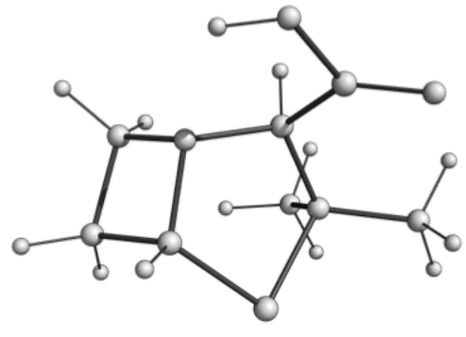
CS 6824: Molecules

Acknowledgement:
Many of the images in the slides are derived from images.google.com or other publicly available sources.

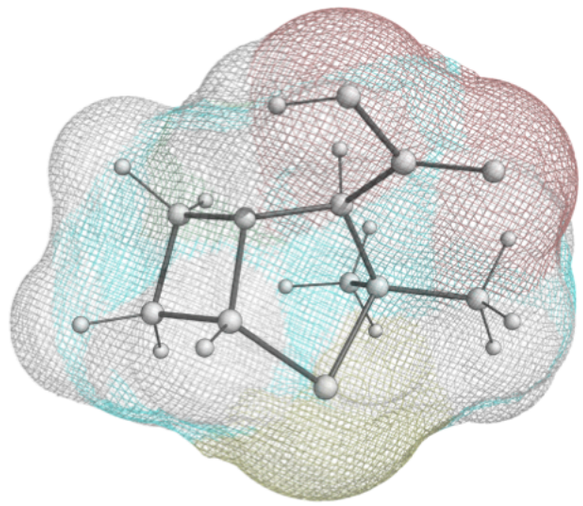
Molecules

Chemical View



A group of atoms bonded together, representing the smallest fundamental unit of a chemical compound that can take part in a chemical reaction.

Geometric View

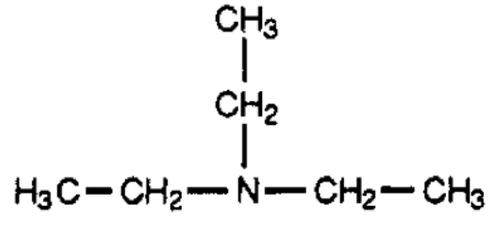


Molecular systems (and three-dimensional representations thereof) can be considered as objects or graph in Euclidean space.

SMILES (Simplified Molecular Input Line Entry System)

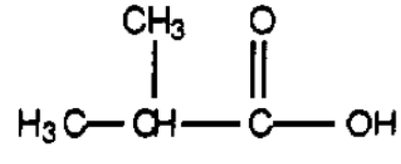
A notation system well suited for high-speed machine processing

- Atoms:
  - Represented by their atomic symbols



CCN(CC)CC

Triethylamine

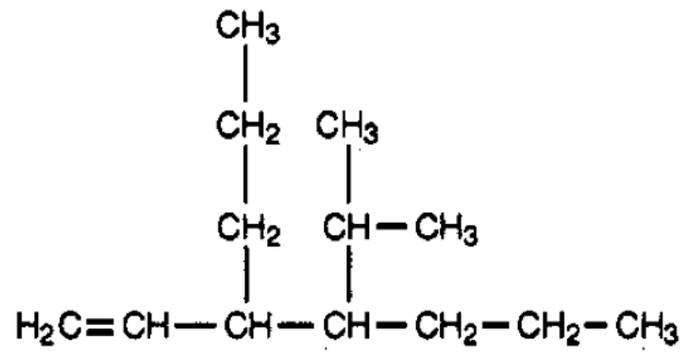


CC(C)C(=O)O

Isobutyric acid

- Bonds
  - Single, double, triple, and aromatic bonds are represented by the symbols —, =, #, and :, respectively.

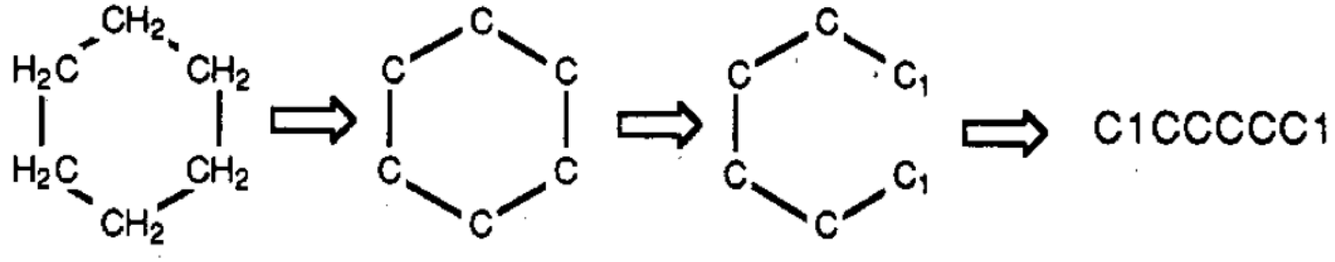
- Branches
  - Branches are specified by enclosures in parentheses
  - Branches can be nested



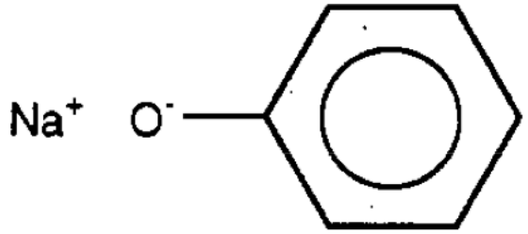
C=CC(CCC)C(C(C)C)CCC

SMILES can represent small molecules

Cyclic structures



Disconnected structures

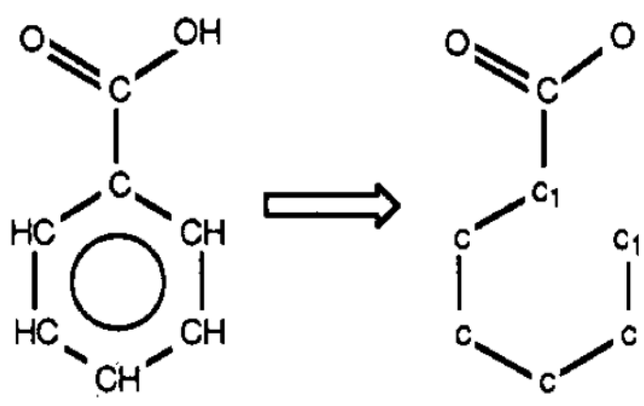


[Na+].[O-]c1ccccc1

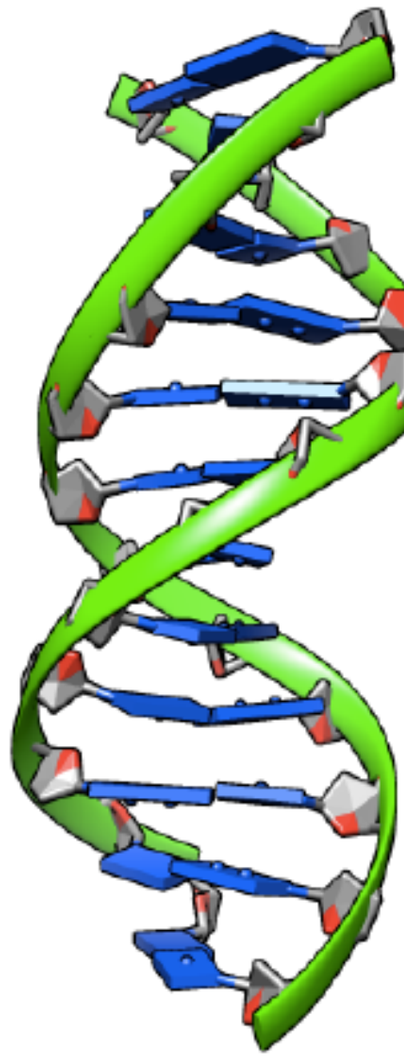
or

c1cc([O-].[Na+])ccc1

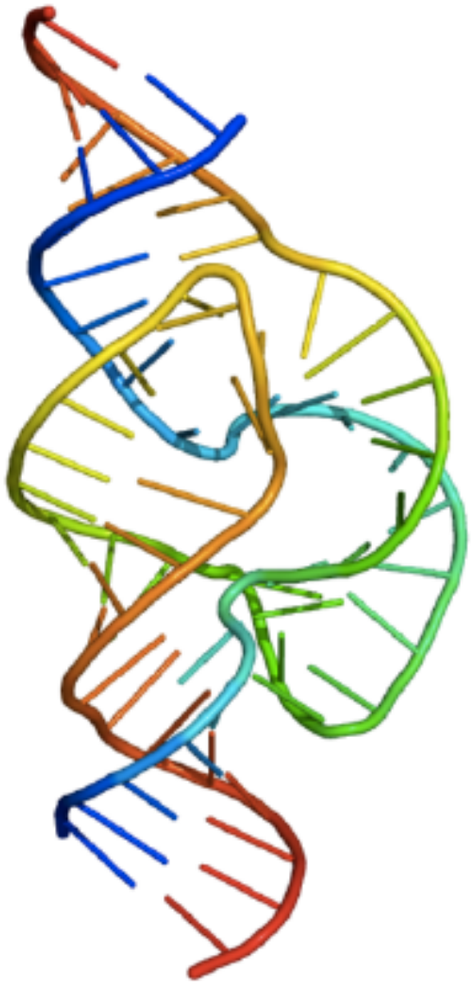
Aromaticity



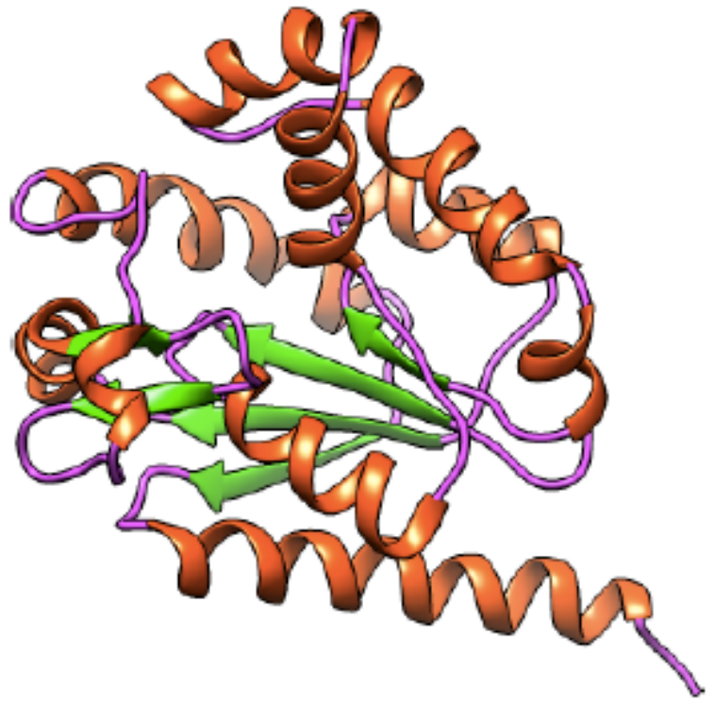
Going beyond small molecules: cellular macromolecules



DNA



RNA



Protein

DNA

The double helix

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] Von Arx, 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.

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

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 easy access to them.

The structure is an open one, and its water content is rather high. At lower water contents we would expect the bases to tilt so that the structure could become more compact.

The novel feature of the structure is the manner in which the two chains are held together by the purine and pyrimidine bases. The planes of the bases are perpendicular to the fibre axis. They are joined together in pairs, a single base from one chain being hydrogen-bonded to a single base from the other chain, so that the two lie side by side with identical z-co-ordinates. One of the pair must be a purine and the other a pyrimidine for bonding to occur. The hydrogen bonds are made as follows: purine position 1 to pyrimidine position 1; purine position 6 to pyrimidine position 6.

If it is assumed that the bases only occur in the structure in the most plausible tautomeric forms (that is, with the keto rather than the enol configurations) it is found that only specific pairs of bases can bond together. These pairs are: adenine (purine) with thymine (pyrimidine), and guanine (purine) with cytosine (pyrimidine).

In other words, if an adenine forms one member of 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[5,6] on deoxyribose 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 stereochemical 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.

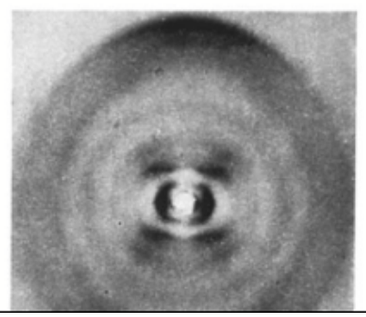
We are much indebted to Dr. Jerry Donohue for constant advice and criticism, especially on inter-atomic distances. We have also been stimulated by a knowledge of the general nature of the unpublished experimental results and ideas of Dr. M. H. F. Wilkins, Dr. R. E. Franklin and their co-workers at

King's College, London. One of us (J. D. W.) has been aided by a fellowship from the National Foundation for Infantile Paralysis.

J. D. WATSON
F. H. C. CRICK

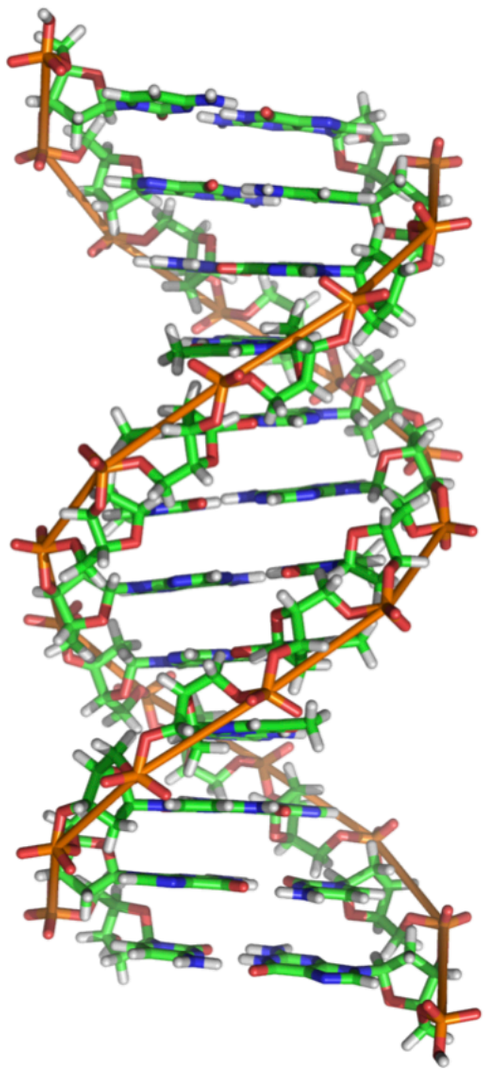
Medical Research Council Unit for the Study of the Molecular Structure of Biological Systems, Cavendish Laboratory, Cambridge. April 2.

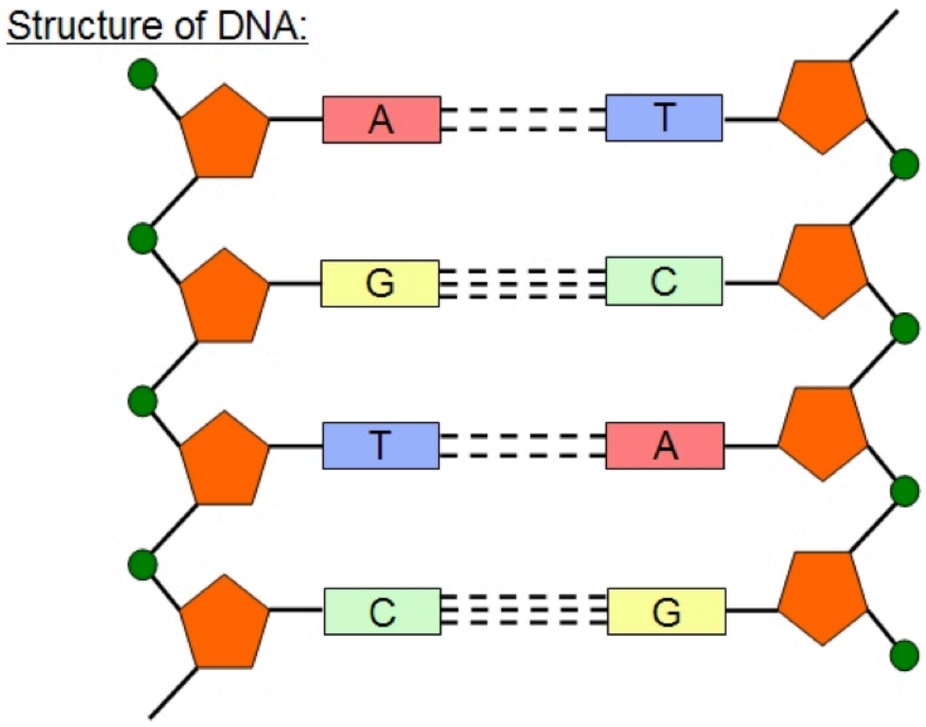
[1] Pauling, L., and Corey, R. B., *Nature*, **171**, 346 (1953); *Proc. U.S. Nat. Acad. Sci.*, **39**, 84 (1953).
[2] Furberg, S., *Acta Chem. Scand.*, **6**, 634 (1952).
[3] Chargaff, E., for references see Zamenhof, S., Brawerman, G., and Chargaff, E., *Biochim. et Biophys. Acta*, **9**, 402 (1952).
[4] Wyatt, G. R., *J. Gen. Physiol.*, **36**, 201 (1952).
[5] Astbury, W. T., Symp. Soc. Exp. Biol. 1, Nucleic Acid, 66 (Camb. Univ. Press, 1947).
[6] Wilkins, M. H. F., and Randall, J. T., *Biochim. et Biophys. Acta*,



DNA

Each strand composed of sequence of covalently bonded nucleotides (bases).





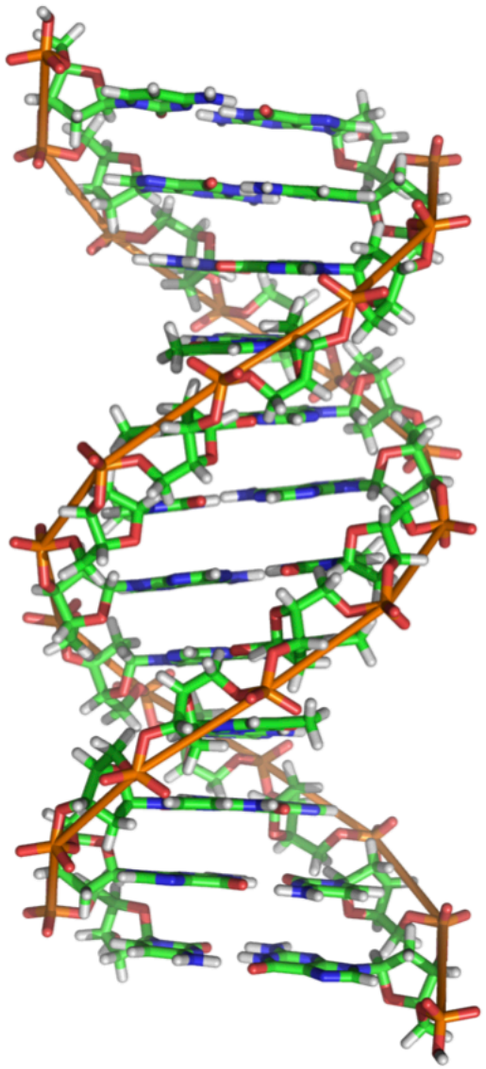
Four nucleotides:

A (adenine)
C (cytosine)
T (thymine)
G (guanine)

A—T, C—G Watson-Crick base-pairing

DNA

Each strand composed of sequence of covalently bonded nucleotides (bases).



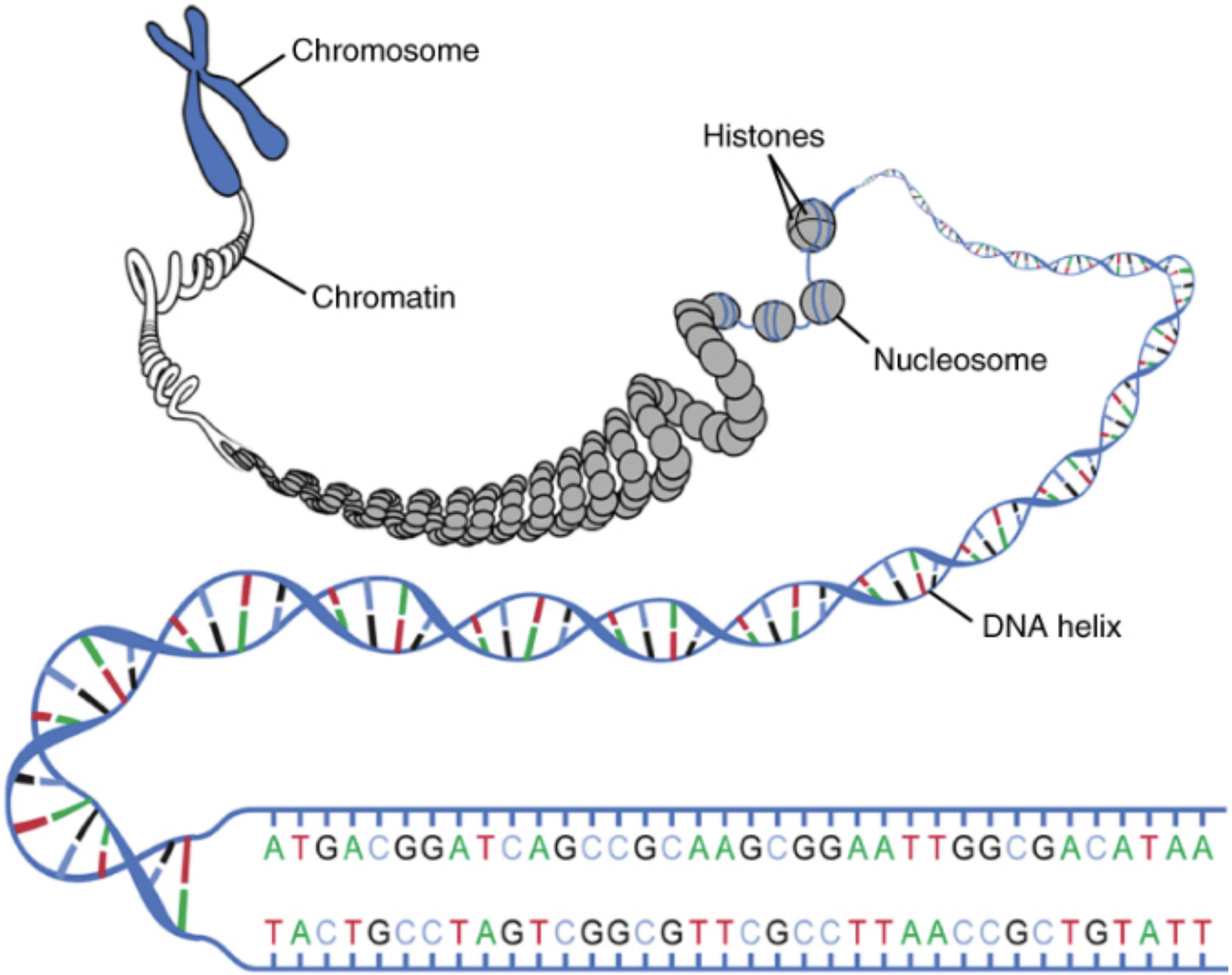
```
5’ ...ACGTGACTGAGGACCGTG... 3’
   ...|||||||||||||||||||...
3’ ...TGCACTGACTCCTGGCAC... 5’
```

- Pair of strings from 4 character alphabet

```
5’ ...ACGTGACTGAGGACCGTG
   CGACTGAGACTGACTGGGT
   CTAGCTAGACTACGTTTTA
   TATATATATACGTCGTCGT
   ACTGATGACTAGATTACAG
   TGATTTTAAAAAAATATT... 3’
```

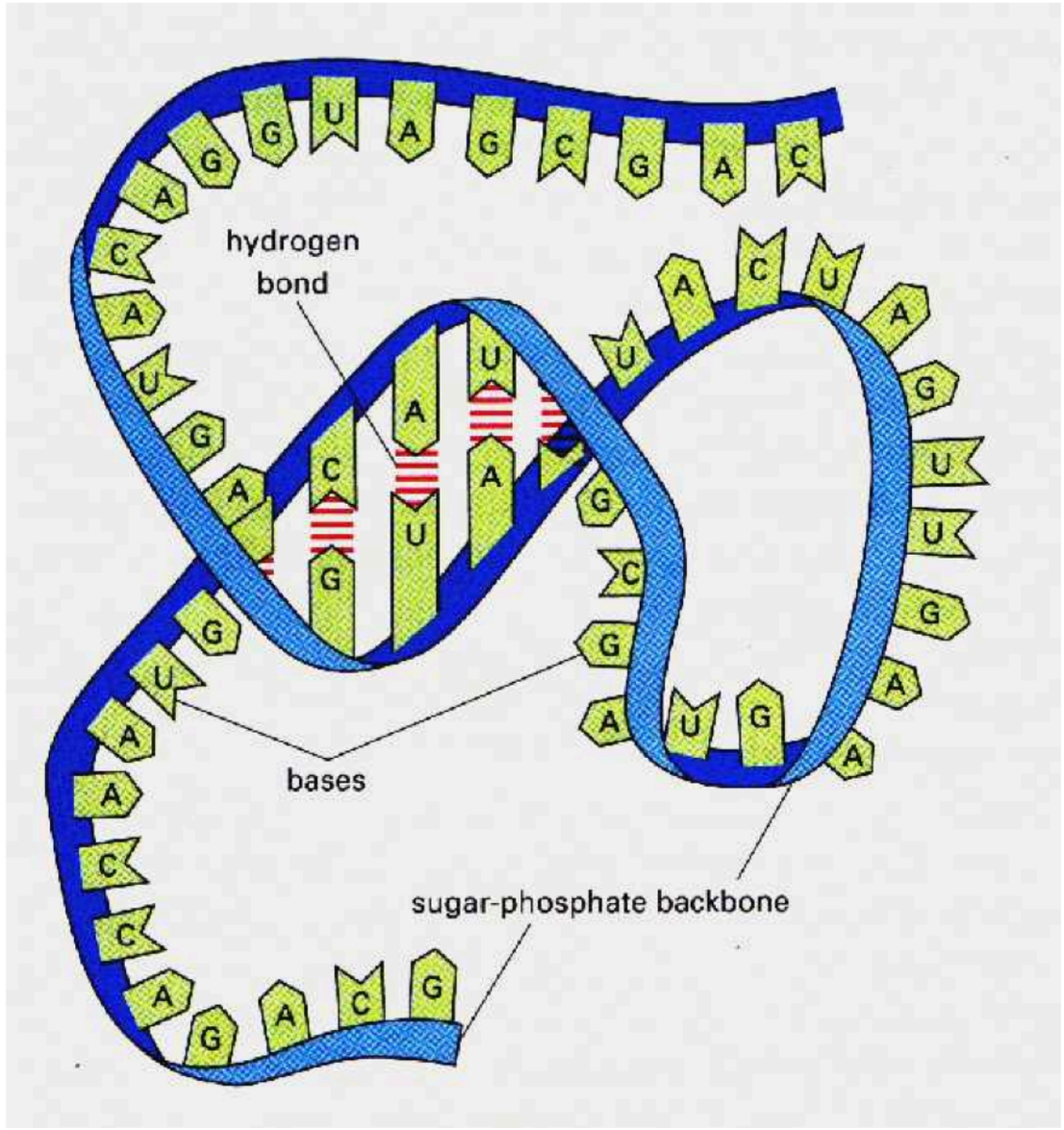
- Single string from 4 character alphabet

DNA forms Chromosomes



RNA

Chemically similar to DNA, but **single-stranded**

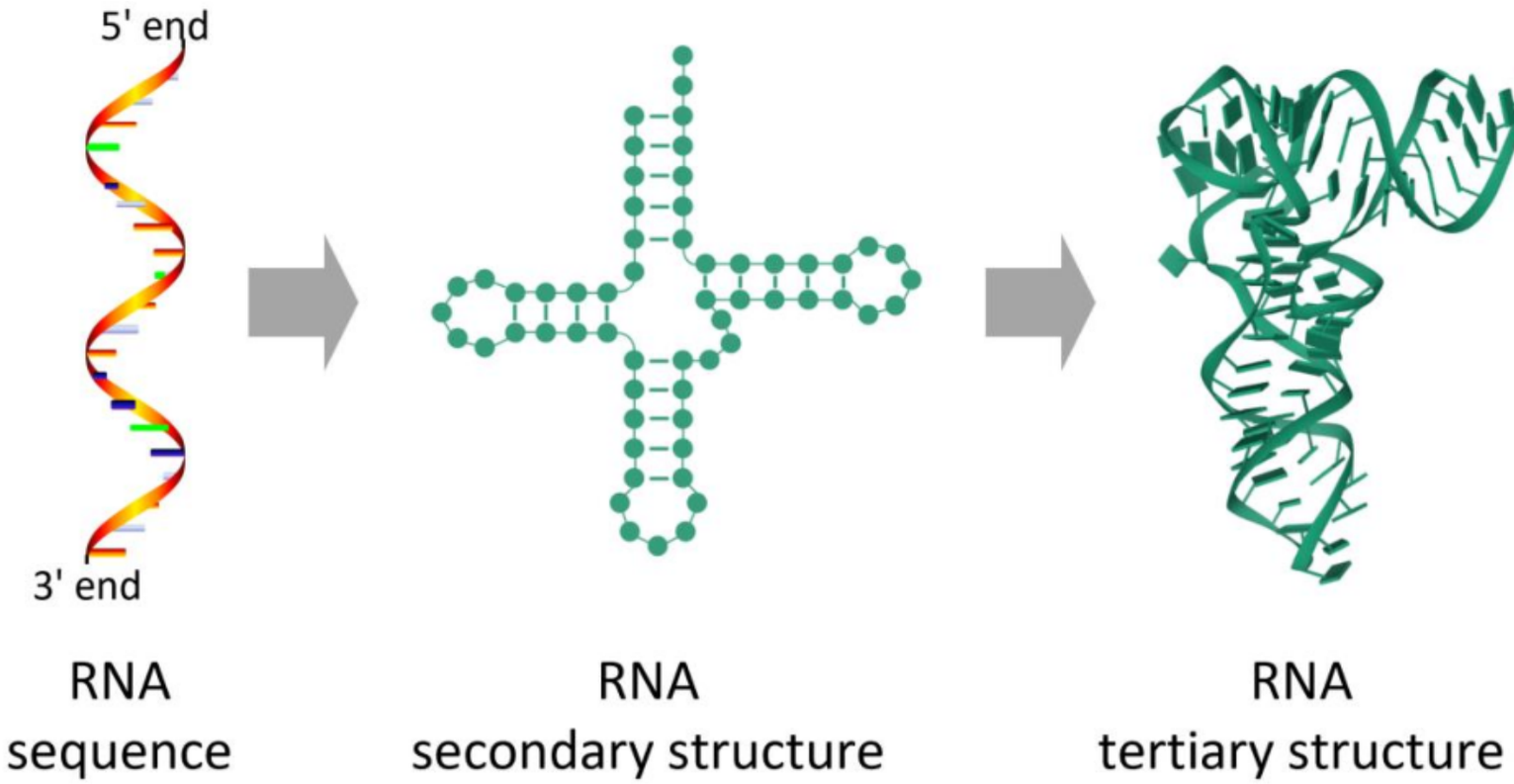


- Bases
  - A(adenine)
  - C(cytosine)
  - U(uracil)
  - G(guanine)

- RNA base complementarity (A—U, C—G)

RNA can fold

RNA can fold into structures due to base complementarity



Types of RNA

- **mRNA** (messenger RNA)
  - carries a gene's information out of nucleus
- **tRNA** (transfer RNA)
  - transfer's mRNA's information onto a protein chain of amino acids
- **rRNA** (ribosomal RNA)
  - part of the ribosome, where proteins are synthesized)

Protein

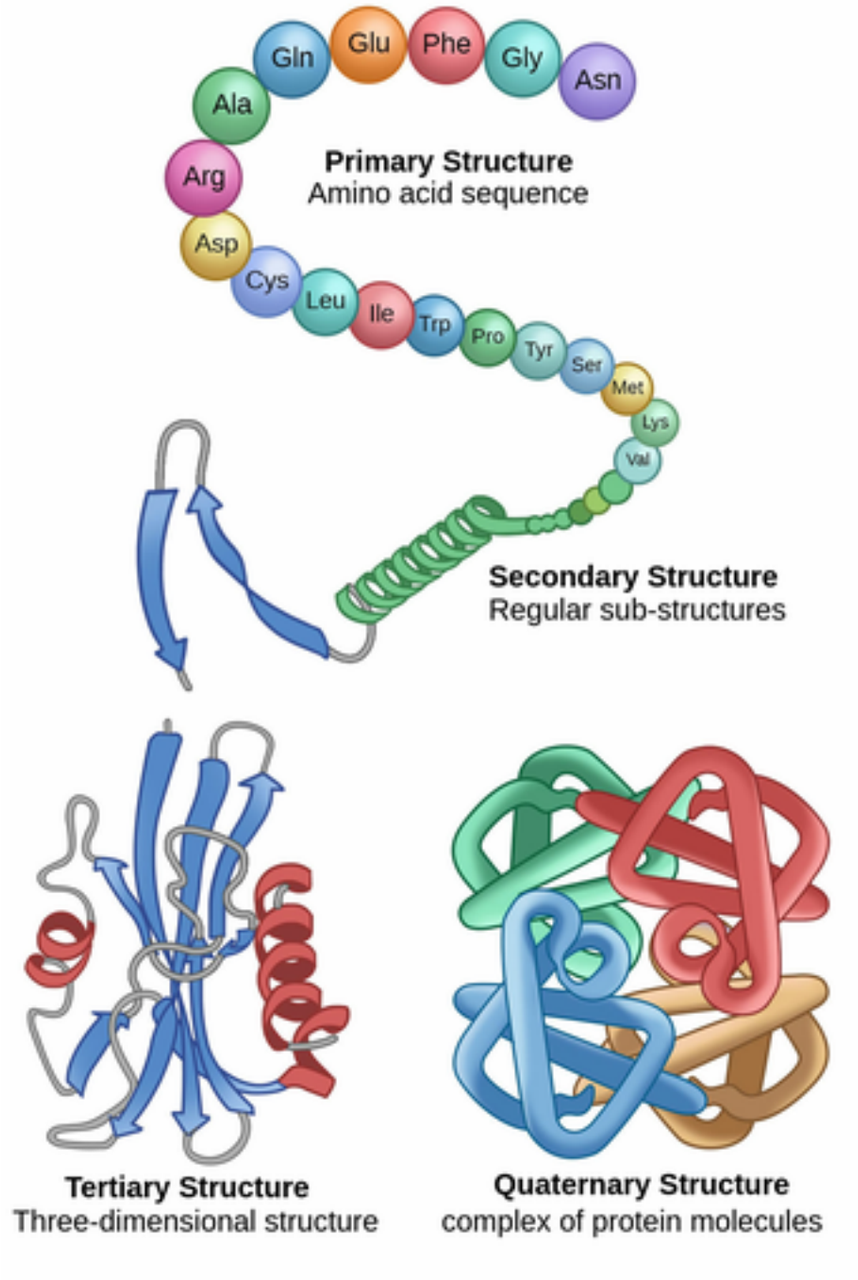
String of amino acids (20 letter alphabet)

…DTIGDWNSPSFFGIQLVSSVHT
TLWYRENAFPVLGGFSWLSWFNW
HNMGYYYPVYHIGYPMIRCGTHL
VPMQFAFQSIARSFALVHWNAPM
VLKINPHERQDPVFWPCLYYSVD
IRSMHIGYPMIRCYQA…

Amino Acid	3-Letters	1-Letter
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Proteins spontaneously fold

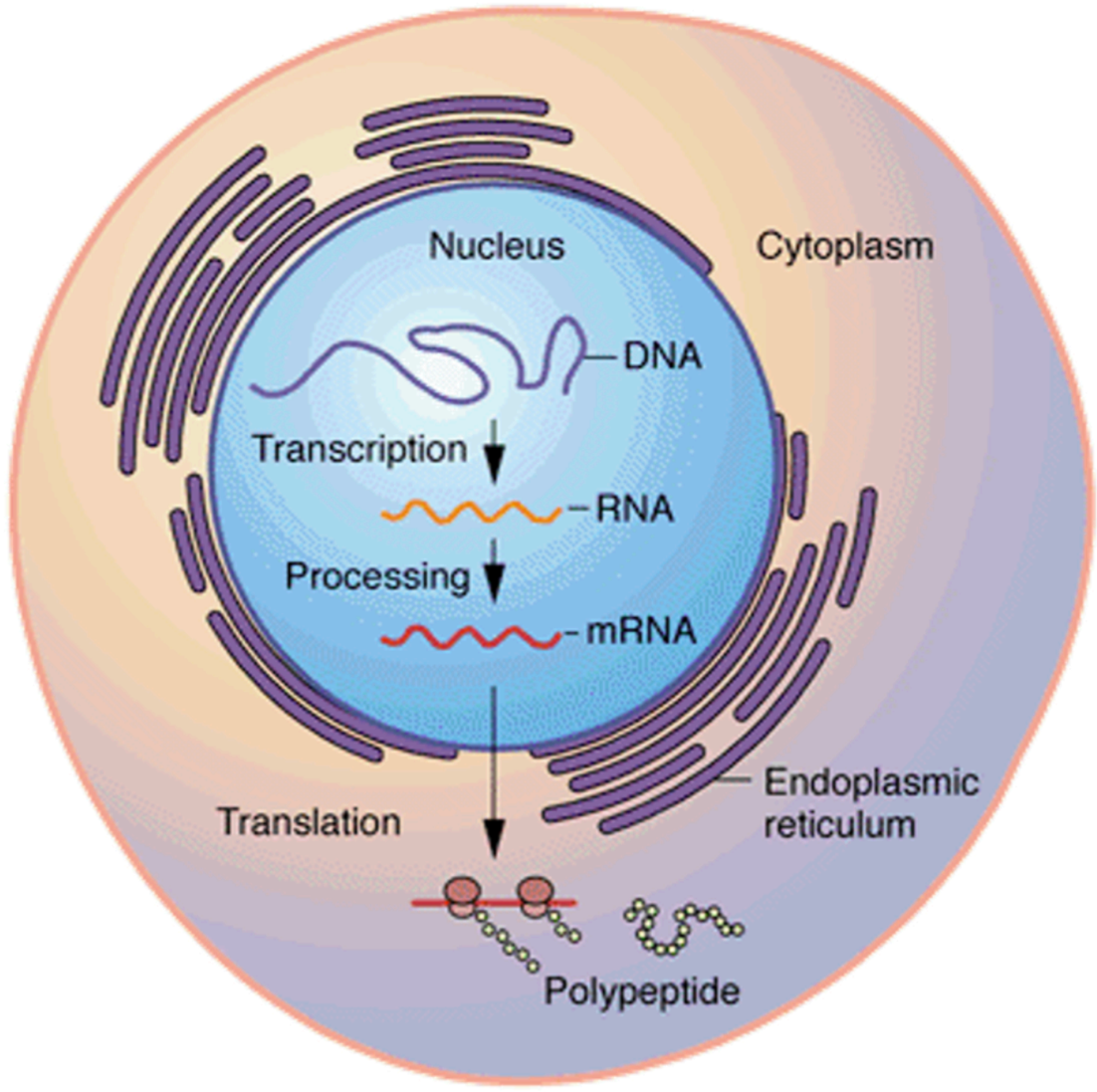
Proteins fold into 3D structures to perform various functions in cells



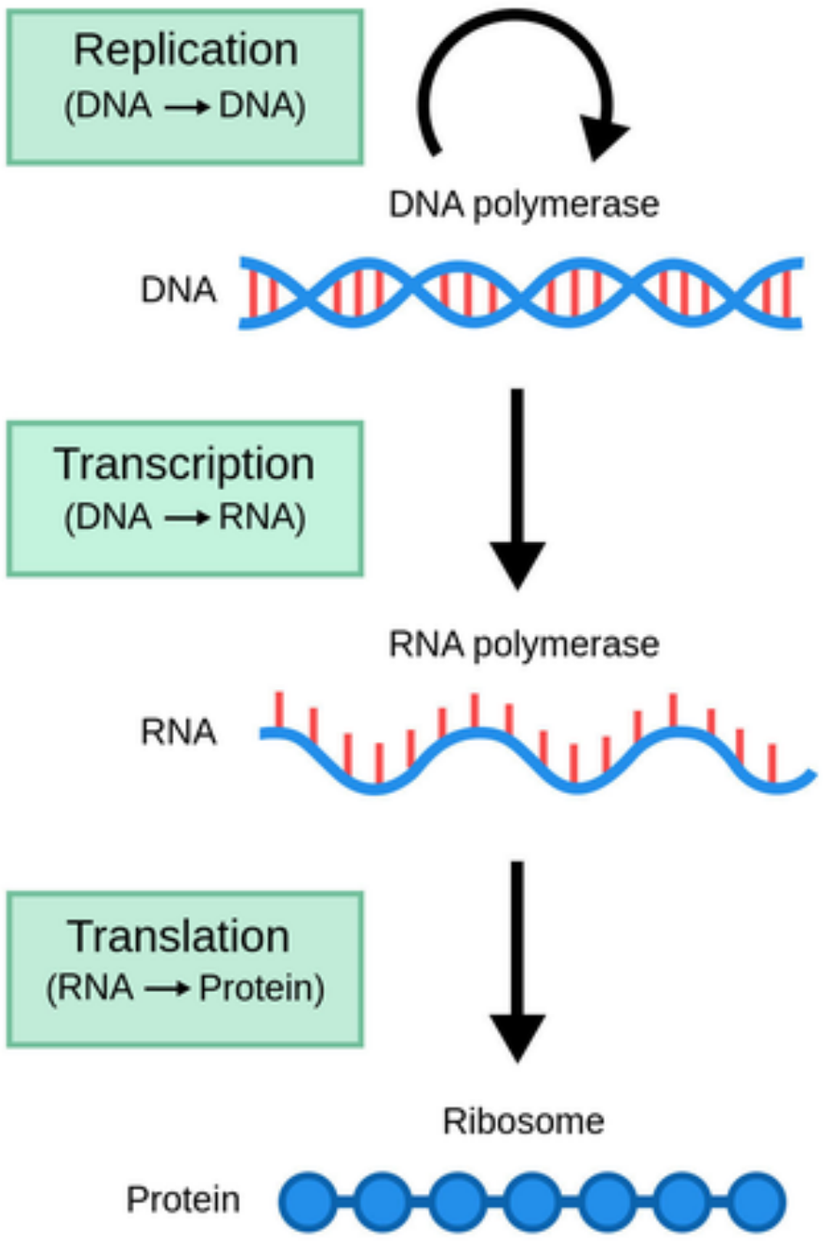
https://theory.labster.com/protein-structure/

Three fundamental macromolecules

- **DNA**
  - Holds information on how the cell works
- **RNA**
  - Acts to transfer short pieces of information to different parts of the cell
  - Provides templates to synthesize into proteins
- **Proteins**
  - Form the body's major components (hair, skin, etc.)
  - Often referred to as the "workhorse of the cell"



Central dogma of molecular biology

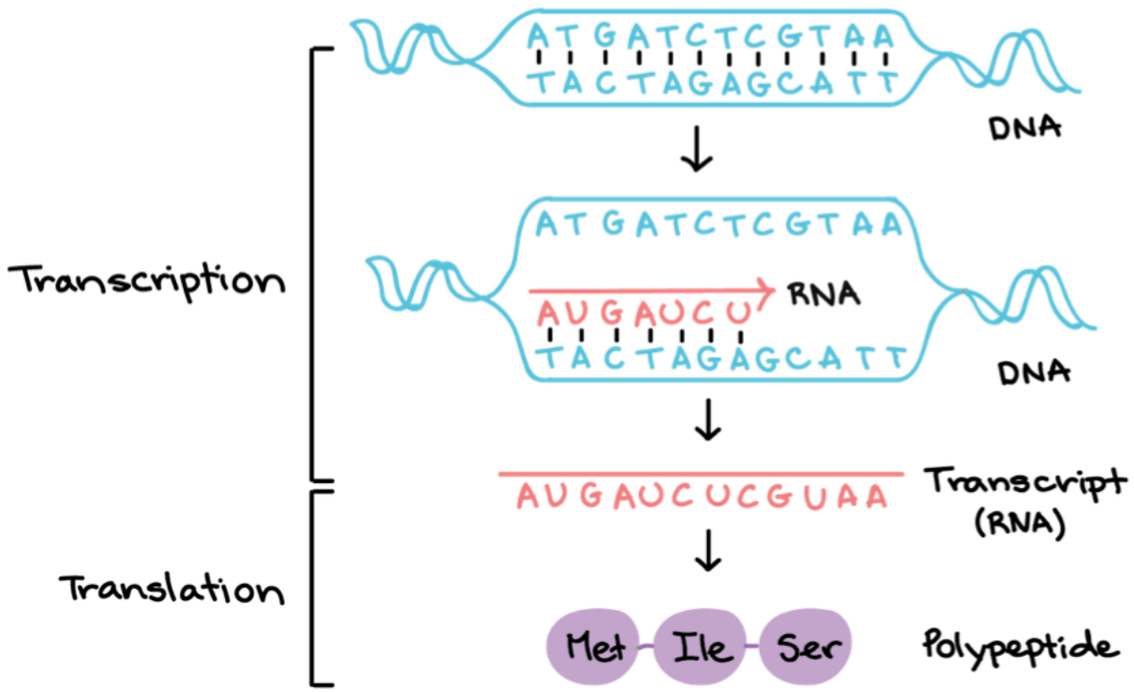


DNA→ RNA → Protein

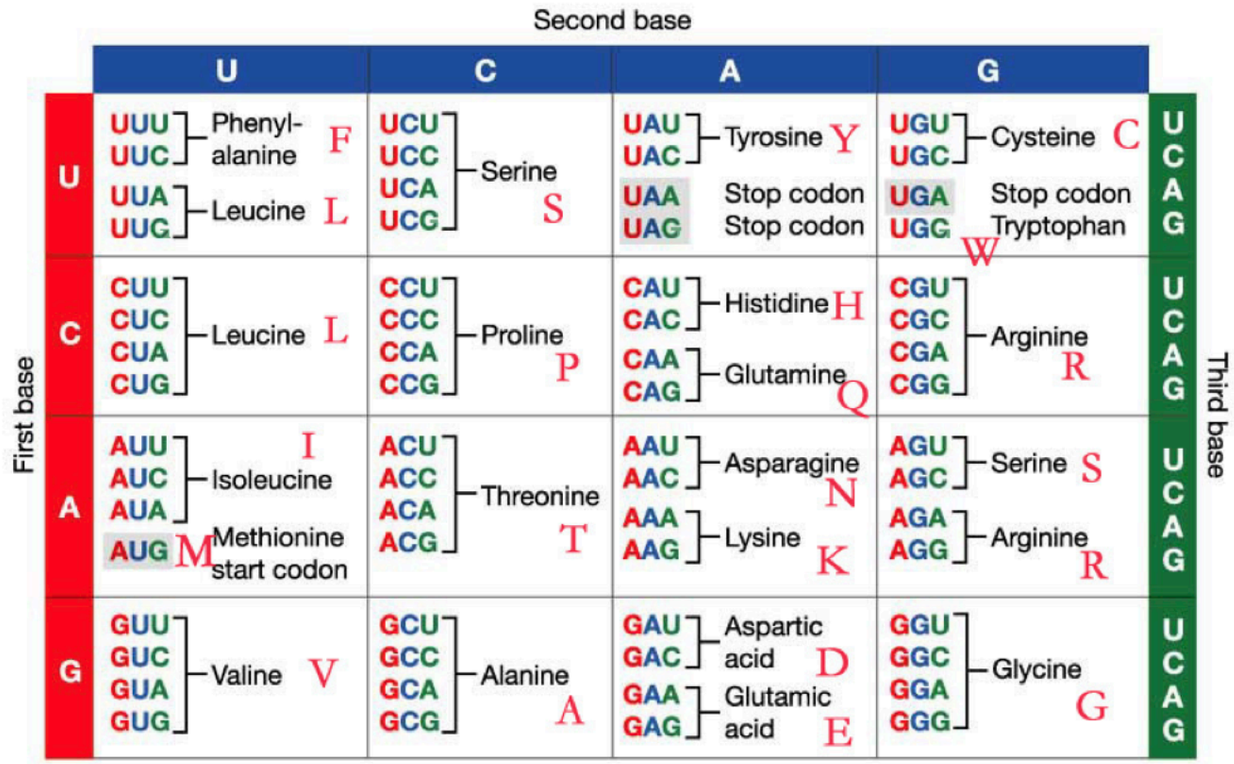
Information flows from DNA through RNA to Synthesize Proteins in cells

https://theory.labster.com/central_dogma_molecular_biology_pre/

Transcription and Translation



https://www.khanacademy.org/science/biology/gene-expression-central-dogma/transcription-of-dna-into-rna/a/overview-of-transcription



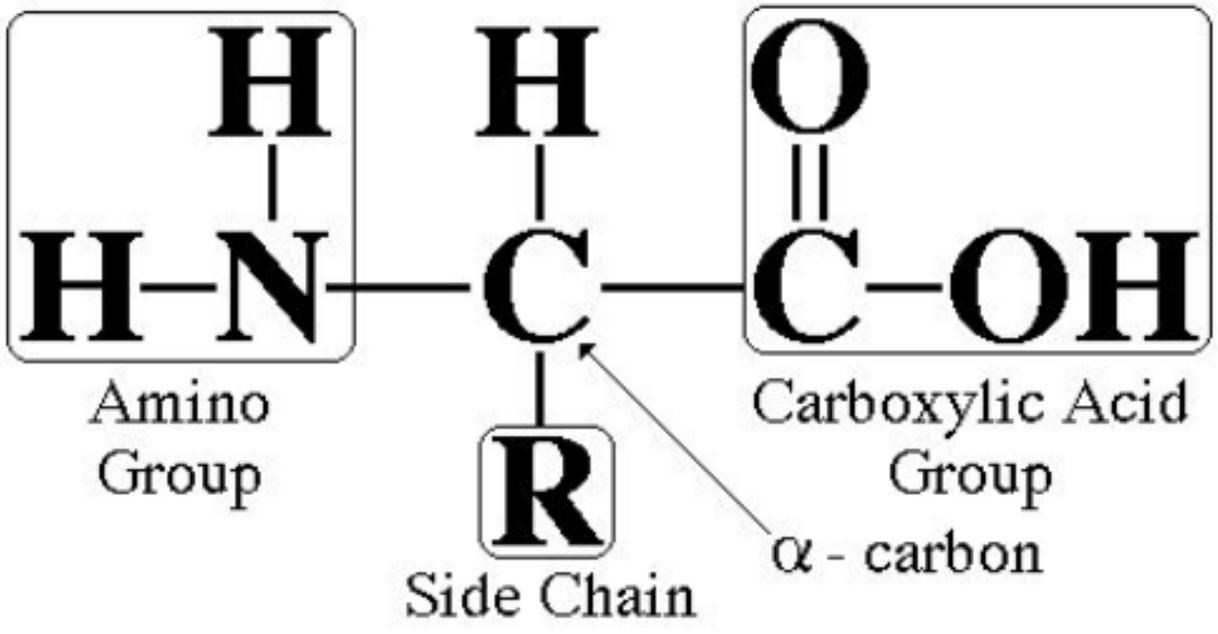
http://bioinfo.bisr.res.in/project/crat/pictures/codon.jpg

Bacterial ribosome translating RNA into protein

https://www.youtube.com/watch?v=q_n0Ij3K_Ho

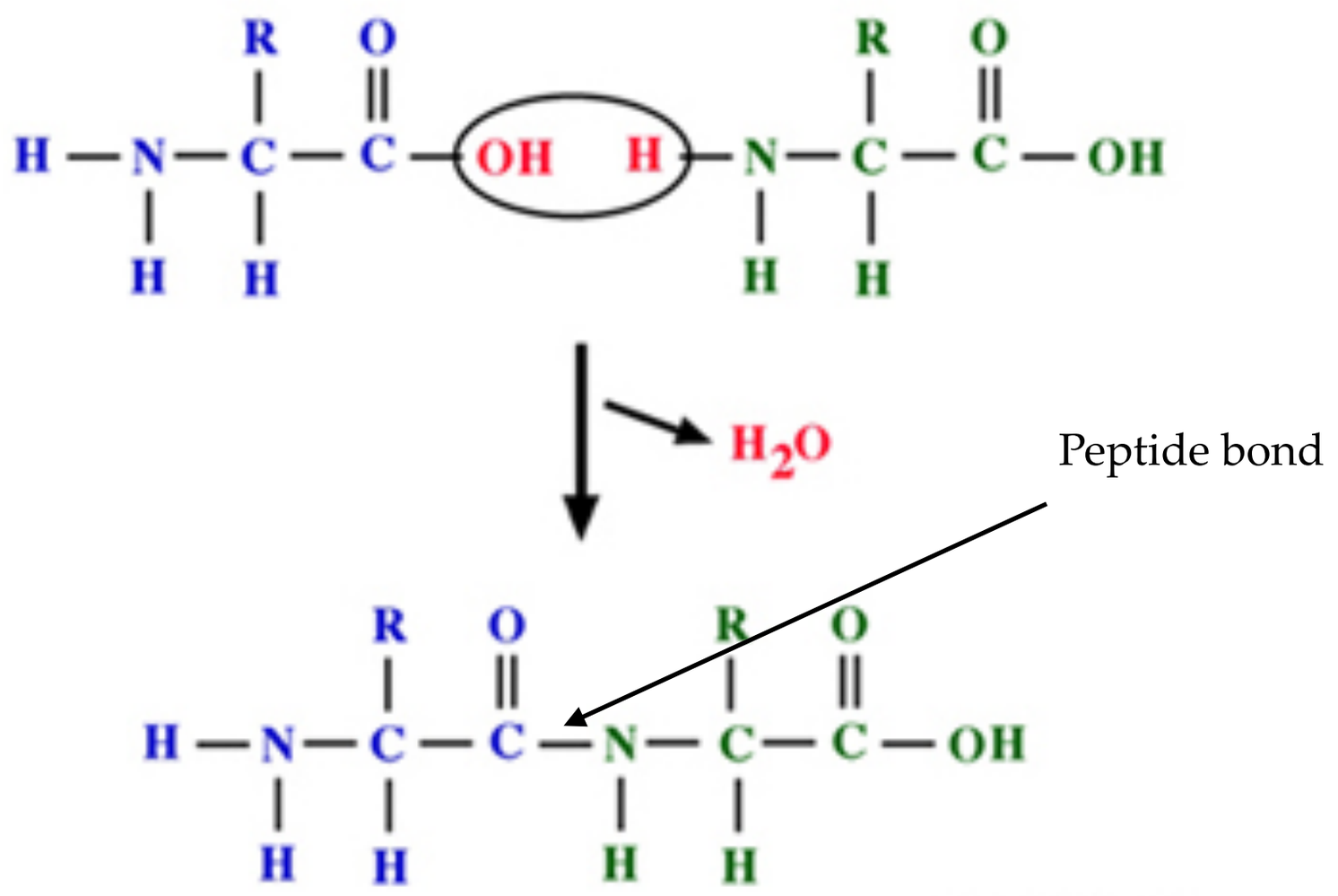
Amino acid

Amino acids are the building block of proteins



Condensation Reaction

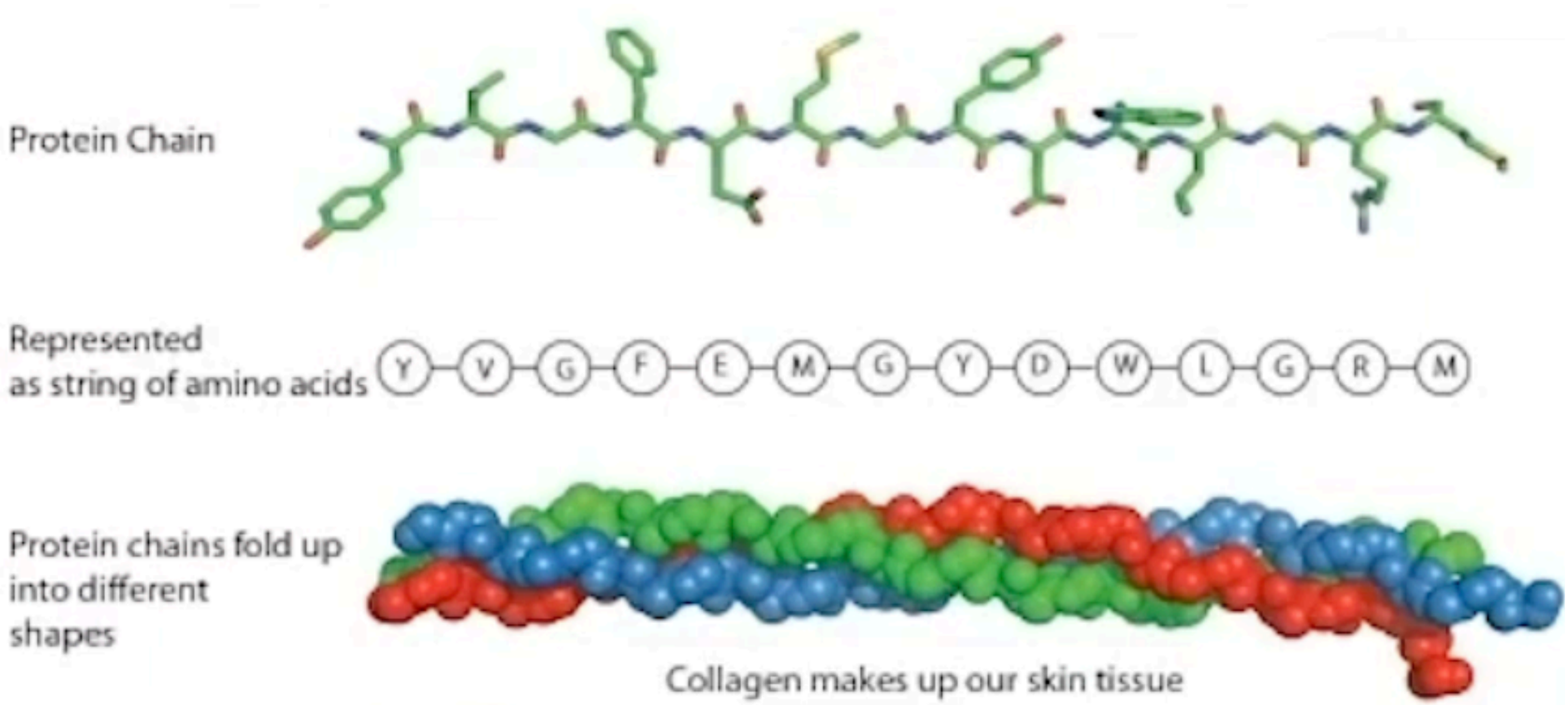
Amino acid residues link together via peptide bonds to form polypeptide chain



http://home.cc.umanitoba.ca/~mshaw/BIOL_1020/lab2/biolab2_4.html

Polypeptide chain

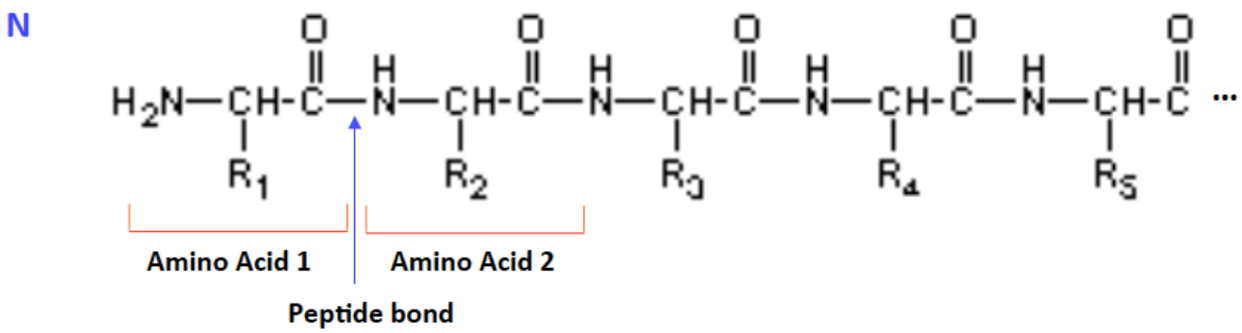
Amino acid residues link together via peptide bonds to form **polypeptide chain**

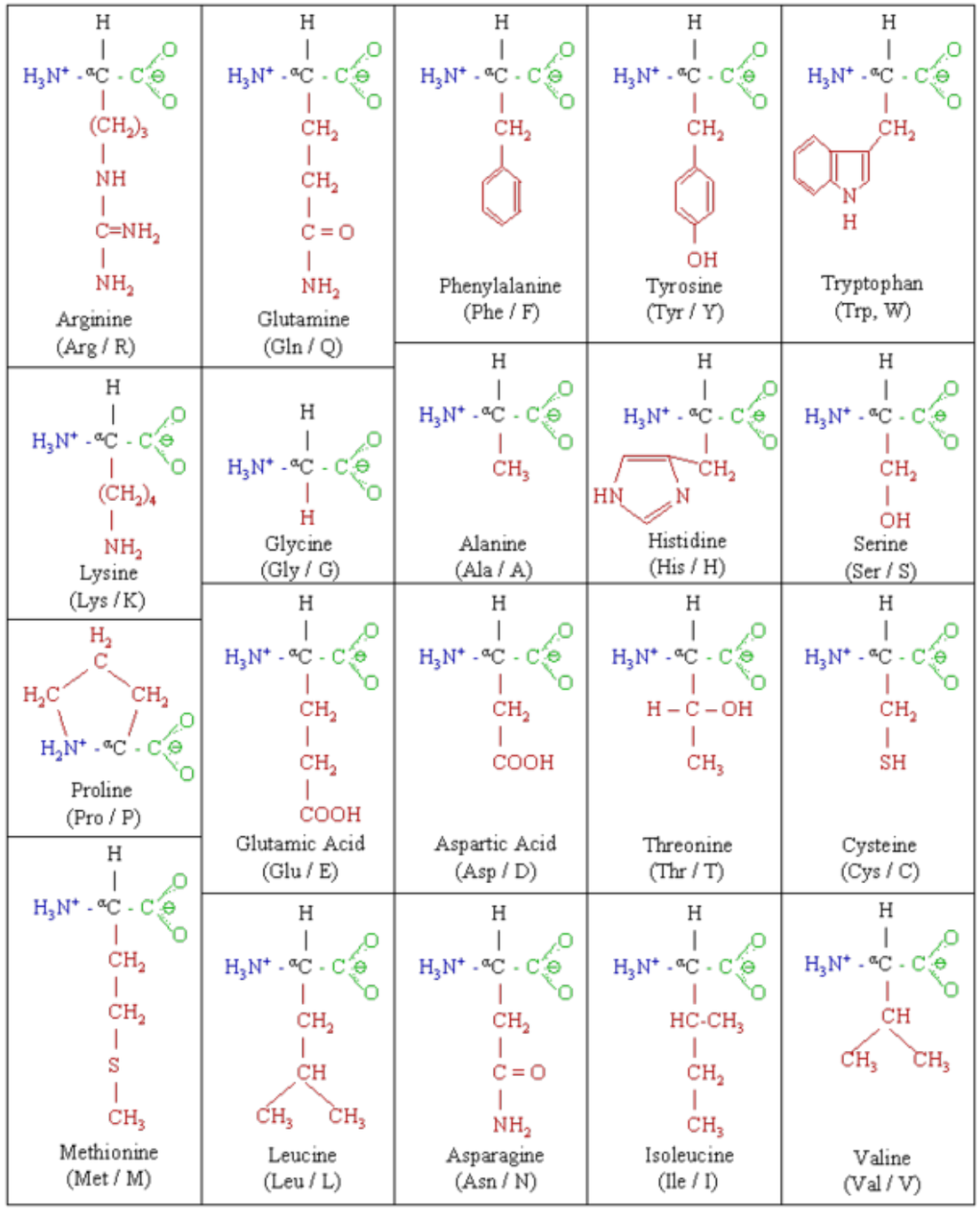


The Gene Machine - V. Ramakrishnan

Proteins

A directional sequence of amino acids/residues





20 naturally occurring amino acid residues