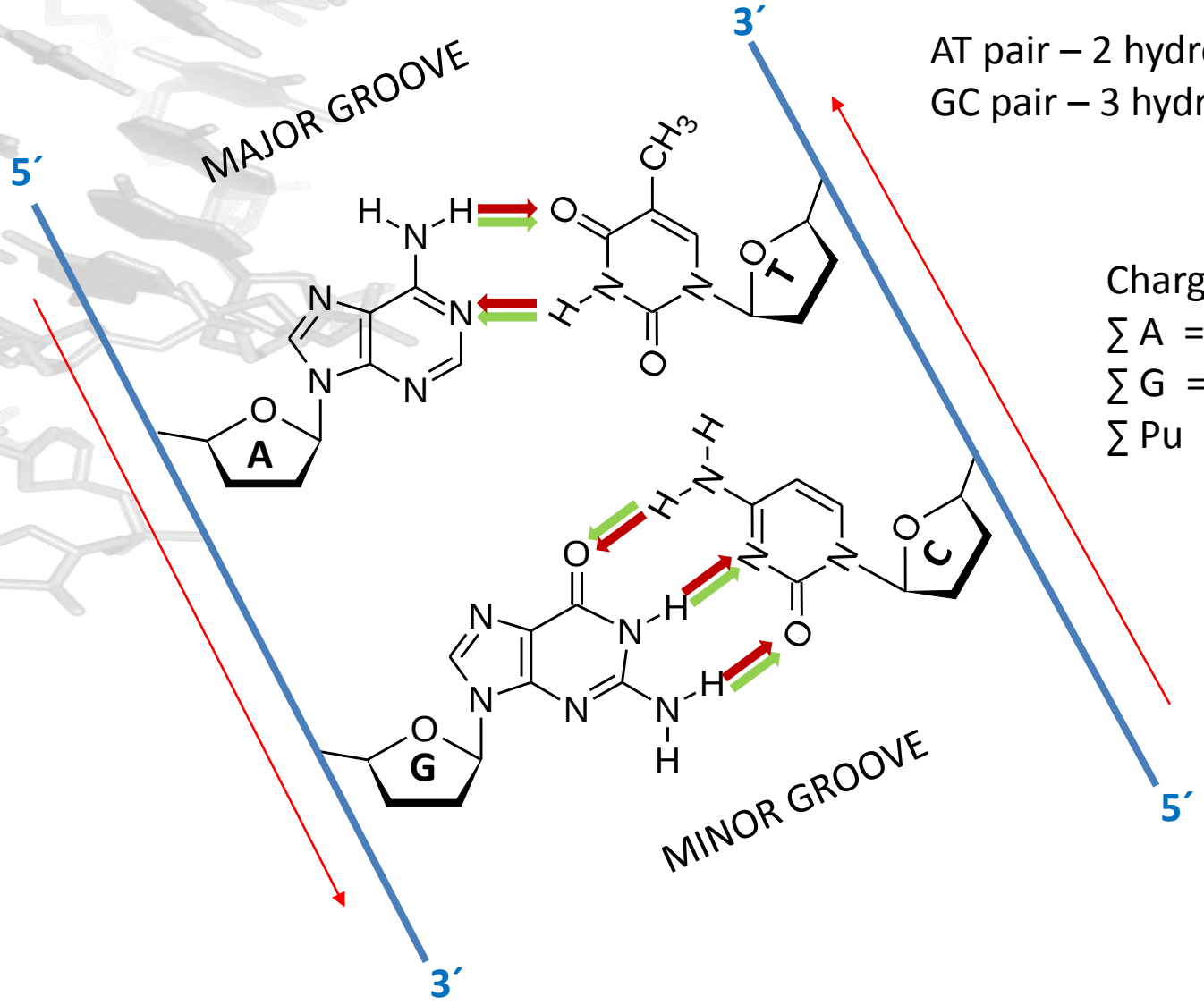


Guanine



Tymine

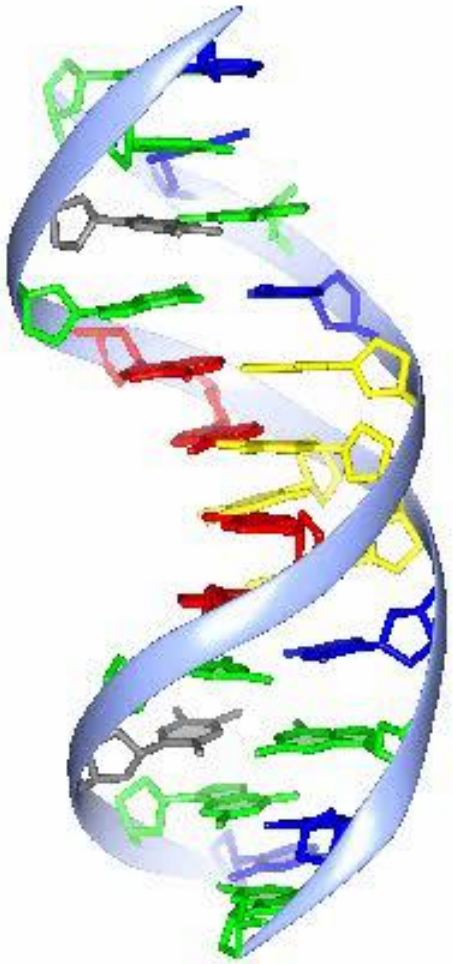
Watson-Crick base pairing

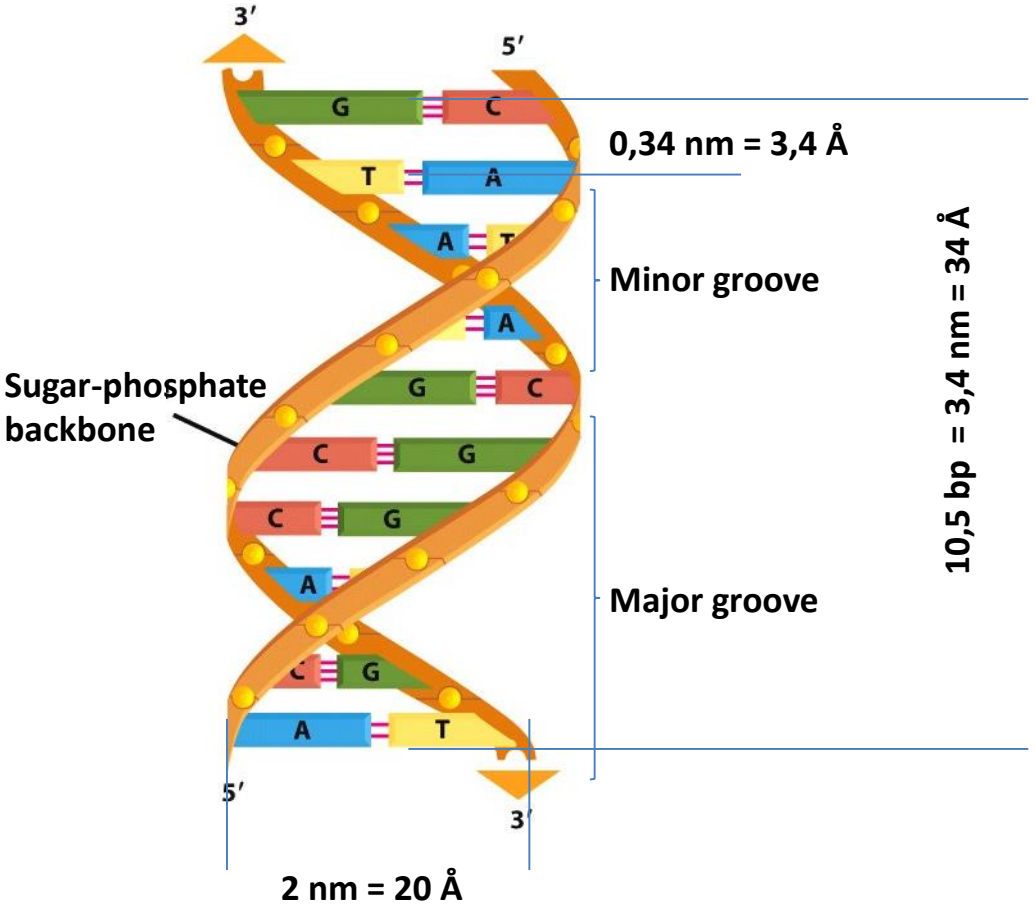


AT pair – 2 hydrogen bonds
GC pair – 3 hydrogen bonds

Chargaff rules:
$\sum A = \sum T$
$\sum G = \sum C$
$\sum Pu = \sum Py$

DNA double helix





DNA double helix

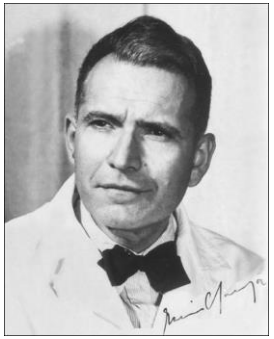
- **two** molecules (**strands**) of DNA
- the helix is **right-handed**
- the strands go **antiparalel** – their 5´-3´ direction is opposite in context of the double helix direction
- similar content of purines and pyrimidines; content of A = T, G = C (Chargaff rules)
- result – the strands are **complementary** – i.e. according to the Watson-Crick base pairing rules we can predict/create the sequence of one strand according to the sequence of the other
- on average the double helix contains **10,5 base pairs** per turn of the helix, which is about **3,4 nm** in length

DNA structure –Watson and Crick model

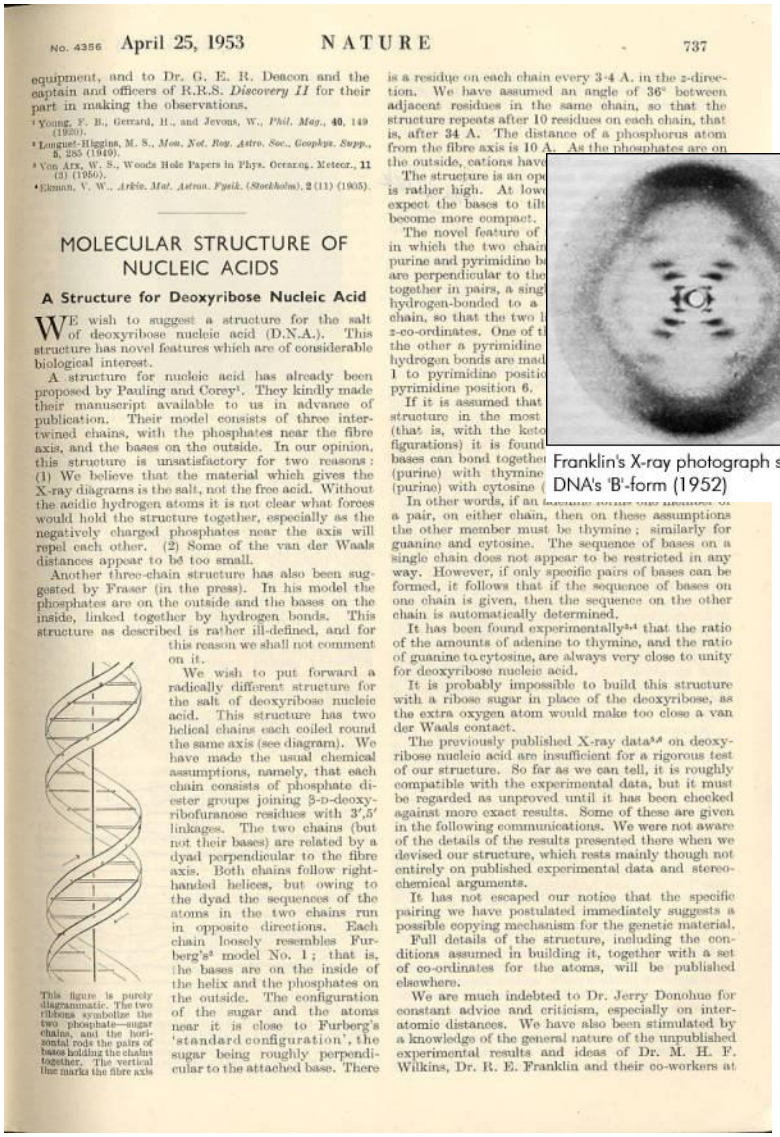
F. Crick J.D. Watson

M. Wilkins

R. Franklin



E. Chargaff



No. 4356 April 25, 1953 NATURE 737

equipment, and to Dr. G. E. R. Deacon and the captain and officers of R.R.S. *Discovery II* for their part in making the observations.

[1] Young, F. B., Gerrard, H., and Jevons, W., *Phil. Mag.*, **40**, 149 (1920).
[2] Longuet-Higgins, M. S., *Mon. Not. Roy. Astro. Soc., Geophys. Supp.*, **5**, 285 (1949).
[3] Van Atta, W. S., Woods Hole Papers in Phys. Oceanog. Meteor., **11** (3) (1950).
[4] Ekman, V. W., *Arkiv. Mat. Astron. Fysik. (Stockholm)*, **2** (11) (1905).

MOLECULAR STRUCTURE OF NUCLEIC ACIDS

A Structure for Deoxyribose Nucleic Acid

WE wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest.

A structure for nucleic acid has already been proposed by Pauling and Corey[1]. They kindly made their manuscript available to us in advance of publication. Their model consists of three inter-twined chains, with the phosphates near the fibre axis, and the bases on the outside. In our opinion, this structure is unsatisfactory for two reasons: (1) We believe that the material which gives the X-ray diagrams is the salt, not the free acid. Without the acidic hydrogen atoms it is not clear what forces would hold the structure together, especially as the negatively charged phosphates near the axis will repel each other. (2) Some of the van der Waals distances appear to be too small.

Another three-chain structure has also been suggested by Fraser (in the press). In his model the phosphates are on the outside and the bases on the inside, linked together by hydrogen bonds. This structure as described is rather ill-defined, and for this reason we shall not comment on it.

We wish to put forward a radically different structure for the salt of deoxyribose nucleic acid. This structure has two helical chains each coiled round the same axis (see diagram). We have made the usual chemical assumptions, namely, that each chain consists of phosphate di-ester groups joining β-D-deoxy-ribofuranose residues with 3′,5′ linkages. The two chains (but not their bases) are related by a dyad perpendicular to the fibre axis. Both chains follow right-handed helices, but owing to the dyad the sequences of the atoms in the two chains run in opposite directions. Each chain loosely resembles Furberg's[2] model No. 1; that is, the bases are on the inside of the helix and the phosphates on the outside. The configuration of the sugar and the atoms near it is close to Furberg's 'standard configuration', the sugar being roughly perpendicular to the attached base. There is a residue on each chain every 3·4 A. in the z-direction. We have assumed an angle of 36° between adjacent residues in the same chain, so that the structure repeats after 10 residues on each chain, that is, after 34 A. The distance of a phosphorus atom from the fibre axis is 10 A. As the phosphates are on the outside, cations have

The structure is an ope
is rather high. At low
expect the bases to tilt
become more compact.

The novel feature of
in which the two chair
purine and pyrimidine b
are perpendicular to the
together in pairs, a sing
hydrogen-bonded to a
chain, so that the two l
z-co-ordinates. One of t
the other a pyrimidine
hydrogen bonds are mad
1 to pyrimidine positio
pyrimidine position 6.

If it is assumed that
structure in the most
(that is, with the keto
figurations) it is found
bases can bond together
(purine) with thymine
(purine) with cytosine (

In other words, if an
a pair, on either chain, then on these assumptions the other member must be thymine; similarly for guanine and cytosine. The sequence of bases on a single chain does not appear to be restricted in any way. However, if only specific pairs of bases can be formed, it follows that if the sequence of bases on one chain is given, then the sequence on the other chain is automatically determined.

It has been found experimentally[3,4] that the ratio of the amounts of adenine to thymine, and the ratio of guanine to cytosine, are always very close to unity for deoxyribose nucleic acid.

It is probably impossible to build this structure with a ribose sugar in place of the deoxyribose, as the extra oxygen atom would make too close a van der Waals contact.

The previously published X-ray data[3,4] on deoxy-ribose nucleic acid are insufficient for a rigorous test of our structure. So far as we can tell, it is roughly compatible with the experimental data, but it must be regarded as unproved until it has been checked against more exact results. Some of these are given in the following communications. We were not aware of the details of the results presented there when we devised our structure, which rests mainly though not entirely on published experimental data and stereo-chemical arguments.

It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.

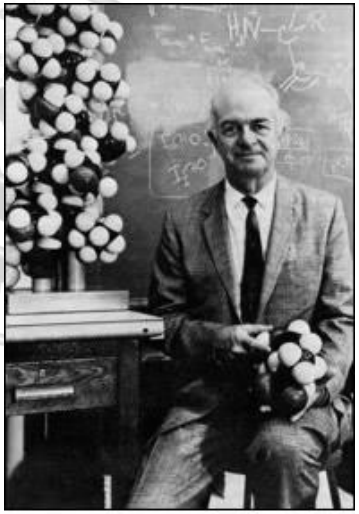
Full details of the structure, including the conditions assumed in building it, together with a set of co-ordinates for the atoms, will be published elsewhere.

We are much indebted to Dr. Jerry Donohue for constant advice and criticism, especially on inter-atomic distances. We have also been stimulated by a knowledge of the general nature of the unpublished experimental results and ideas of Dr. M. H. F. Wilkins, Dr. R. E. Franklin and their co-workers at

This figure is purely diagrammatic. The two ribbons symbolize the two phosphate—sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis

Franklin's X-ray photograph shows DNA's 'B'-form (1952)

DNA structure – Pauling model



Linus Pauling

A PROPOSED STRUCTURE FOR THE NUCLEIC ACIDS

By Linus Pauling and Robert B. Corey

Gates and Crellin Laboratories of Chemistry,* California Institute of Technology

Communicated December 31, 1952

The nucleic acids, as constituents of living organisms, are comparable in importance to the proteins. There is evidence that they are involved in the processes of cell division and growth, that they participate in the transmission of hereditary characters, and that they are important constituents of viruses. An understanding of the molecular structure of the nucleic acids should be of value in the effort to understand the fundamental phenomena of life.

which are involved in ester linkages. This distortion of the phosphate group from the regular tetrahedral configuration is not supported by direct experimental evidence; unfortunately no precise structure determinations have been made of any phosphate di-esters. The distortion, which corresponds to a larger amount of double bond character for the inner oxygen atoms than for the oxygen atoms involved in the ester linkages, is a reason-

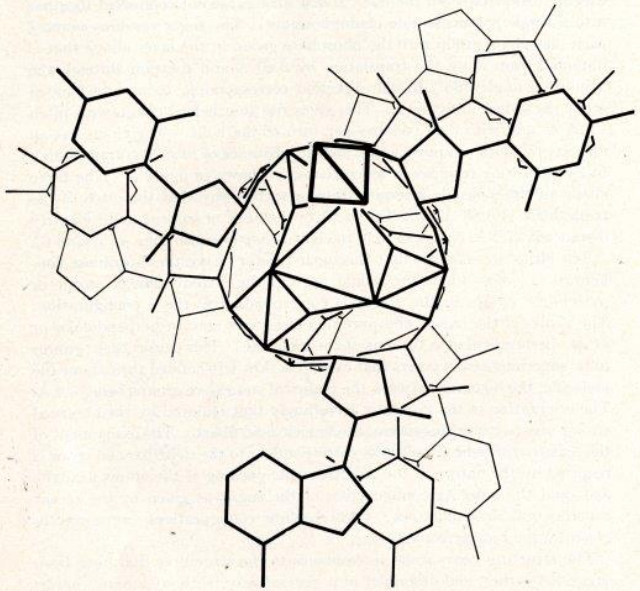


FIGURE 6

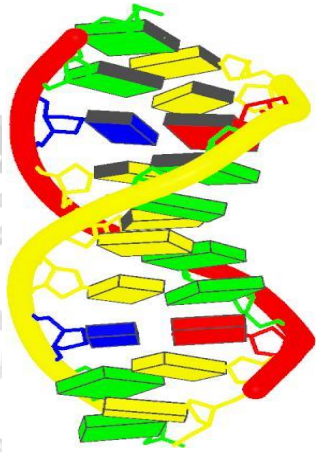
Plan of the nucleic acid structure, showing several nucleotide residues.

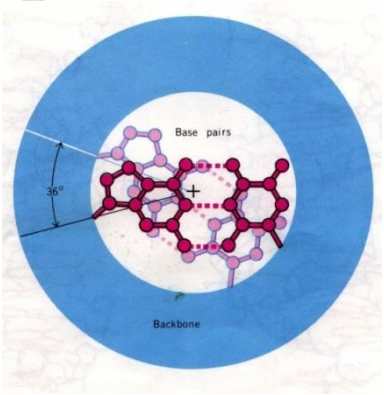
able one, and the assumed distances are those indicated by the observed values for somewhat similar substances, especially the ring compound S_3O_9, in which each sulfur atom is surrounded by a tetrahedron of four oxygen atoms, two of which are shared with adjacent tetrahedra, and two unshared. The O—O distances within the phosphate tetrahedron are 2.32 Å (between the two inner oxygen atoms), 2.46 Å, 2.55 Å, and 2.60 Å. The

Various types of double helix

B-DNA

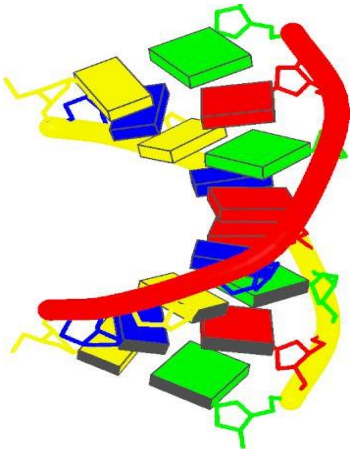
- DNA in water/salt soulutions

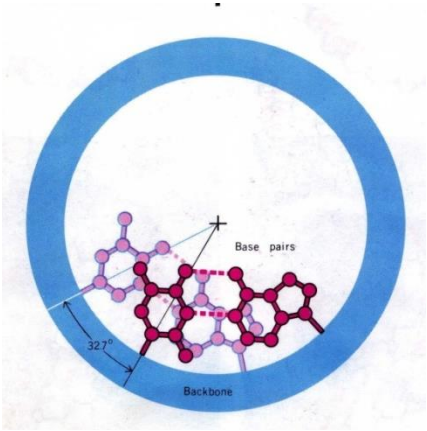




A-DNA

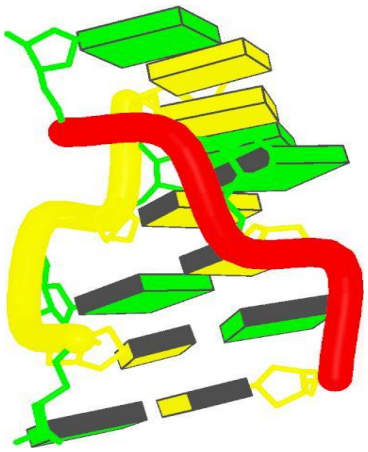
- DNA in crowding solutions
- RNA





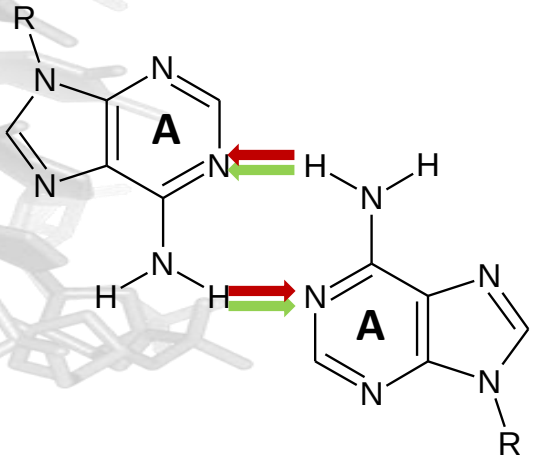
Z-DNA

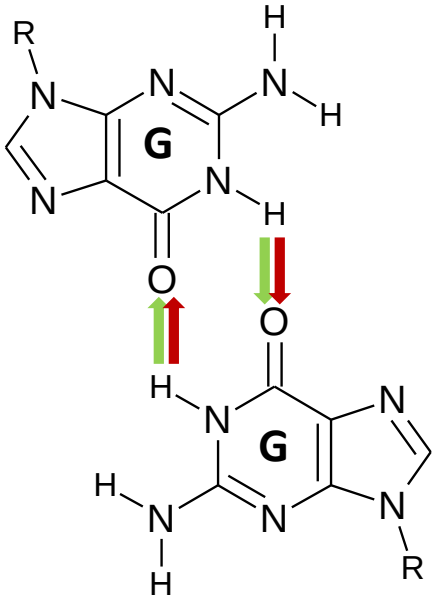
- CpG sequances in crowding conditions

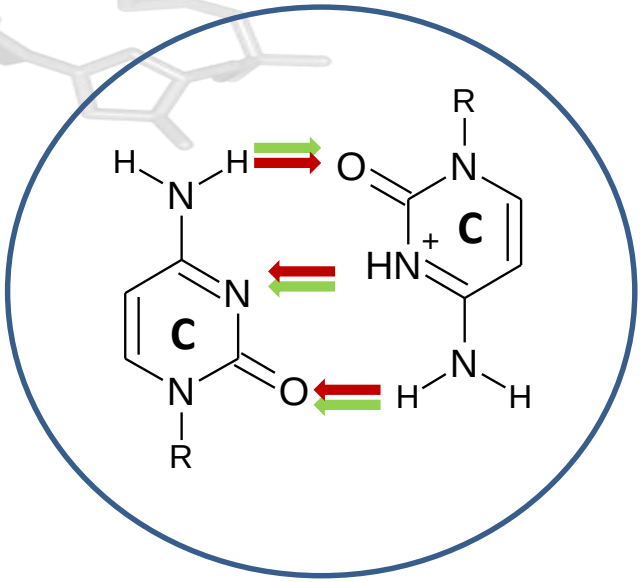


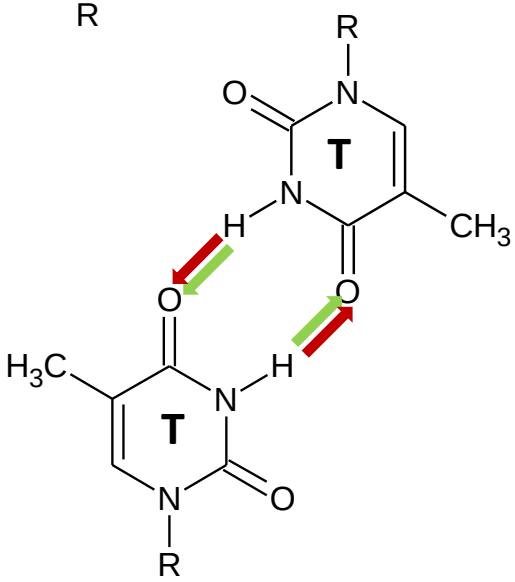
Left handed
Zig-zag step

Reversed Watson-Crick pairing

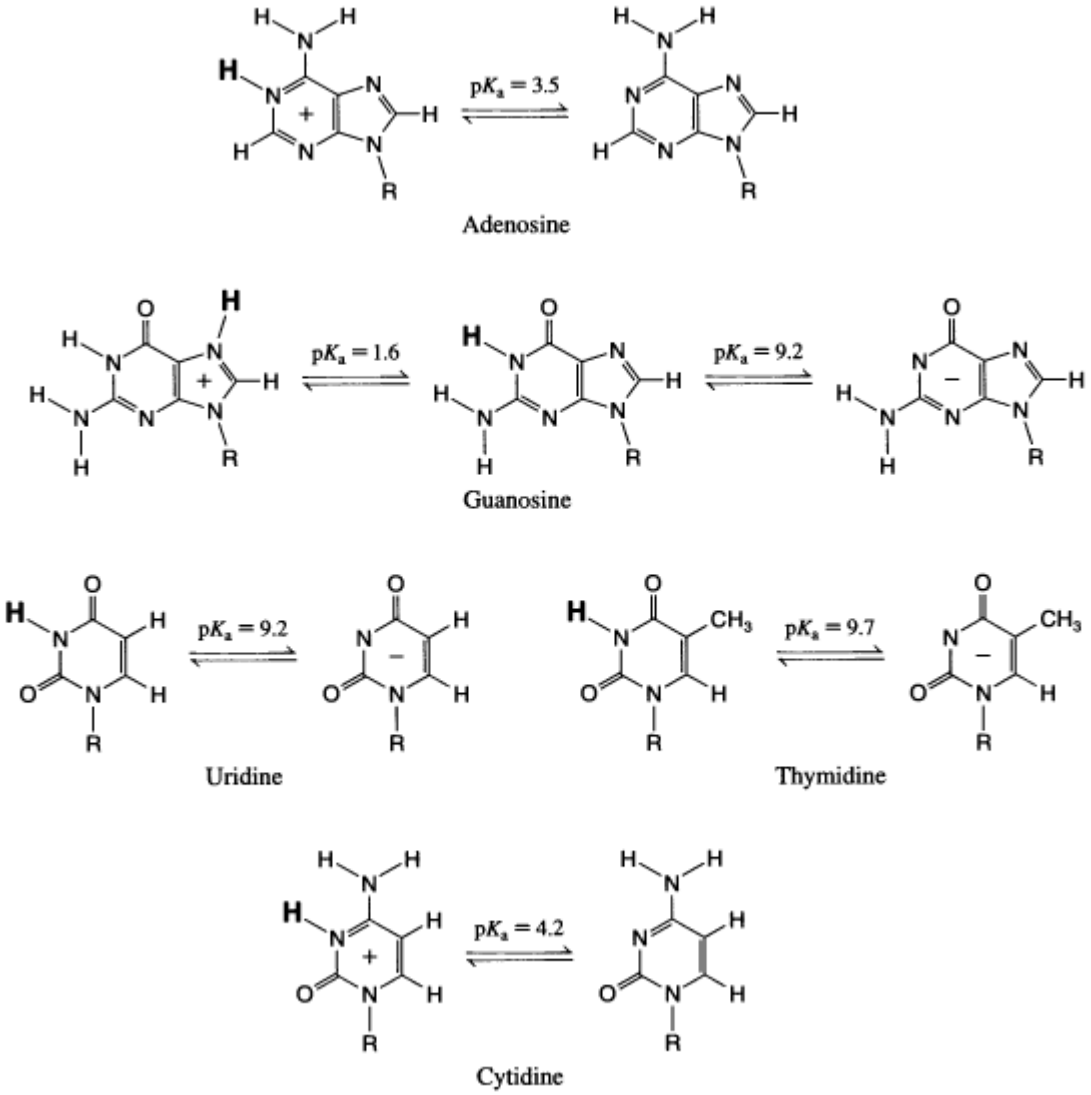








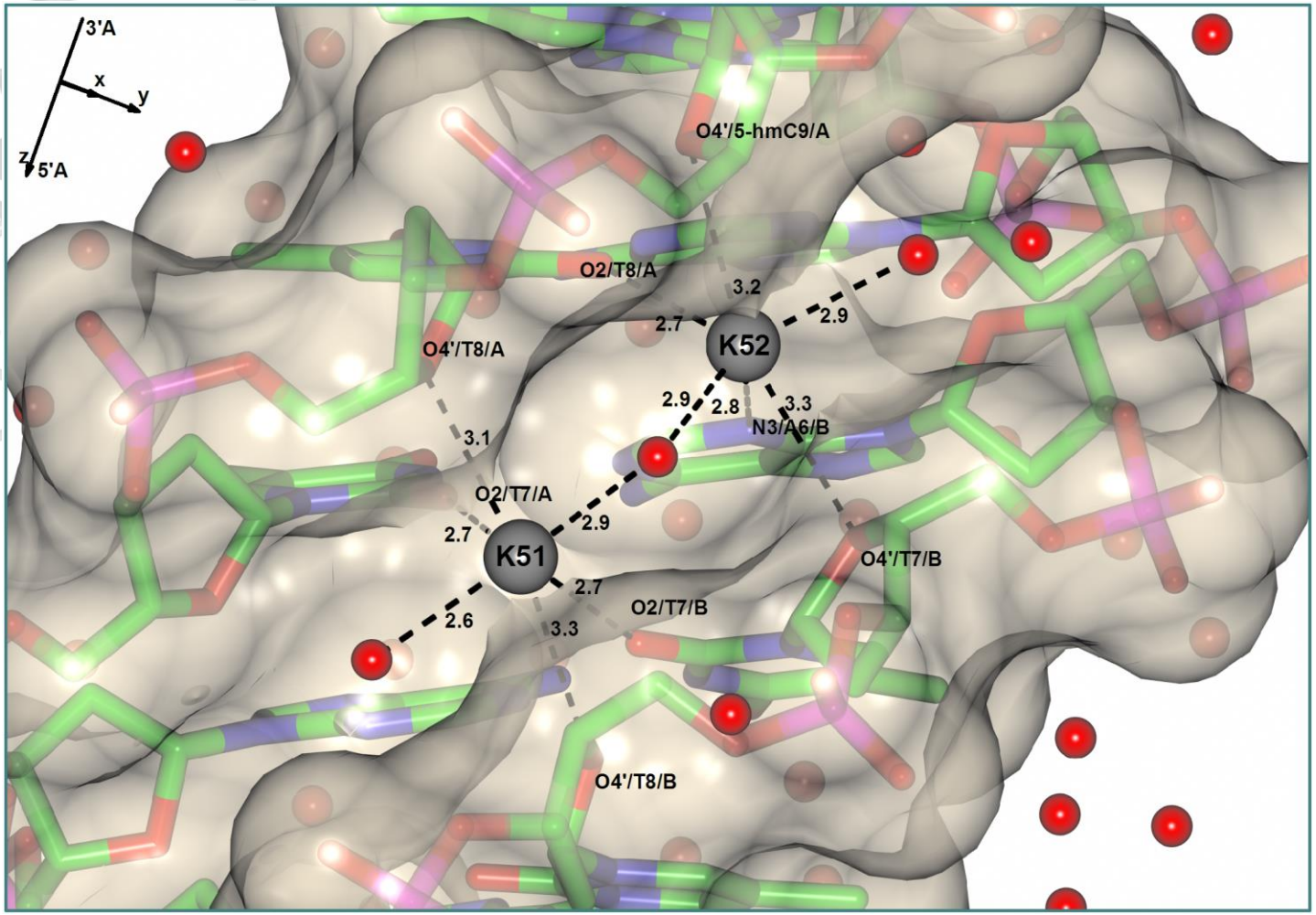
Base protonation



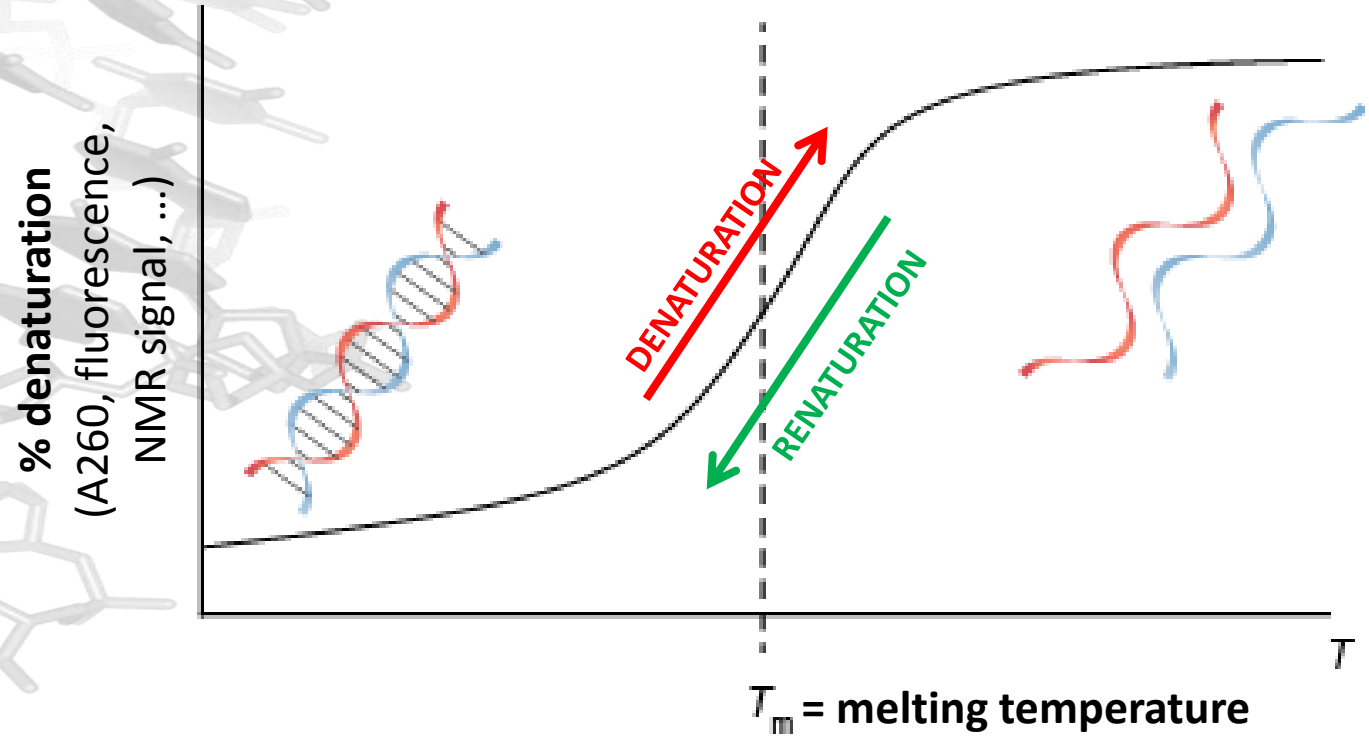
- base protonation might alter the base reactivity
- free bases have pK far from physiological
- pK of bases in DNA might be closer to pH 7.4
- cytosine in C_n sequences has pK~7 – cytosine i-motif

DNA double helix x ions / water

- phosphates in DNA backbone are negatively charged – repulsion
- this is compensated by interaction with ions (Na^+, K^+, Mg^{2+}, …) or water (H-H bonds)



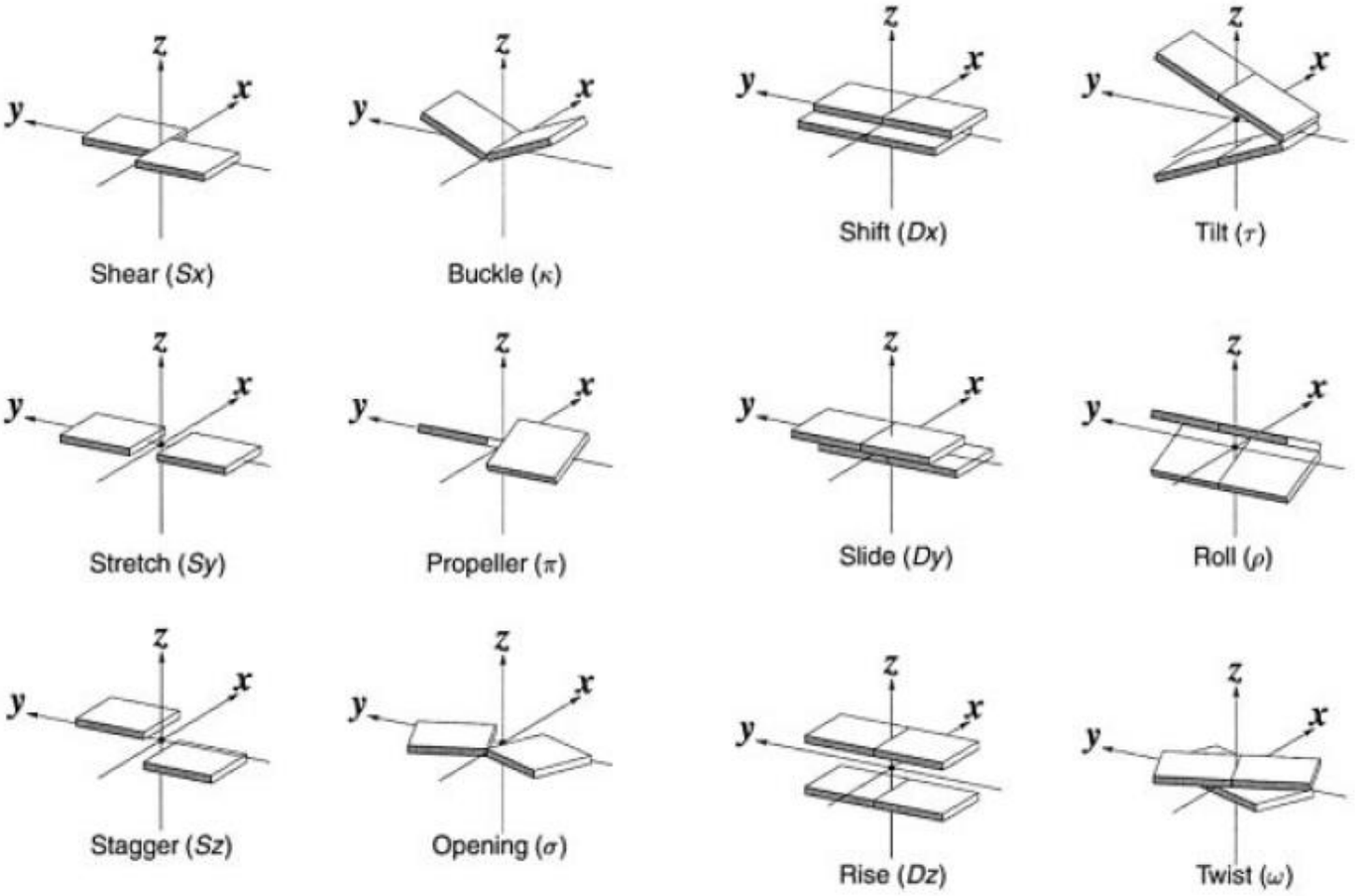
Stability of DNA double helix

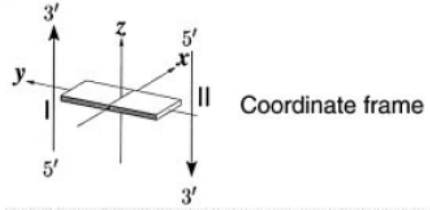


- hydrogen bonds — AT = 2 x GC = 3
- base stacking — various
- repulsion of backbone phosphates — Mg^{2+}>Na^{+}

Tm increases with GC and length
Tm increases with length and ions
Tm increases with ions

Base-pair parameters in double helix



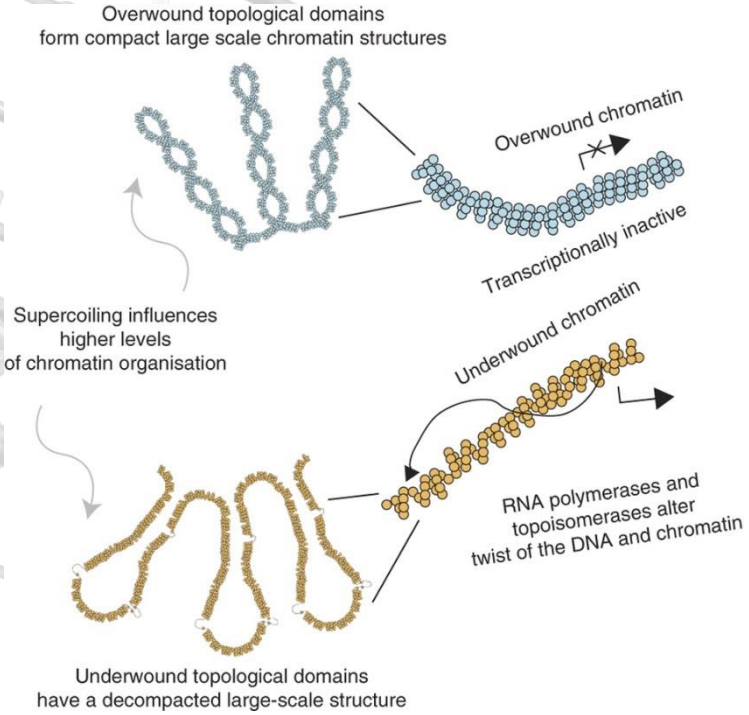


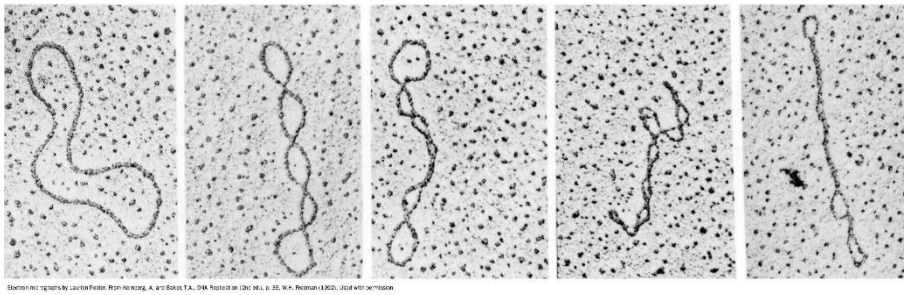
Lu et al., 2003, NAR

Types of nucleic acids

- **linear** (human chromosome) **x circular** (bacterial genome)
- **single-stranded** (most RNAs) **x double-stranded** (human DNA)

Superhelicity





Superhelicity happen mostly as a result of transition of polymerase complex and unwinding of DNA (helicase, …) during replication and transcription.

Topoisomerases

- Enzymes that relax the superhelicity
- Topo I – works on 1 DNA strand
- Topo II – works on 2-strand DNA

Reactivity of bases with amino acids

Double-stranded NA:
Interaction of Hoogsteen side
with amino acid in major groove.

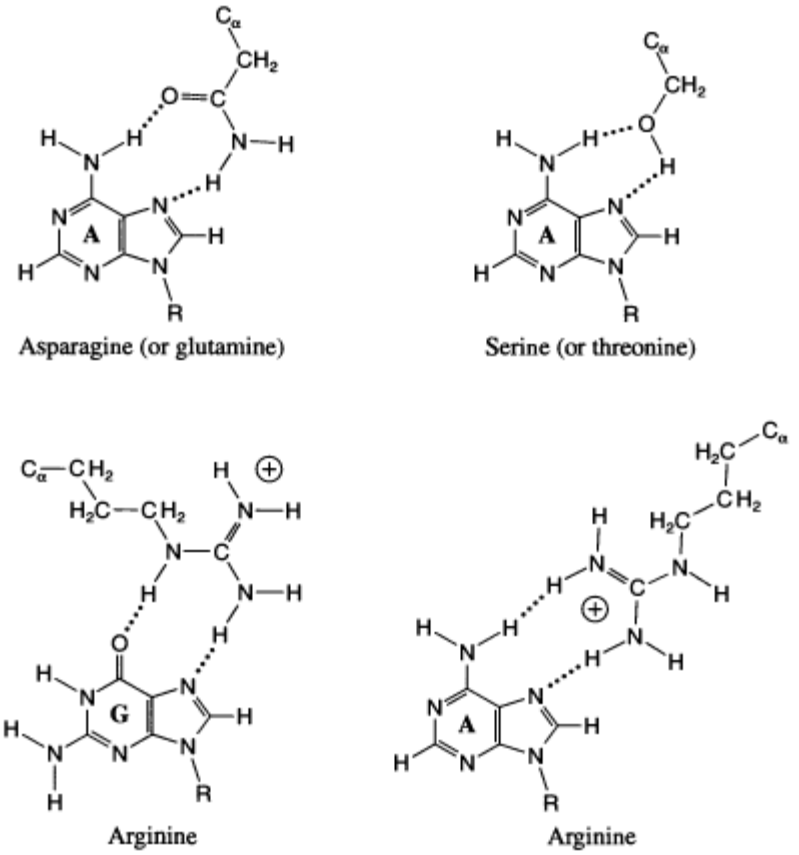


Figure 2-16
Interactions involving two hydrogen bonds between amino acids and bases that can occur through the major groove of a double helix.

Reactivity of bases with amino acids

Single-stranded NA:
Interaction of Watson-Crick side with amino acid.

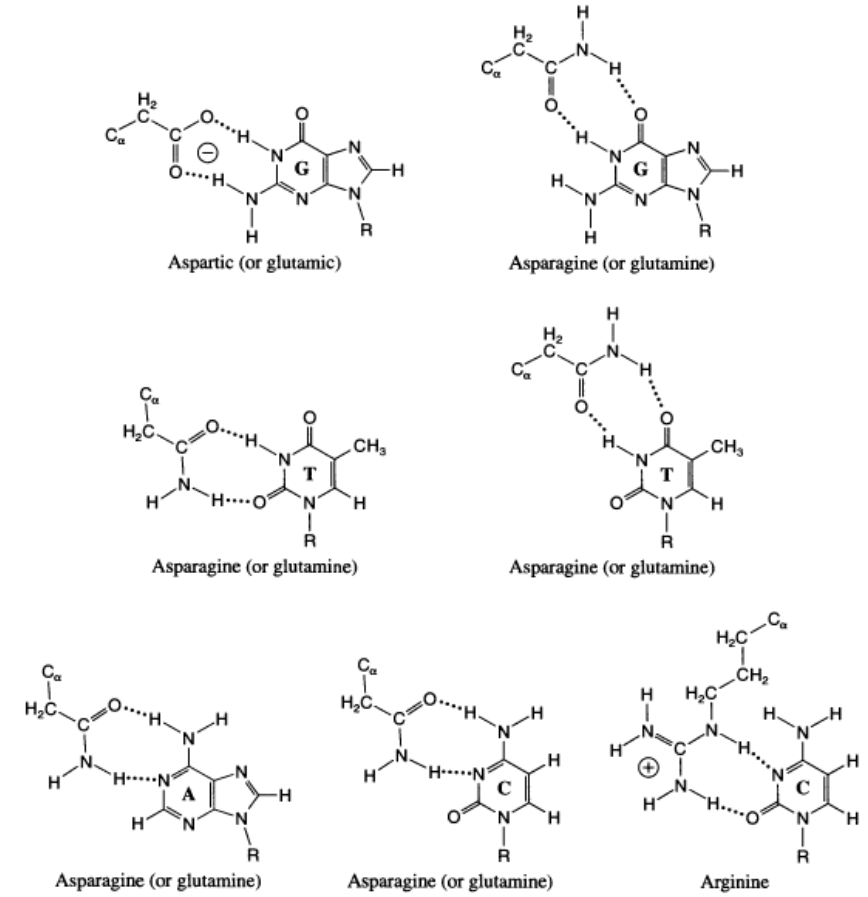
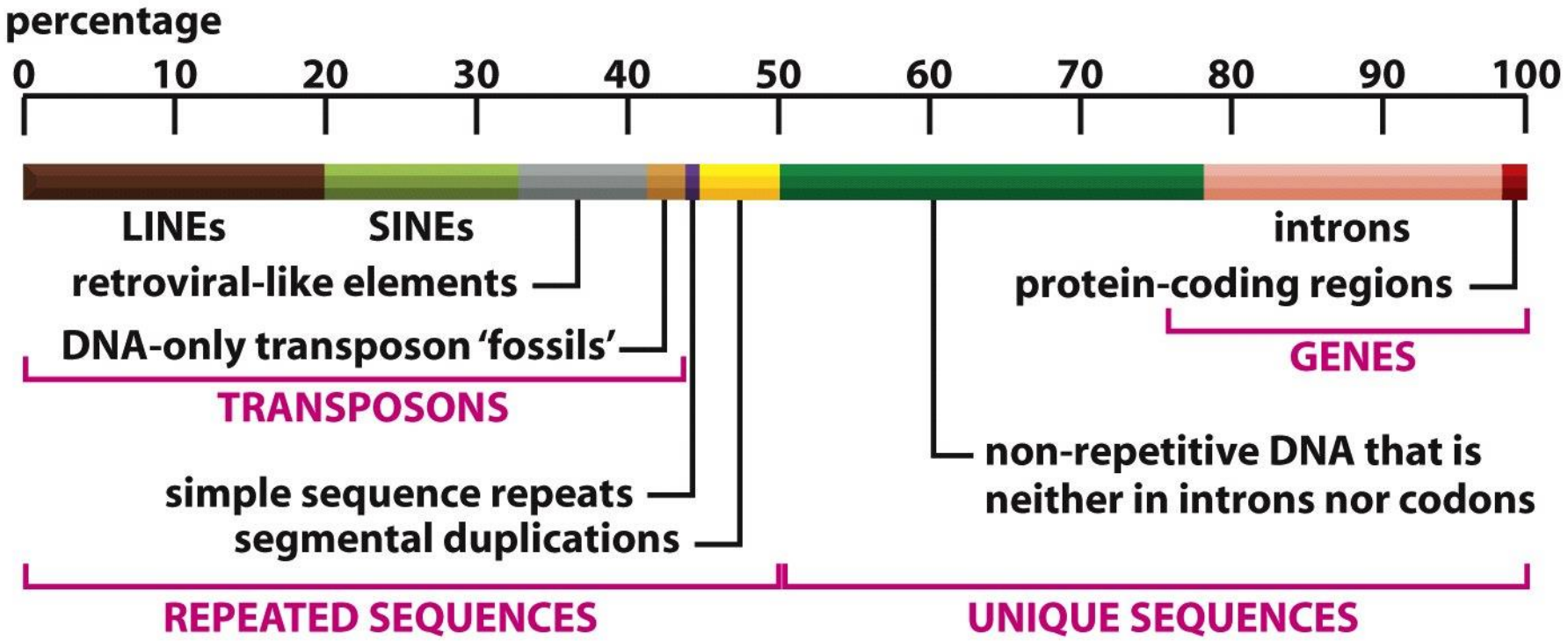


Figure 2-17
Interactions involving two hydrogen bonds between amino acids and bases that take the place of Watson–Crick base pairing.

Genome composition



Repetitive sequences - repeats

Some sequences in genome are **unique**, usually the genomic sequences (both coding and non-coding). In contrast, other sequences exist in many copies – **repetitive sequences (repeats)**. The length of repeat (microsatelites 2-6 bp x LINE 6-7000 bp), as well as the number of copies (several – 1.5M SINE in human) is highly variable.

Structure:

- **direct** repeats

```
5´...AGTC...AGTC...3´
3´...TCAG...TCAG...5´
```

```
5´...AGTC...CTGA...3´
3´...TCAG...GACT...5´
```

- **inverted** repeats + **palindromes**

```
5´...AGTC...GACT...3´
3´...TCAG...CTGA...5´
```

Position:

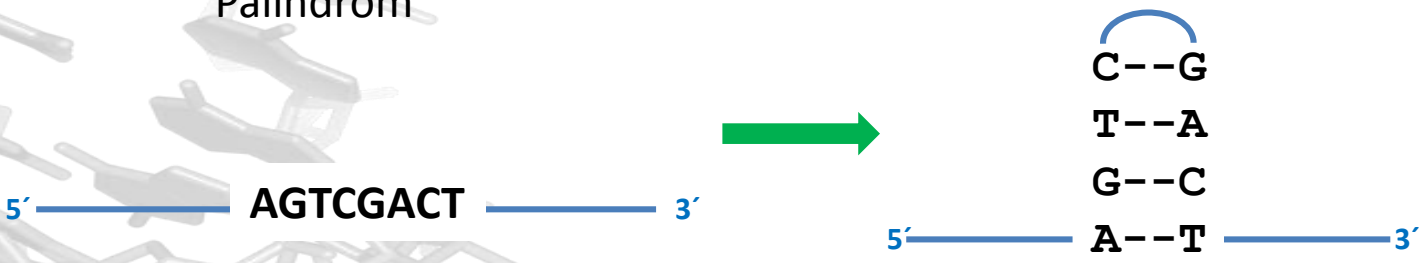
- **Tandem** repeats

- **Interspersed** repeats

Inverted repeats

Hairpin

Palindrom



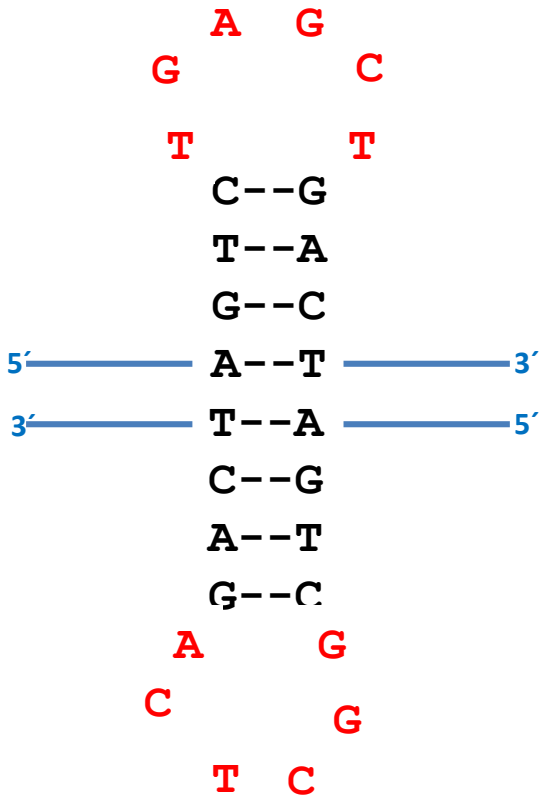
Hairpin with loop

Inverted repeat

Inverted repeats

Inverted repeat

Cruciform



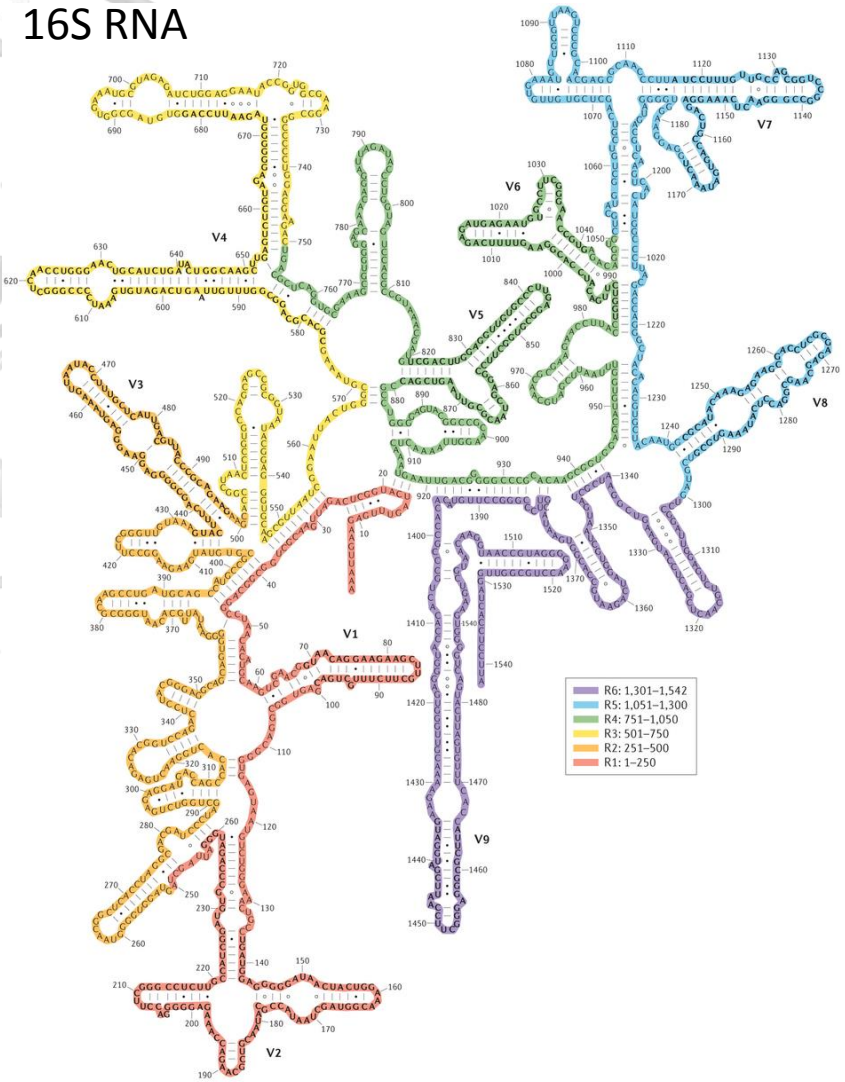
Special types of repetitions - transposons

Interspersed repetitions with various lengths and number of copies.

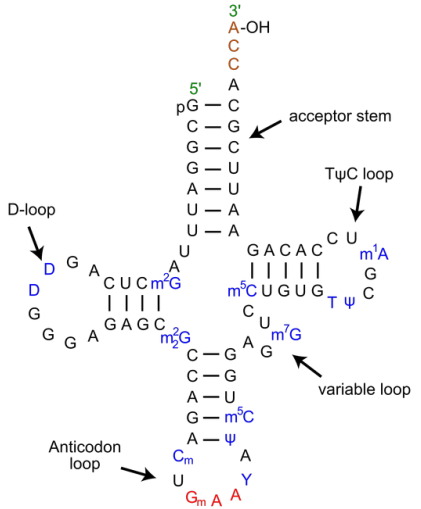
LTR – long terminal repetitions - 100 bp – 5 kbp – variant of retrotransposons
LINE – long interspersed nuclear elements – up to 6 kbp – human> 500k copies
- 3 types (L1, L2, L3) – only some L1 are able to transpose
SINE – (Alu, …) short interspersed nuclear elements – up to 500 bp – human ~ 1,5M copies

Loops and hairpins in RNA

16S RNA

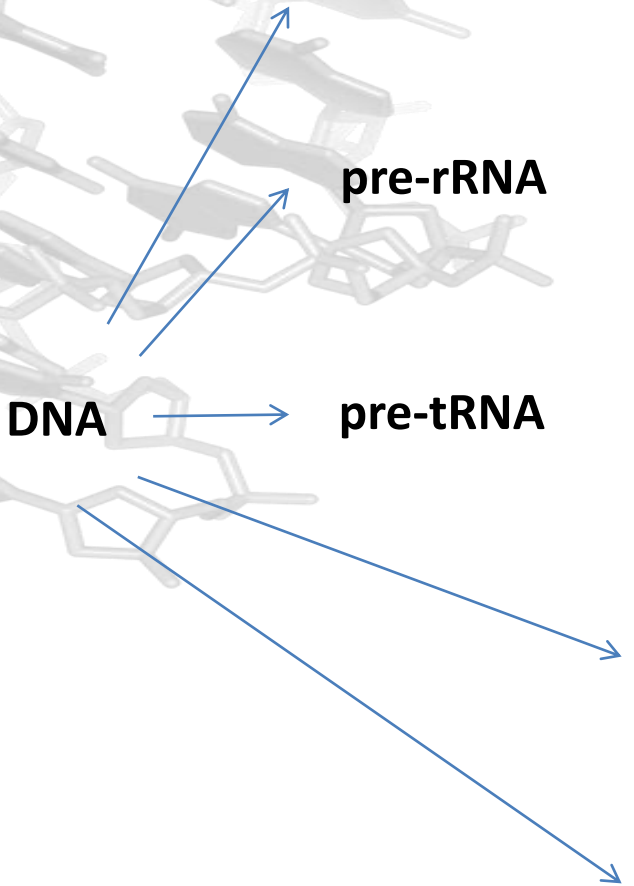


tRNA (Lys)



Nature Reviews | Microbiology

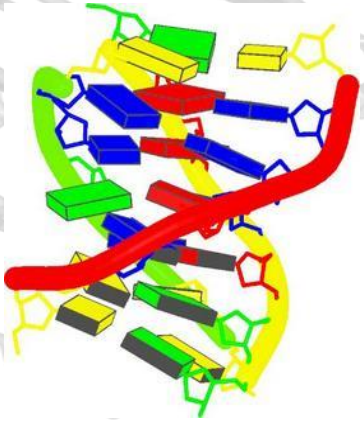
Functional types of RNA



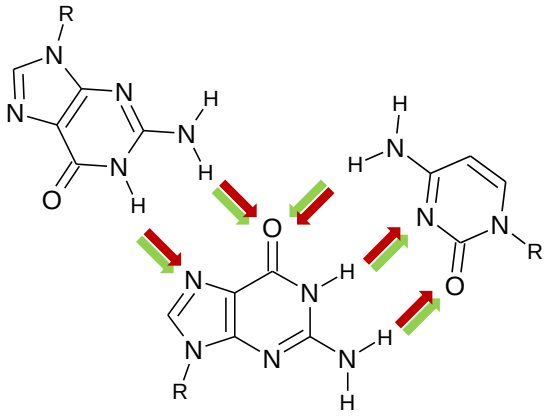
hnRNA	**mRNA**	Translation; codes for protein sequence
pre-rRNA	**rRNA**	Translation; part of ribosome
pre-tRNA	**tRNA**	Translation; amino acid carrier
	snRNA, snoRNA, …	Splicing/modification of RNA
	siRNA, miRNA, …	Gene expression regulation

Hoogsteen pairing - triplexes

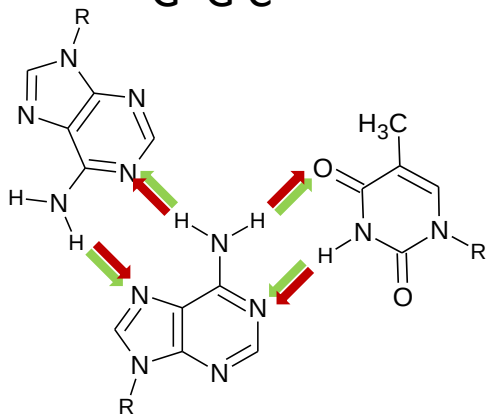
- gene expression regulation
- therapy



Pu*Pu Py

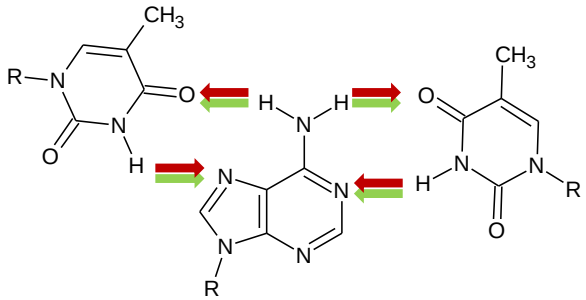


G*G C

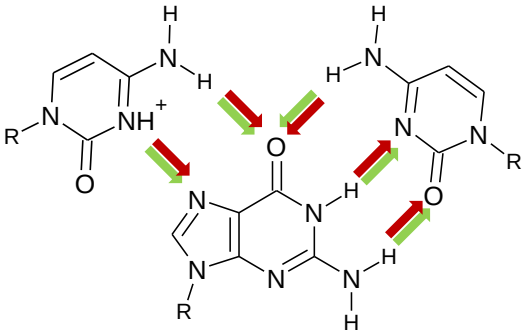


A*A T

Py*Pu Py

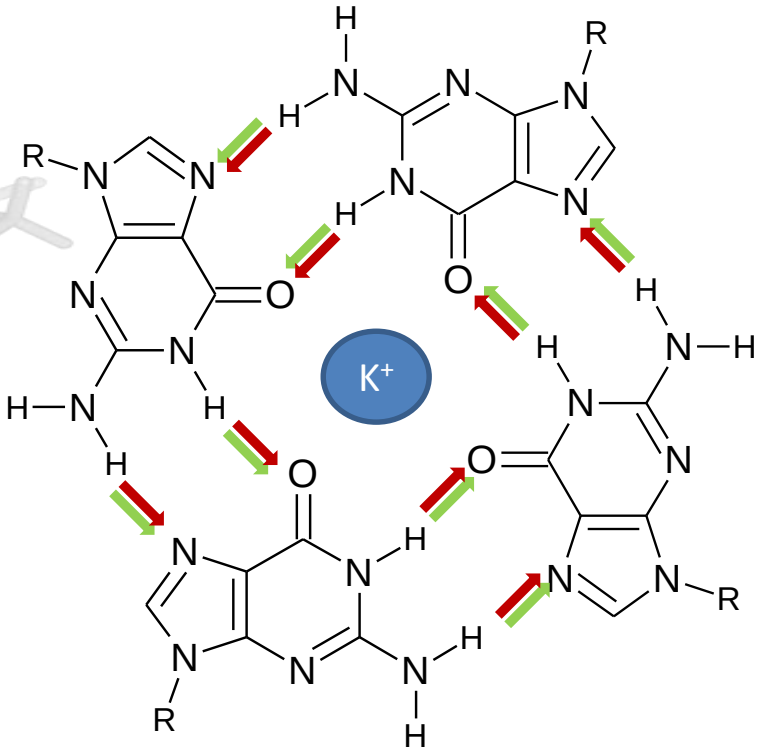


T*A T



C⁺*G C

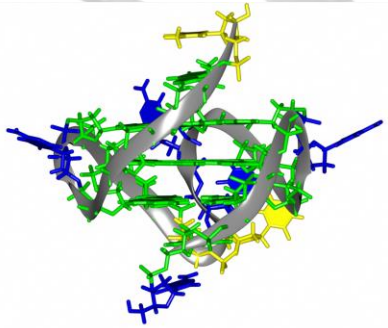
Hoogsteen pairing - quadruplexes

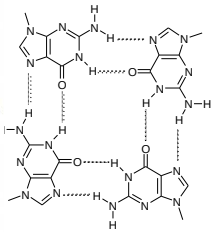


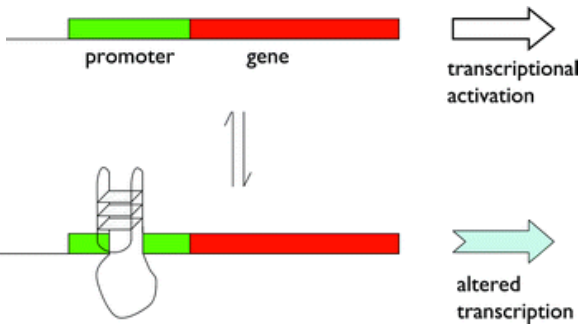
Guanine quadruplexes

$GGGN_nGGGN_nGGGN_nGGG$

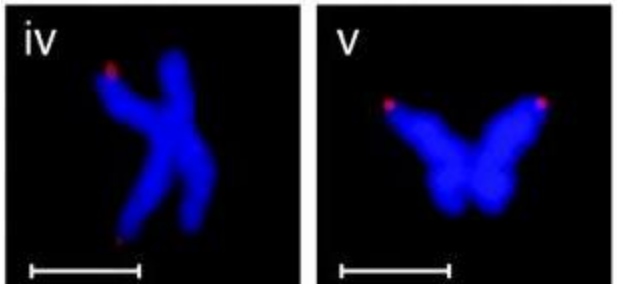
- gene expression regulation
- telomere structure



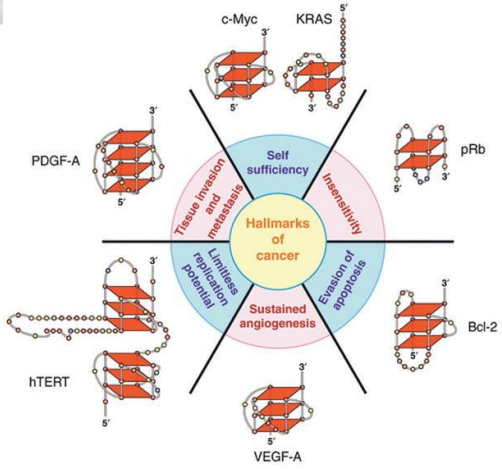




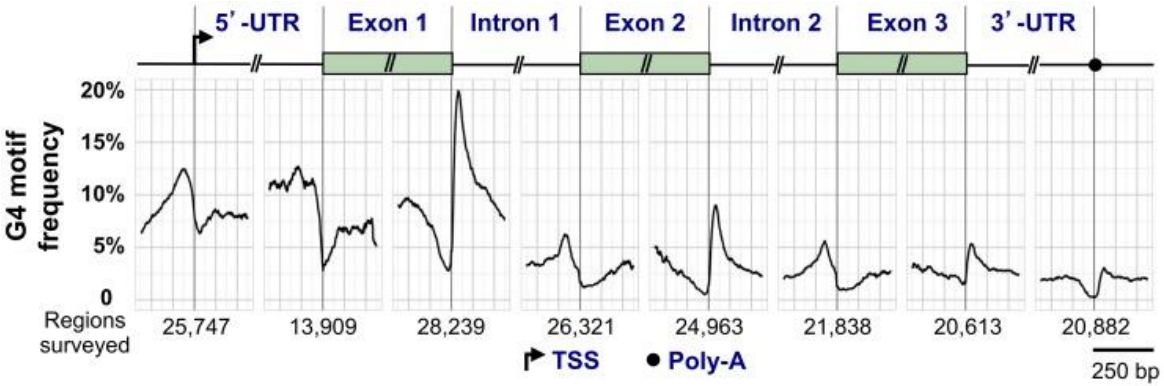
(Huppert J.L., Chem Soc Rev, 2008)



(Biffi G., et al., Nat Chem, 2013)



(Brooks T. A., et al., FEBS J, 2010)

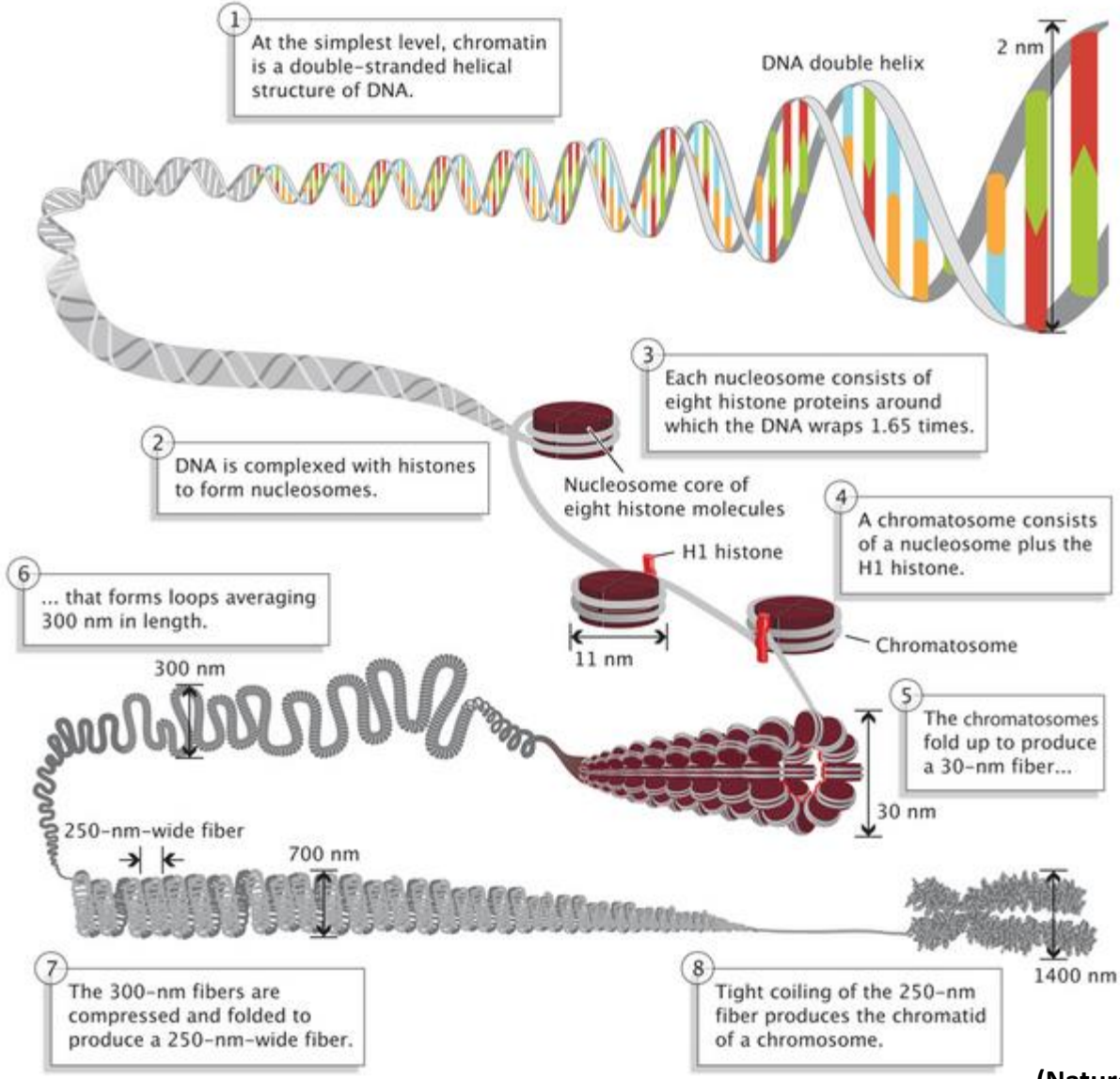


(Maizels N. and Gray L.T., PloS Genet., 2013)

Base reactivity

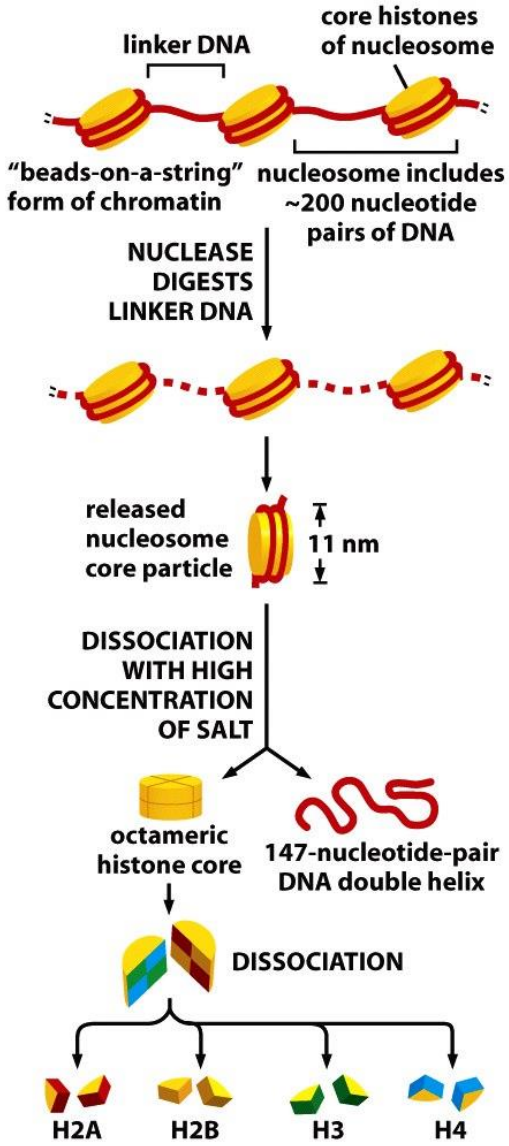
Hydrofobic bases with high ability to form hydrogen bonds are reluctant to be freely expressed into water environment around – if there is any chance to avoid this and lower the base exposition to the environment by any type of base pairing or base stacking, the bases tend to form a structure. Even the “single-stranded” RNA or DNA forms, in fact, compact structure with number of base pairs.

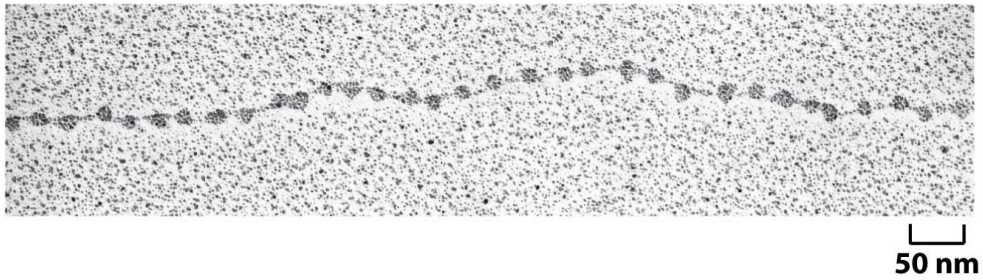
Packing of DNA into chromosome

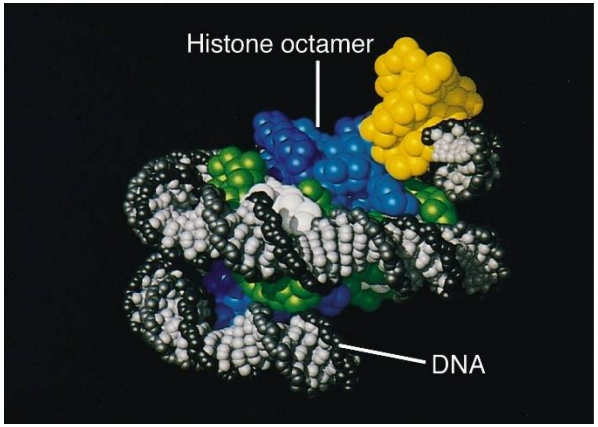


(Nature Education, 2013)

Binding of DNA to a histone octamer

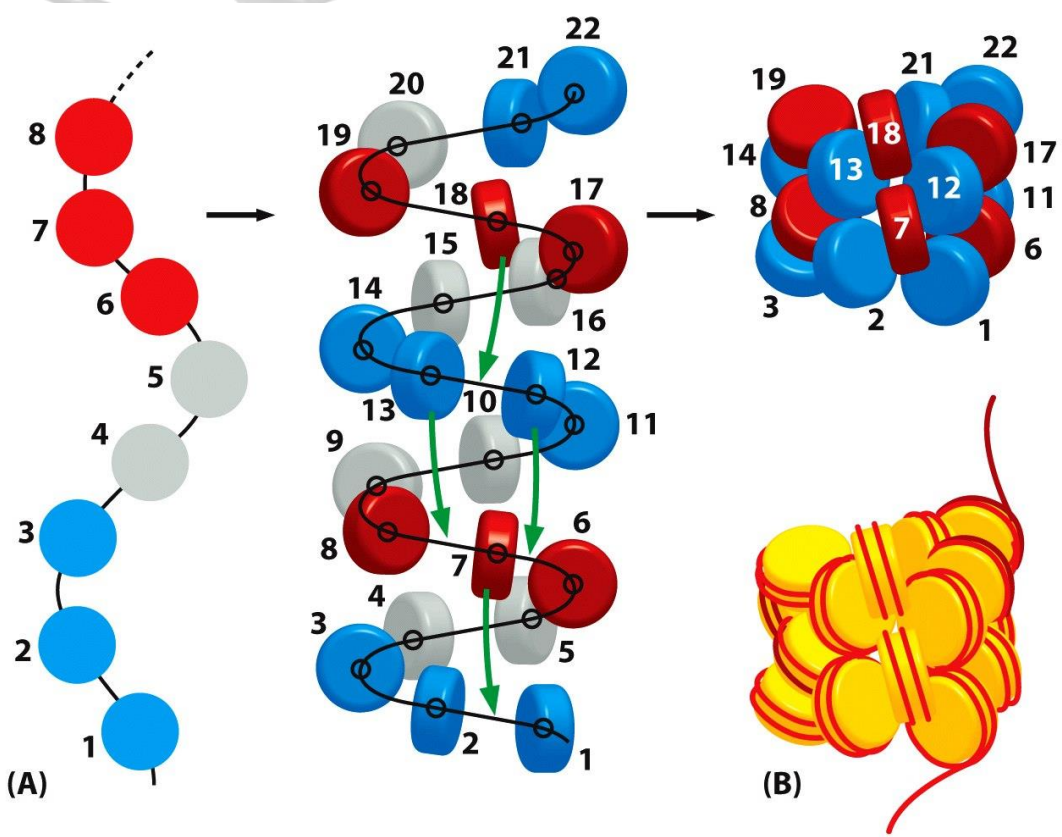


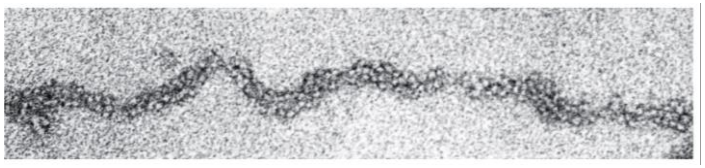


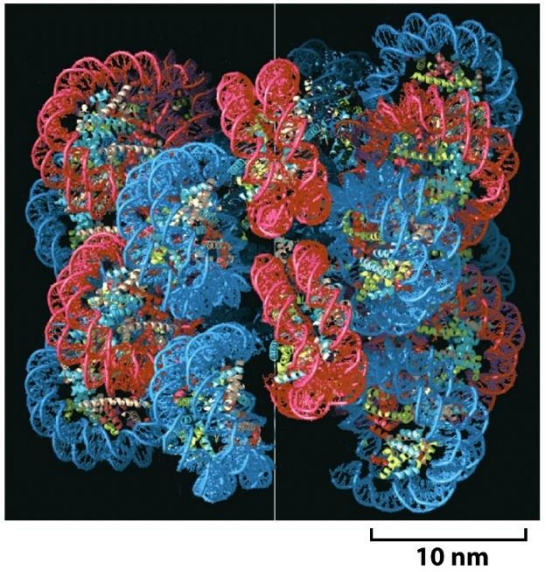


Molecular Biology of the Cell (© Garland Science 2008)

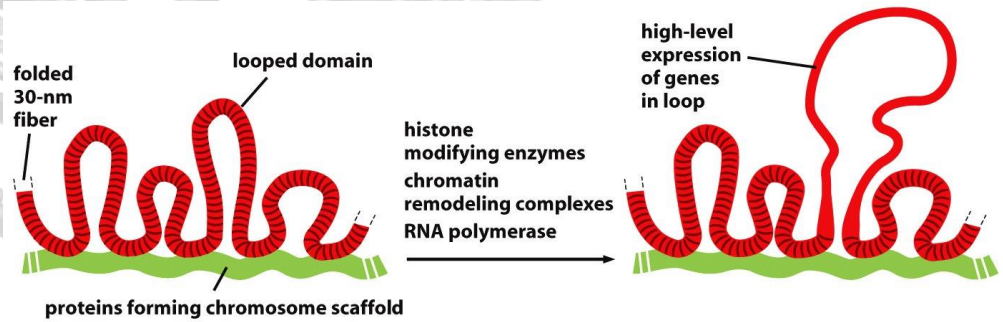
Folding of nucleosomes into 30 nm fiber

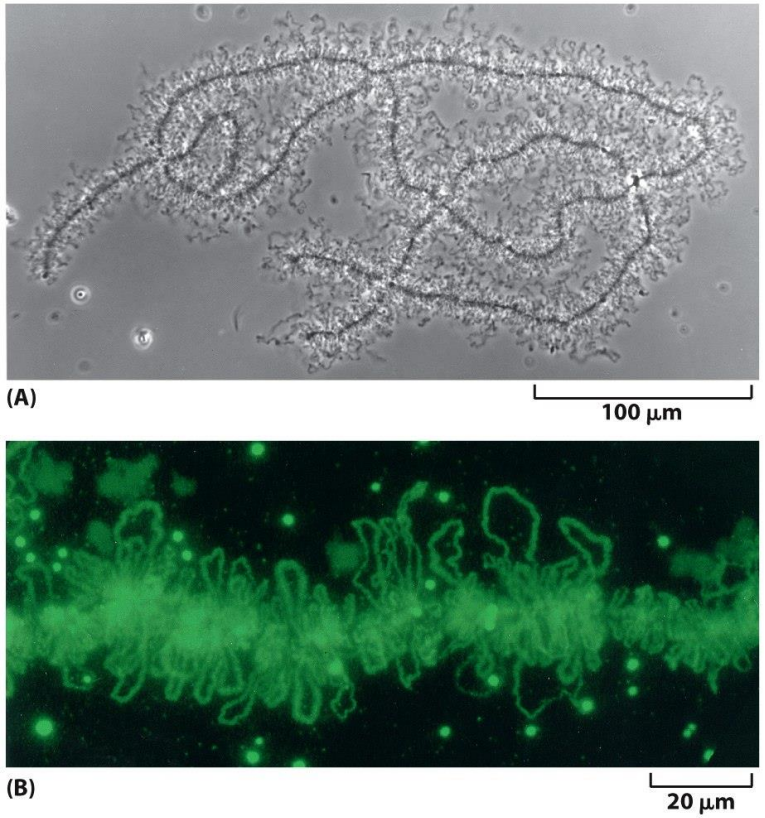




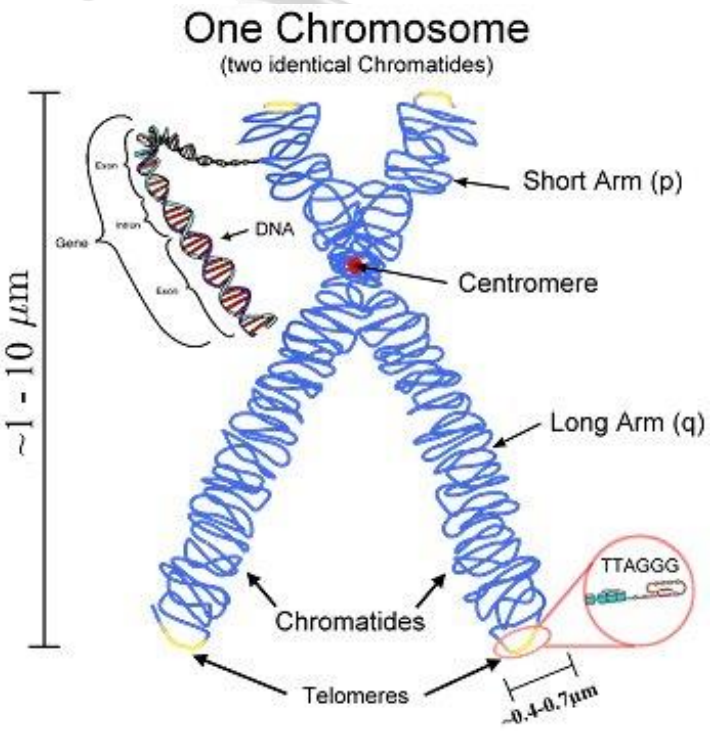


30 nm fiber binds to protein scaffold



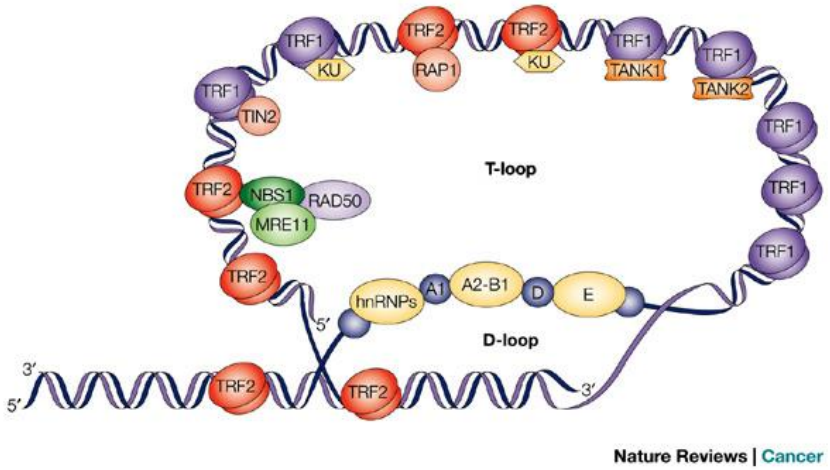


Chromosome



Centromere – here are the chromosomes connected to the system of cellular microtubules – important for chromosome segregation during cell division

Telomere – terminal part of chromatides that protect the end from being recognised as a double-strand break by a DNA repair machinery



Chromosome

Fully condensed chromosomes are present only during the cell division, otherwise they are more or less decondensed to a lower levels of structure, especially in transcriptionally active sites (**euchromatin**). Transcriptionally inactive parts of DNA, as well as repetitive regions or telomeres are much more condensed (**heterochromatin)**. Various types of chromatin differ in **epigenetic** markers of both DNA (5-metyl cytosine) and histones (methylation a acetylation).

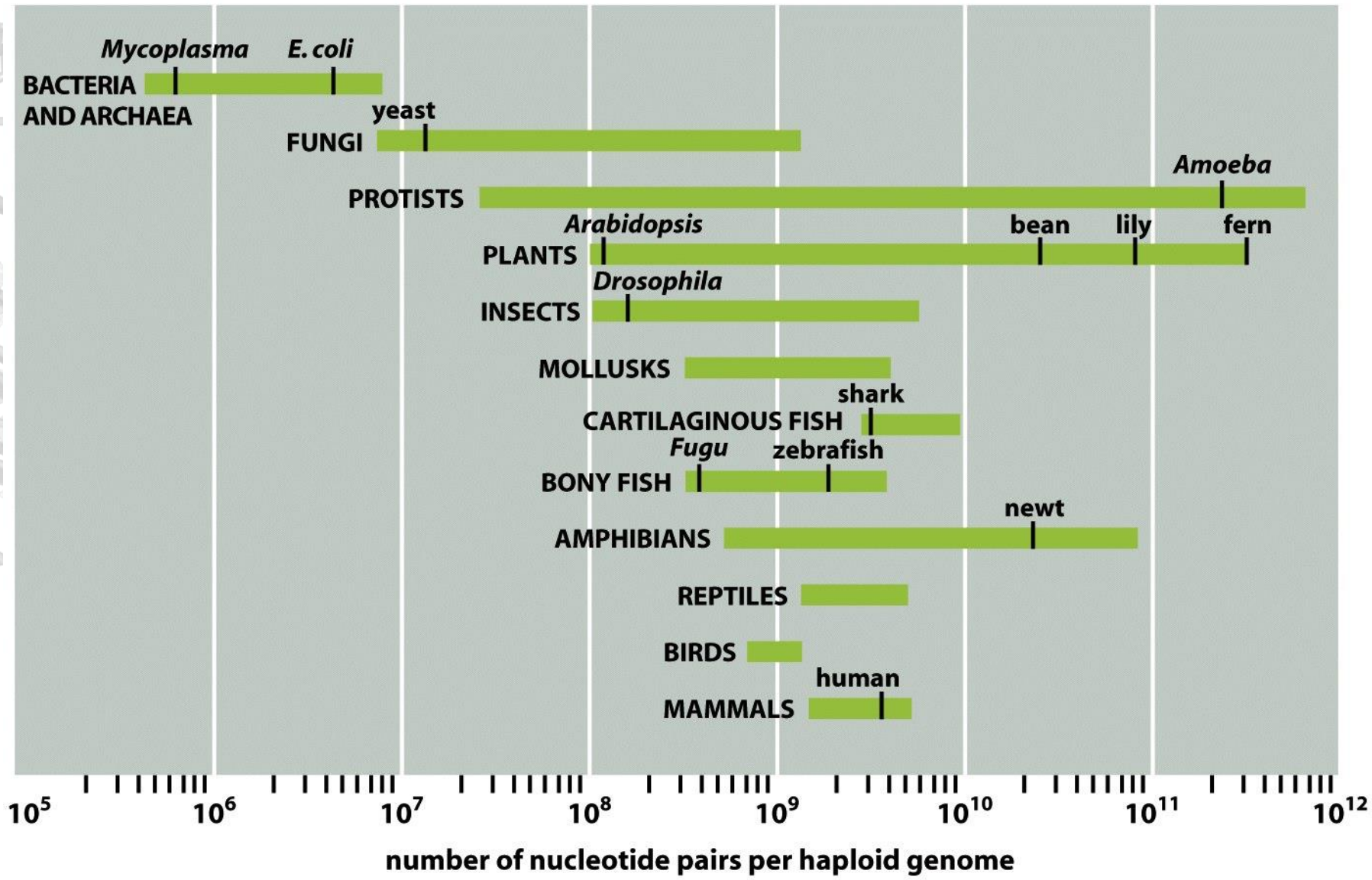


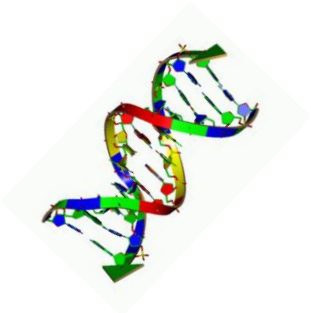
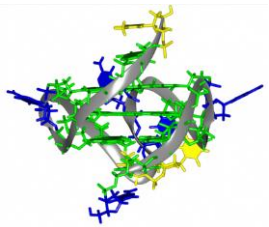
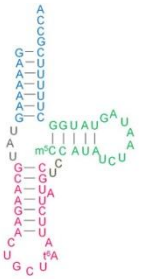
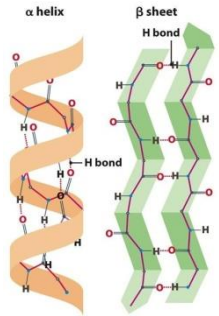
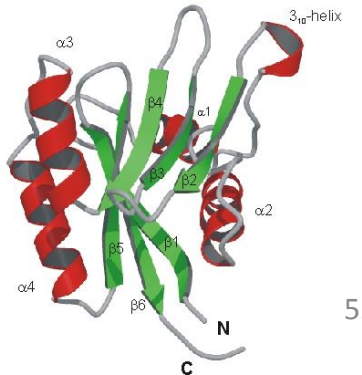
Figure 1-37 *Molecular Biology of the Cell*, Fifth Edition (© Garland Science 2008)

Table 1–1 Some Genomes That Have Been Completely Sequenced

SPECIES	SPECIAL FEATURES	HABITAT	GENOME SIZE (1000s OF NUCLEOTIDE PAIRS PER HAPLOID GENOME)	ESTIMATED NUMBER OF GENES CODING FOR PROTEINS
ARCHAEA				
Methanococcus jannaschii	lithotrophic, anaerobic, methane-producing	hydrothermal vents	1664	1750
Archaeoglobus fulgidus	lithotrophic or organotrophic, anaerobic, sulfate-reducing	hydrothermal vents	2178	2493
Nanoarchaeum equitans	smallest known archaean; anaerobic; parasitic on another, larger archaean	hydrothermal and volcanic hot vents	491	552
EUCARYOTES				
Saccharomyces cerevisiae (budding yeast)	minimal model eucaryote	grape skins, beer	12,069	~6300
Arabidopsis thaliana (Thale cress)	model organism for flowering plants	soil and air	~142,000	~26,000
Caenorhabditis elegans (nematode worm)	simple animal with perfectly predictable development	soil	~97,000	~20,000
Drosophila melanogaster (fruit fly)	key to the genetics of animal development	rotting fruit	~137,000	~14,000
Homo sapiens (human)	most intensively studied mammal	houses	~3,200,000	~24,000

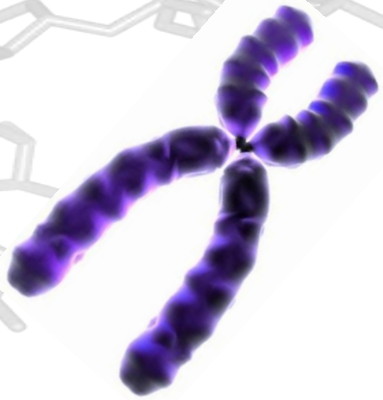
Genome size and gene number vary between strains of a single species, especially for bacteria and archaea. The table shows data for particular strains that have been sequenced. For eucaryotes, many genes can give rise to several alternative variant proteins, so that the total number of proteins specified by the genome is substantially greater than the number of genes.

Levels of structure of biopolymers

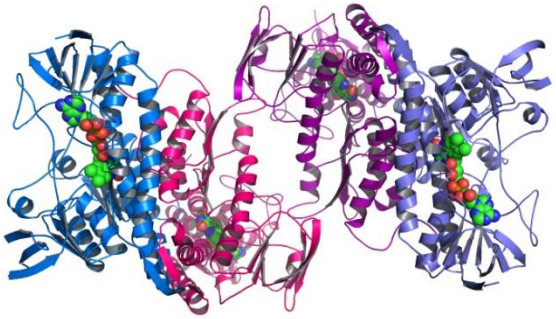
	DNA	RNA	Protein
Primary	AGGCTGCAAGTCGAT	UCCGACGUUCAGCUA	Ser-Asp-Val-Gln-Leu
Secondary			 
Tertiary			

Genetic code

Set of rules that assign a sequence of aminoacids in the protein to the sequence of nucleotides in DNA or RNA.



Transcription
+
Translation



M.W.Nirenberg

RNA CODEWORDS AND PROTEIN SYNTHESIS, III.
ON THE NUCLEOTIDE SEQUENCE OF A CYSTEINE
AND A LEUCINE RNA CODEWORD

By Philip Leder and Marshall W. Nirenberg

National Heart Institute, National Institutes of Health

Communicated by Richard B. Roberts, October 1, 1964

Previous studies utilizing randomly ordered synthetic polynucleotides to direct amino acid incorporation into protein in *E. coli* extracts indicated that RNA codewords corresponding to valine, leucine, and cysteine contain the bases (UUG).[1–4] The activity of chemically defined trinucleotides in stimulating the binding of a specific C^{14}-aminoacyl-sRNA to ribosomes, prior to peptide bond formation,[5] provided a means of investigating base sequence of RNA codewords and showed that the sequence of a valine RNA codeword is GpUpU.[6]

Properties of genetic code

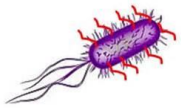
• genetic code is based on **triplets** – one aminoacid in protein is coded by a sequence of three nucleotides in DNA (RNA)

mRNA CGUGGUACGAUUGGAUGU

Protein **Arg Gly Thr Ile Gly Cys**

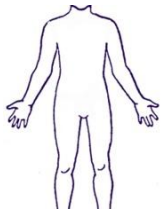
Triplet = Codon x anticodon = complementary sequence on particular tRNA that carries the respective aminoacid

• genetic code is **universal** – individual triplets code for the same aminoacid in almost all organisms (x mitochondria)



CGU = **Arginine**

CGU = **Arginine**



CGU = **Arginine**

• genetic code is **degenerated** – one aminoacid might be code by several different triplets (but the opposite is not true)

Genetic code

	Second nt				
First nt	U	C	A	G	Third nt
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	STOP	STOP/Sel	A
	Leu	Ser	STOP	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met/START	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

Reading frames

Genetic code is based on triplets – three possible ways of reading (**reading frames**), but only one is correct.

mRNA CGUGGUACGAUUGGAUGU

Protein1 **Arg Gly Thr Ile Gly Cys**

mRNA CGUGGUACGAUUGGAUGU

Protein2 **Val Val Arg Leu Asp**

mRNA CGUGGUACGAUUGGAUGU

Protein3 **Trp Tyr Asp Trp Met**

Genetic code

Although the genetic code is universal, the usage of particular codons, as well as the amount of particular tRNAs and aminoacyl transferases differ

Optimization of synthetic genes for recombinant protein production according to the expression system used (Bacteria, human, …) might be highly beneficial.