

Structural Bioinformatics

WS 2021/22

Lecture 1

Introduction to 3D structure of proteins and
nucleic acids



UNIVERSITÄT
DES
SAARLANDES

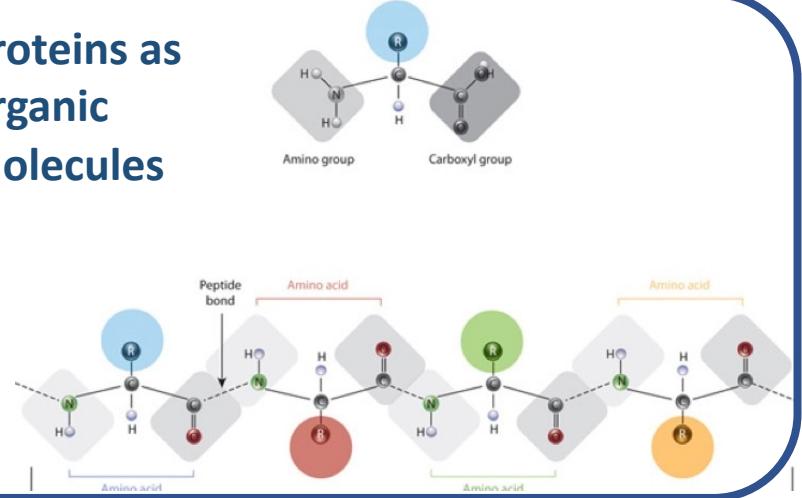


Topics of this course

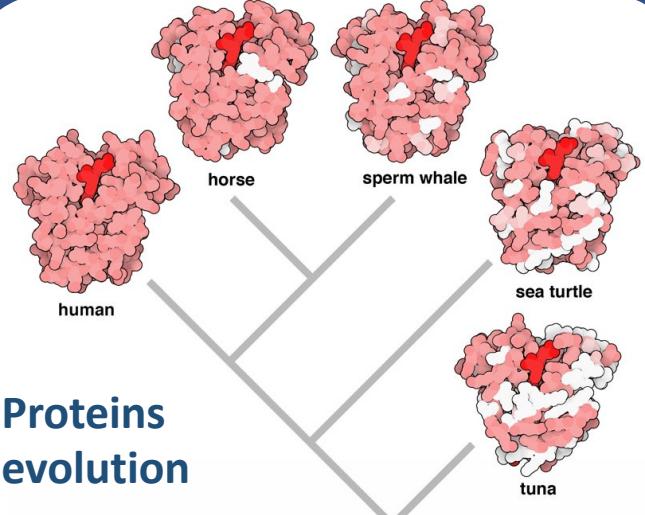
- Three-dimensional organization of biological molecules
- Selected methods for predicting structural and functional features in proteins
- Key concepts in protein evolution
- Prediction of protein 3D structure using homology modelling
- Approaches to comparing protein 3D structures
- Prediction of protein 3D structure without detectable homology
- Applications of protein evolution to prediction of their 3D structure
- Relevant algorithms / computational techniques

Topics of this course

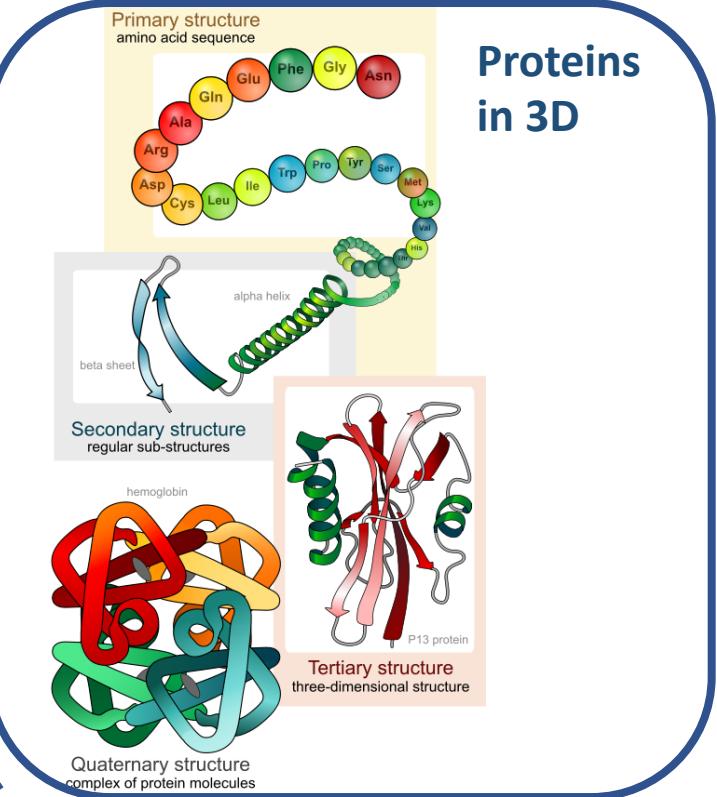
Proteins as organic molecules



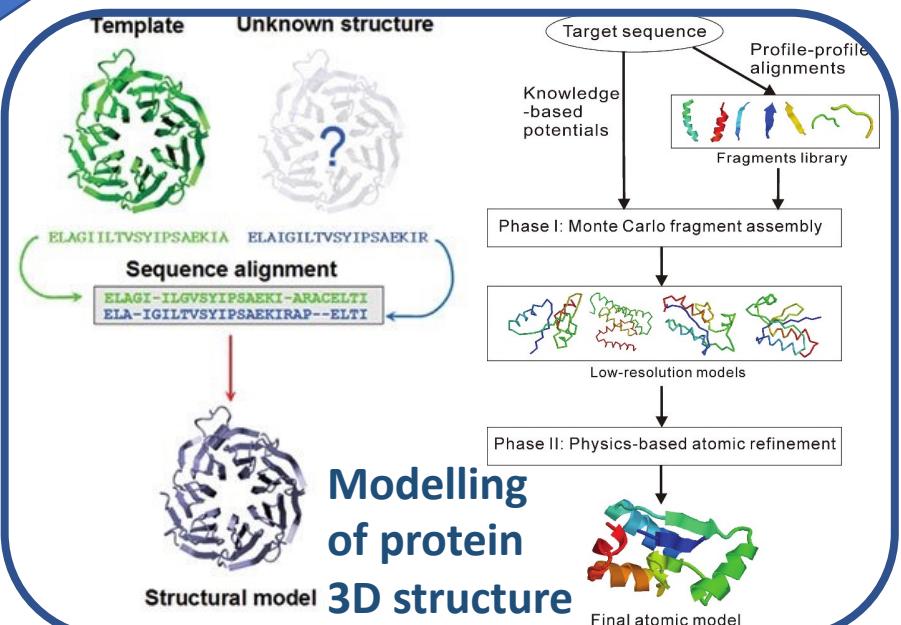
Proteins evolution



Proteins in 3D



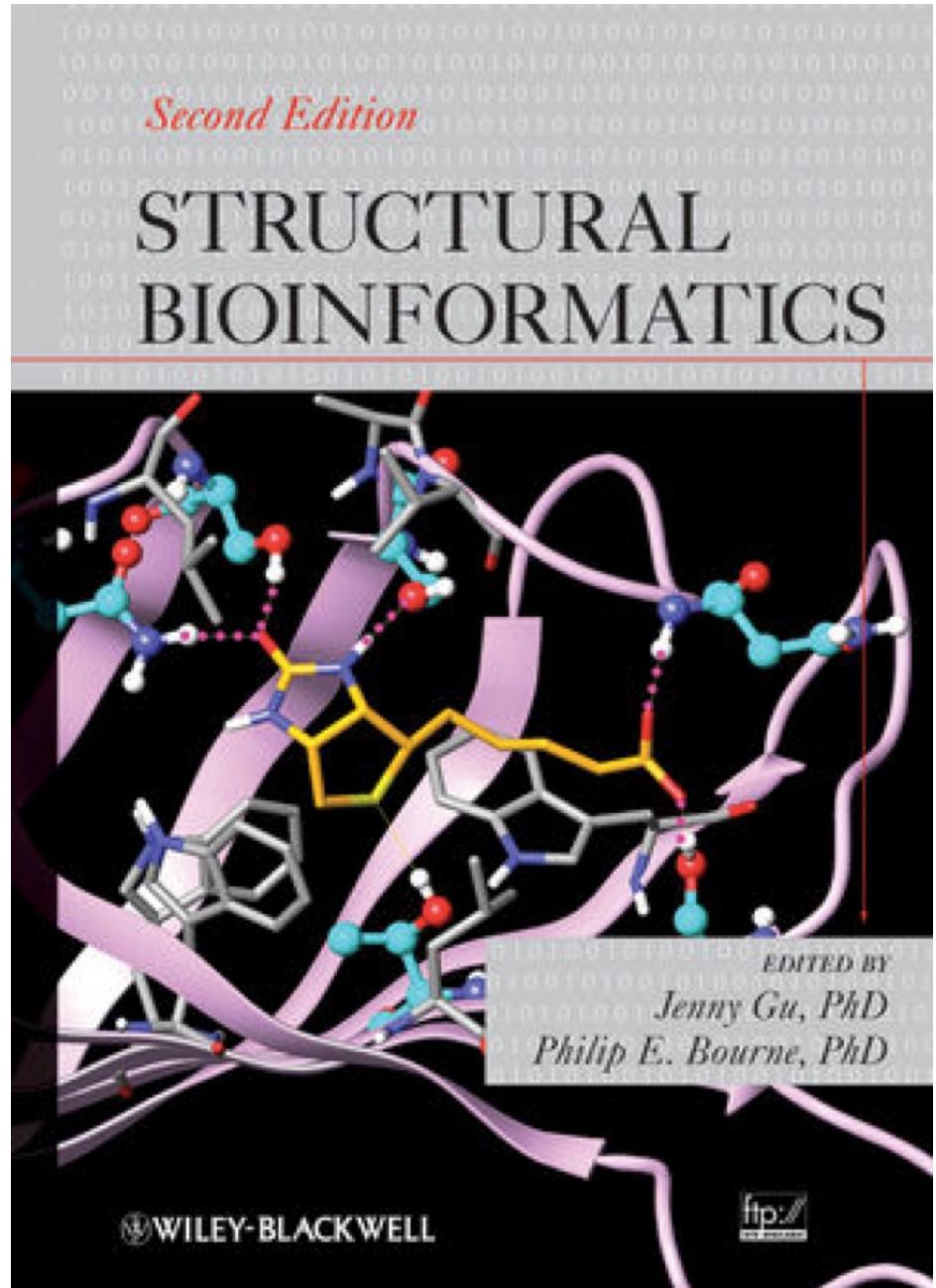
Modelling of protein 3D structure



How all this contributed to AI-based methods and made protein structure prediction possible

Textbook(s)

- Structural Bioinformatics,
2nd Edition
- Jenny Gu (Editor),
Philip E. Bourne (Editor)
- ISBN: 978-0-470-18105-8
- Mar 2009, Wiley-Blackwell
- Semesterapparat:
http://www.infomath-bib.de/tmp/vorlesungen/bioinfo_structural_bioinformatics.html



Structural bioinformatics

Problems

- Predicting protein 3D structure
- Nucleic acids structure
- Structure comparison
- Secondary structure
- Protein function
- Interaction interfaces
- Prediction of complexes
- Drug virtual screening
- Epitope prediction
- Molecular dynamics

Methods

- Statistical hypothesis testing
- Bayesian methods
- Dynamic programming, linear programming
- Optimization techniques
 - Monte Carlo, EM
- Markov chains, hidden Markov models
- Machine learning
 - Clustering, SVMs, decision trees, artificial neural networks
- Deep learning

This course

- **Lectures:** Wednesday 12:15 — 13:45, every week, 001 E2.1 + MS Teams
- **Tutorials:** Tuesday 16:00 — 18:00, biweekly, MS Teams
TA: Alper Yurtseven, Tanya Malygina
First session: Oct 26 (next week)
 - Exercise sheets due **Friday 12:00 (noon)** before the next tutorial session
 - Work in groups of 2-3 (not more!) people, one person represents each group
- **Exam:** 2 attempts, probably written
 - 1st attempt: 1st or 2nd week of February
 - 2nd attempt: end of spring break
- Registration per email: olga.kalinina@helmholtz-hips.de ;
s8alyurt@stud.uni-saarland.de

Topics of lectures

1. 20.10.: Introduction to properties and structure of biological macromolecules
2. 27.10.: Experimental techniques in structural biology; Protein Data Bank
3. 3.11.: Protein structural organization, classification of proteins by structure
4. 10.11.: Prediction of structural features from sequence
5. 17.11.: Evolution and comparison of protein sequences
6. 24.11.: Advanced sequence similarity search using hidden Markov models
7. 1.12.: Homology-based modelling of protein 3D structure
8. 8.12.: Modelling of protein 3D structure by threading
9. 15.12.: Fragment-based prediction of protein 3D structure
10. 5.01.: Comparison of protein 3D structures and of models to experimental structures
11. 12.01.: Prediction of inter-residue contacts and implications for protein 3D structure prediction
12. 19.01.: Introduction to molecular dynamics simulations
13. 26.01.: Backup
14. 2.02.: Q&A
15. 9.02.: Exam, 1st attempt (2nd attempt: end of March / beginning of April)

AlphaFold



Topics of tutorial sessions

1. 26.10.: Getting to know protein structure
 2. 9.11.: Protein structure visualization, introduction to PyMol
 3. 23.11.: Scripting for PyMol
 4. 7.12.: Comparison of protein sequences and structures
 5. 11.01.: Homology-based modelling
 6. 25.01.: Backup or Q&A
-
- Exercise sheet are due **Friday 12 pm (noon)** before the next tutorial; Due dates:
 1. 5.11.
 2. 19.11.
 3. 3.12.
 4. 7.01.
 5. 21.01

Prerequisites

- **Difficult to set, since backgrounds differ, take the following as an estimate...**
- Foundations of bioinformatics
 - Basics of sequence analysis: comparison, alignment, search
 - Common bioinformatics resources
- 101 level in
 - Molecular biology
 - Biochemistry
 - Statistics
- Hands-on experience in programming
 - Preferably Python, but other languages accepted

Exam

- Probably oral; online or in-person
- Prerequisite: all exercise sheets submitted on time, received >50% points on both theoretical and programming parts
- 2 attempts
 - Last lecture slot
 - end of spring break (time and place will be given later)
- 2 parts
 - general questions to the course
 - specific questions to a scientific paper (one student answers questions to one paper, which he/she picks from a list that will be announced around Christmas break)

Expected gained theoretical competencies

- Understanding of physical and chemical basics of three-dimensional structure of biomolecules
- Understanding of relationship between similarity and evolution
- Knowledge of basic principals of algorithms for sequence similarity search
- Knowledge of basic steps of algorithms for modelling of protein 3D structures
- Understanding the concepts of comparison and alignment of protein 3D structures; knowledge of measures for comparison of protein 3D structures
- Understanding of basic principles of molecular dynamics simulation methods

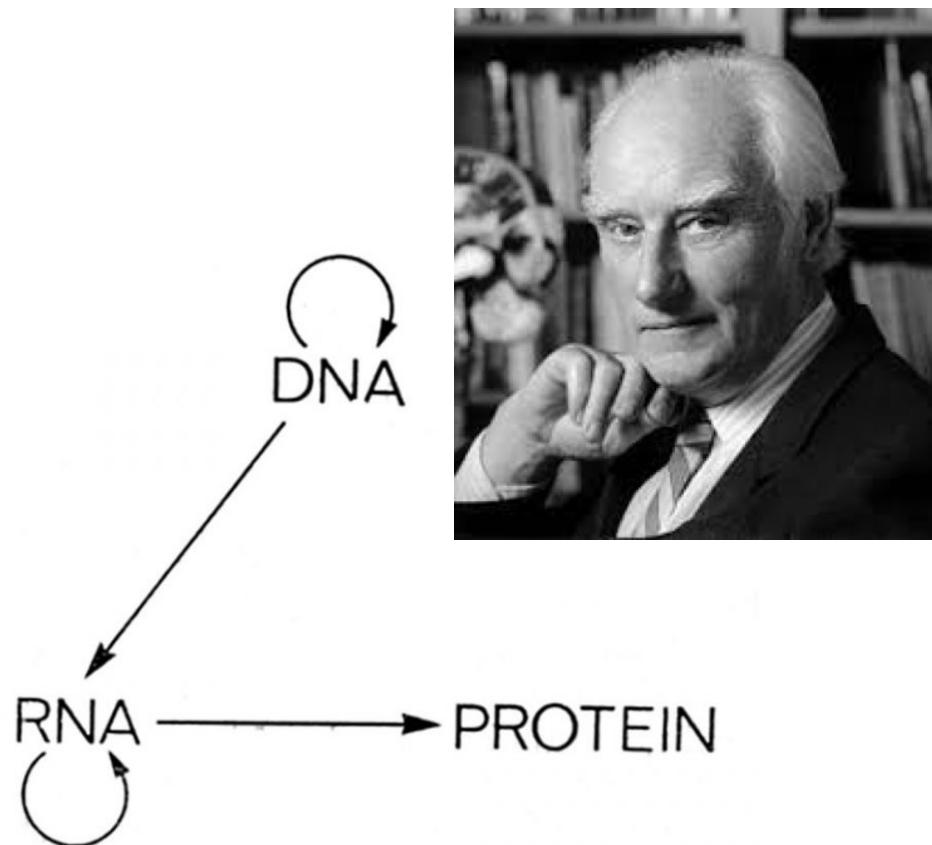
Expected gained practical competencies

- Given a protein sequence, be able to identify its potential functional family, make guesses about cellular localization
- Be able to search for experimentally resolved protein 3D structures, related information in publicly available resources
- Understand the common formats of storing structural data
- Be able to programmatically analyze and manipulate corresponding files
- Be able to visualize protein 3D structures, describe it, analyze it with respect to potential protein function
- Build 3D models of protein structures based on homology-modelling approach using online resources and programmatically

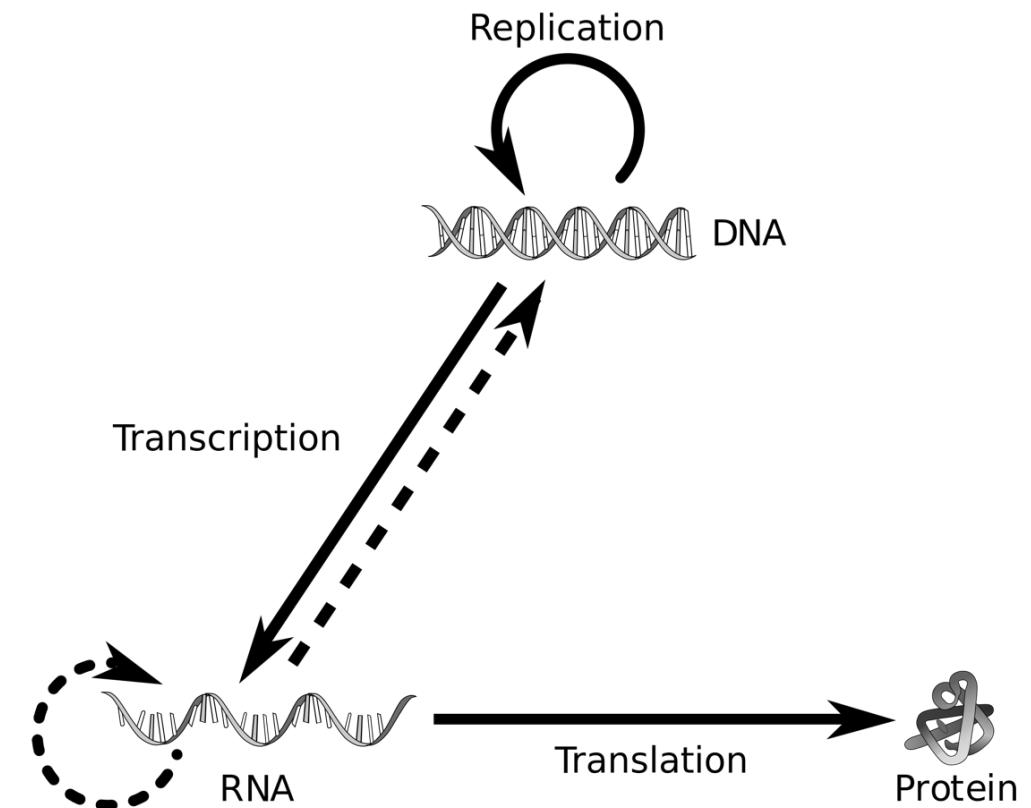
Questions to the organizational part?

The Central Dogma of Molecular Biology

Flow of information in all living systems



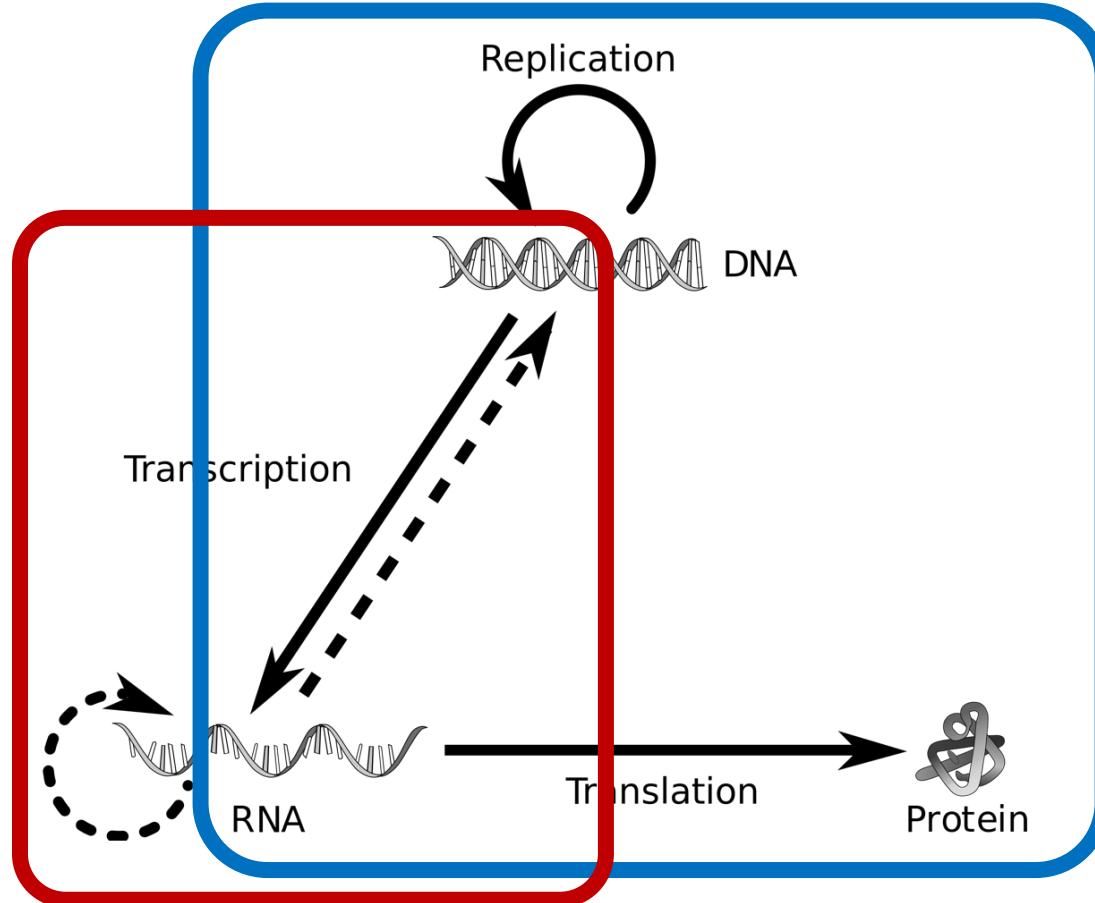
Crick's 1958 version of the
“central dogma” of biology from
Crick (1970, Figure 3, p. 562).



[https://commons.wikimedia.org/wiki/
File:Central_dogma_of_molecular_biology.svg](https://commons.wikimedia.org/wiki/File:Central_dogma_of_molecular_biology.svg)

What living systems?

dsRNA-, ss(+)RNA-,
ss(-)RNA-viruses,
retroviruses

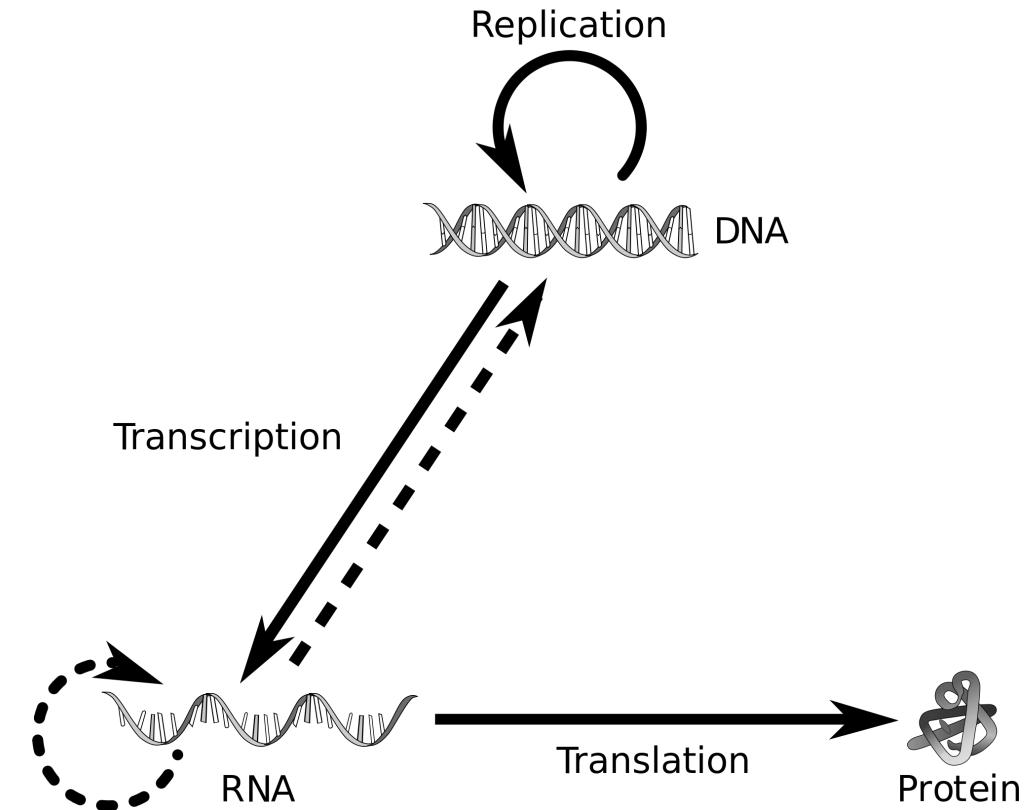


Eukaryotes
Bacteria
Archaea
DNA-containing
viruses

[https://commons.wikimedia.org/wiki/
File:Central_dogma_of_molecular_biology.svg](https://commons.wikimedia.org/wiki/File:Central_dogma_of_molecular_biology.svg)

Physical and chemical properties of the biomolecules involved make it possible

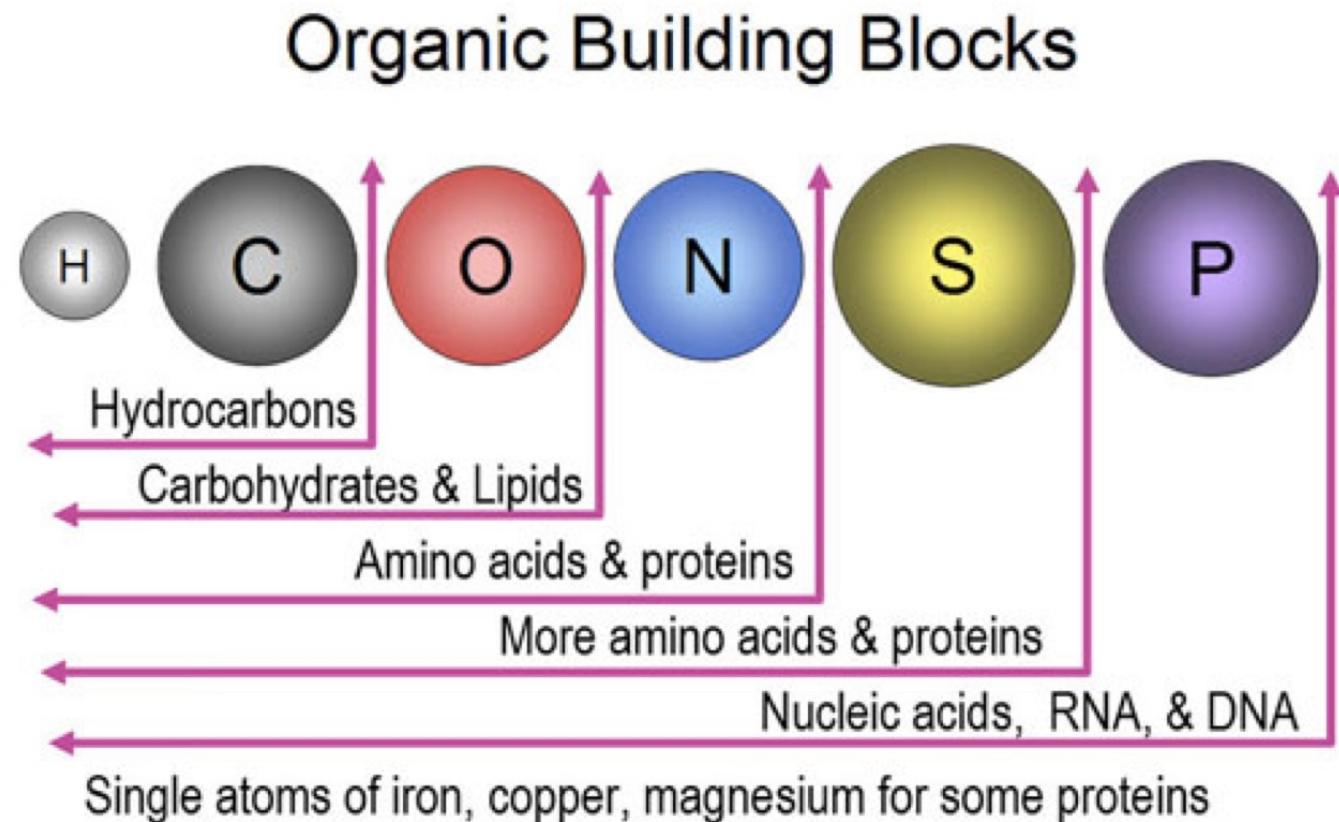
- DNA, RNA, and proteins are polymers
- Base pairing in the DNA (RNA)
- **Stability:** long double-stranded RNA molecules are less stable, that's why genetic information is stored in DNA
- Proteins consist of units (amino acids) with **different chemical properties** and can **fold into complex 3D structures =>** variety of activities



Detour: Basics of organic chemistry

Atoms

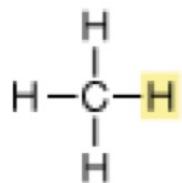
- Organic compounds: contain carbon (C)
 - except carbides, carbonates and simple carbon oxides
- Contain a limited set of elements



Families of organic compounds

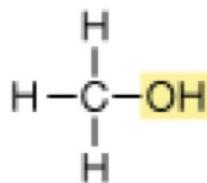
The leading families of organic compounds

Hydrocarbon



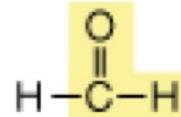
methane
(CH₄)

Alcohol



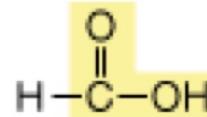
methyl alcohol
(CH₃OH)

Aldehyde



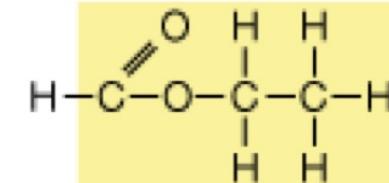
formaldehyde
(HCHO)

Carboxyl group Acid



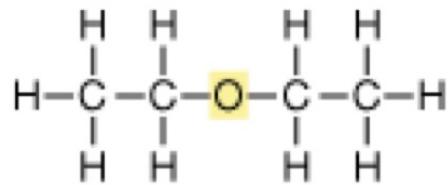
formic acid
(HCOOH)

Ester



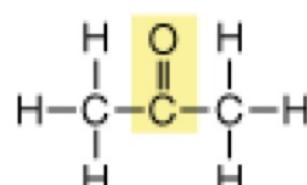
ethyl formate
(C₂H₅COOH)

Ether



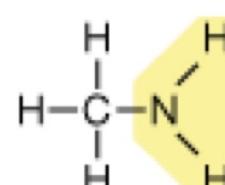
diethyl ether
(C₂H₅OC₂H₅)

Ketone



acetone
(CH₃COCH₃)

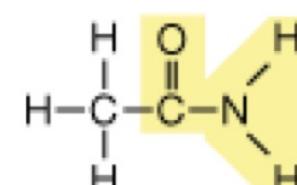
Amine



Amino group

methyl amine
(CH₃NH₂)

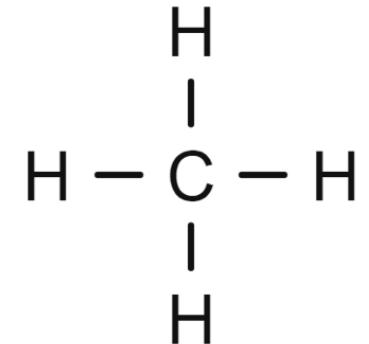
Amide



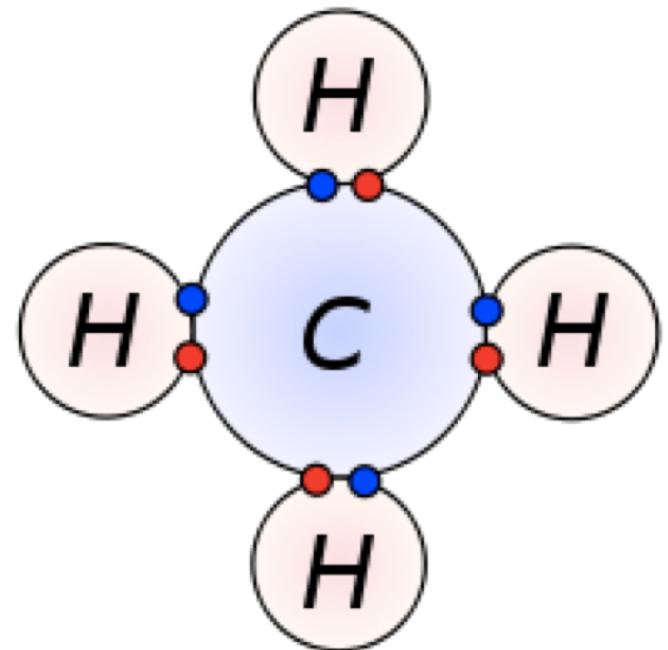
acetamide
(CH₃CONH₂)

© 2012 Encyclopædia Britannica, Inc.

Bonds: covalent

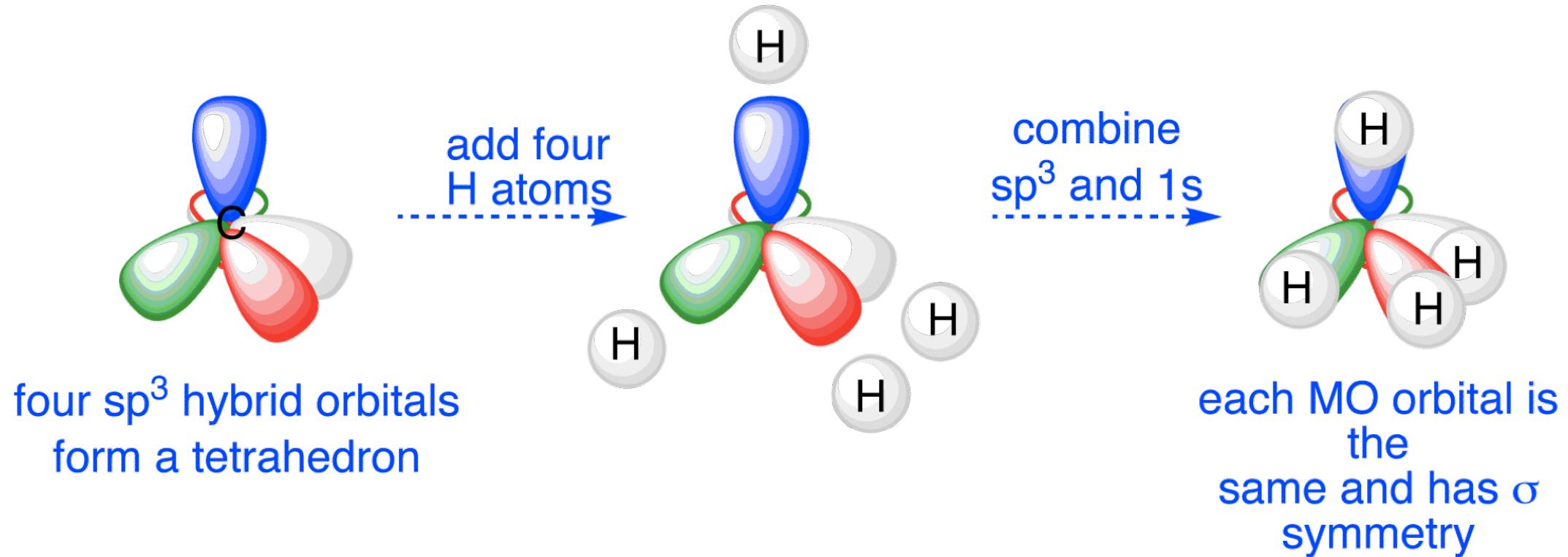


- **Covalent bond:** chemical bond that involves the sharing of electron pairs between atoms
 - Very tight: in practice, needs a catalyst to be broken
 - => chemical reaction



- Electron from hydrogen
- Electron from carbon

Does not represent the reality fully



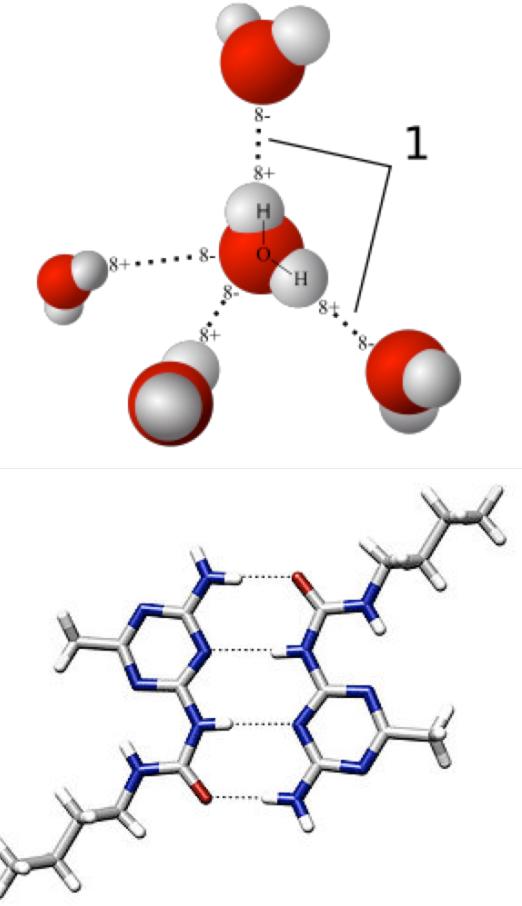
- Molecular orbitals, quantum properties of electrons...
- **Way too complex for this course!**

Bonds: non-covalent

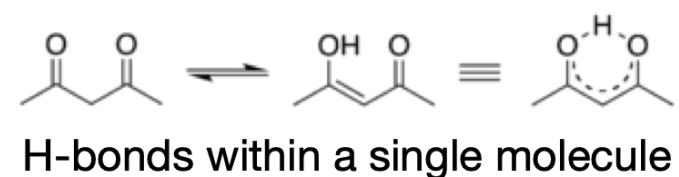
- **Non-covalent bonds:** bonds that do not involve sharing electrons between atoms (or not quite...)
 - => much easier to break, do not require a catalyst
- 3 most important types (for us)
 - hydrogen bonds
 - salt bridges
 - van der Waals interactions

Hydrogen bonds

- Non-covalent interaction between polar molecules
- Electronegative-H ... electronegative
 - H is partially shared between the two molecules
- Hydrogen donor and acceptor
- 5...30 kJ/mol
 - up to 161 kJ/mol between F-H:F

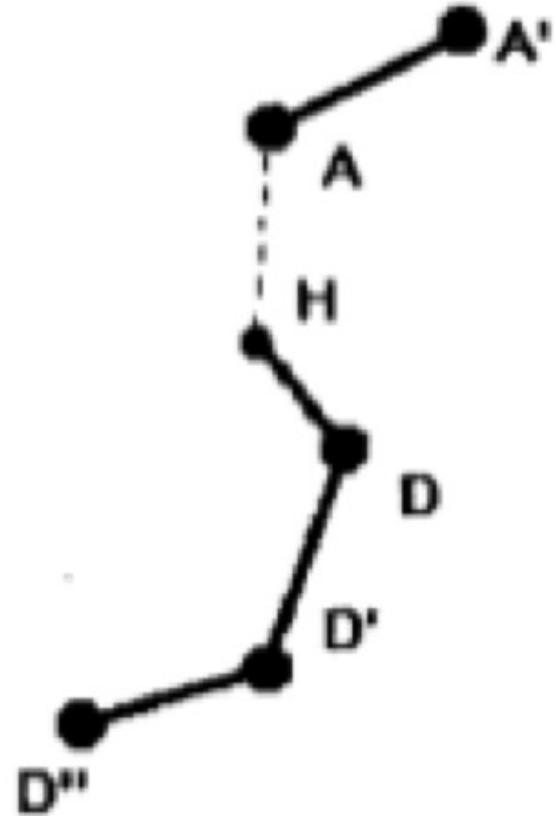


Multiple H-bonds between
two molecules



Geometry of hydrogen bonds

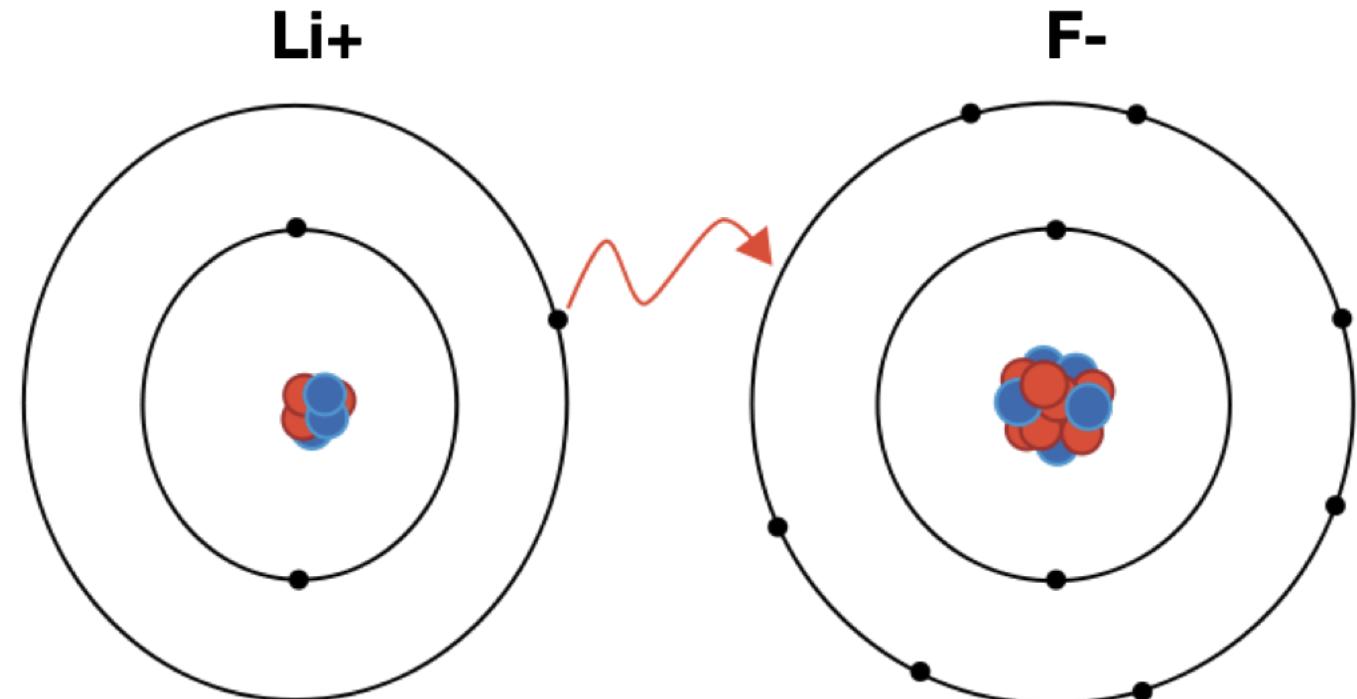
- $d_{HA} < 2.5 \text{ \AA}$
- $\angle DHA > 90^\circ$ (or even 120°)



Troshin et al., 2002

Ionic bond

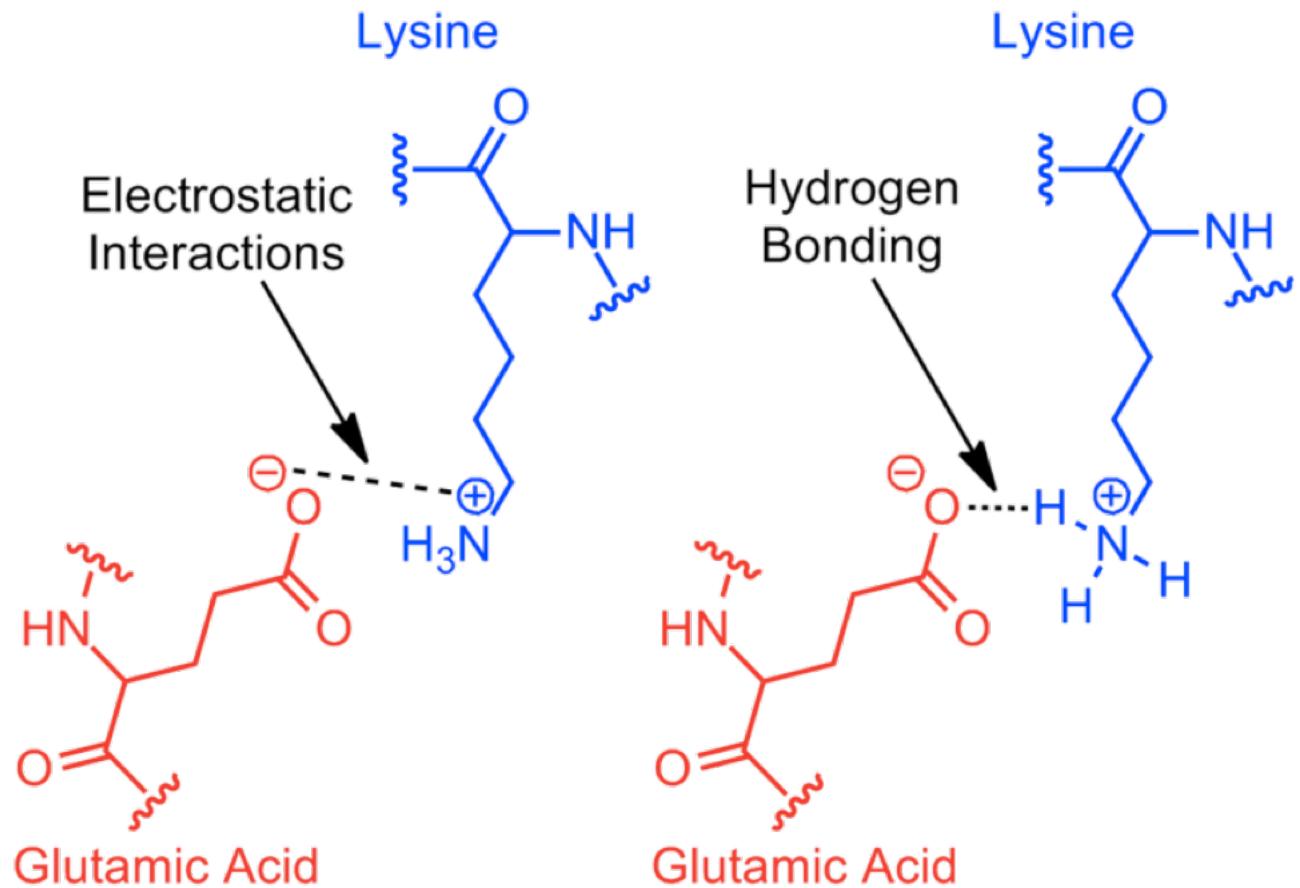
- Chemical bond that involved **electrostatic attraction** between ions
 - involves some electron sharing
 - no directionality



By EliseEtc / vectorised from Ionic bonding.png - Own work, CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=18282119>

Salt bridge

- Non-covalent interaction between electrically charged atoms
 - Combination of H-bond and ionic bond
 - Strength depends on the charge, stronger than H-bond



By Chem540f09grp6 - Own work, Public Domain,
<https://commons.wikimedia.org/w/index.php?curid=8686974>

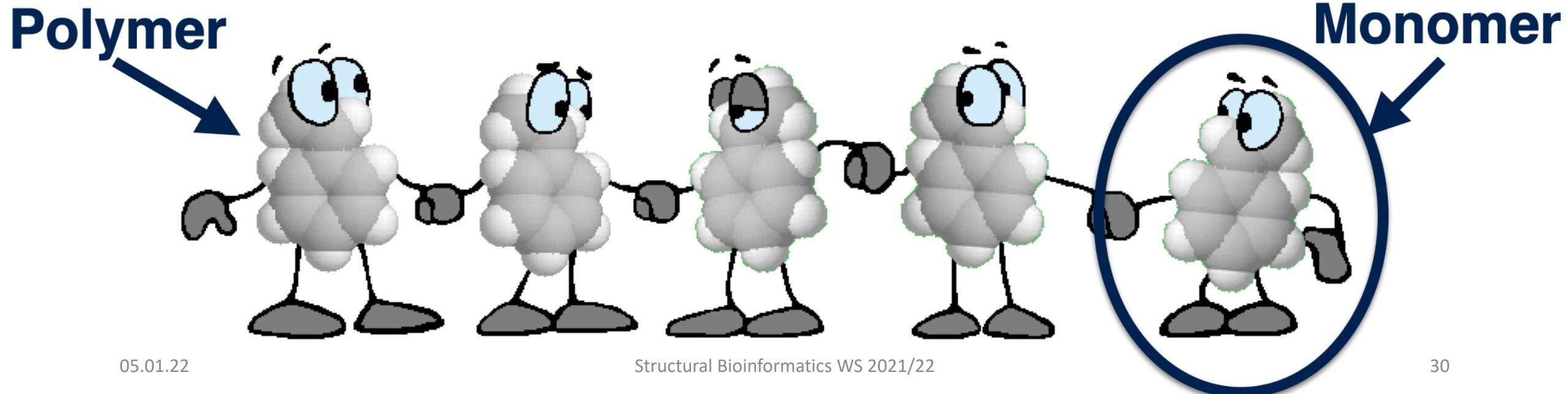
Van der Waals interactions

- Non-covalent interaction between (electrically neutral) atoms
- Very weak (much weaker than H-bonds), depend on the size of the atoms and the shape of the molecules
 - => short-range
- Non-directional
- Caused by various quantum properties of atoms
 - primarily by formation of temporary dipoles in else non-polar atoms

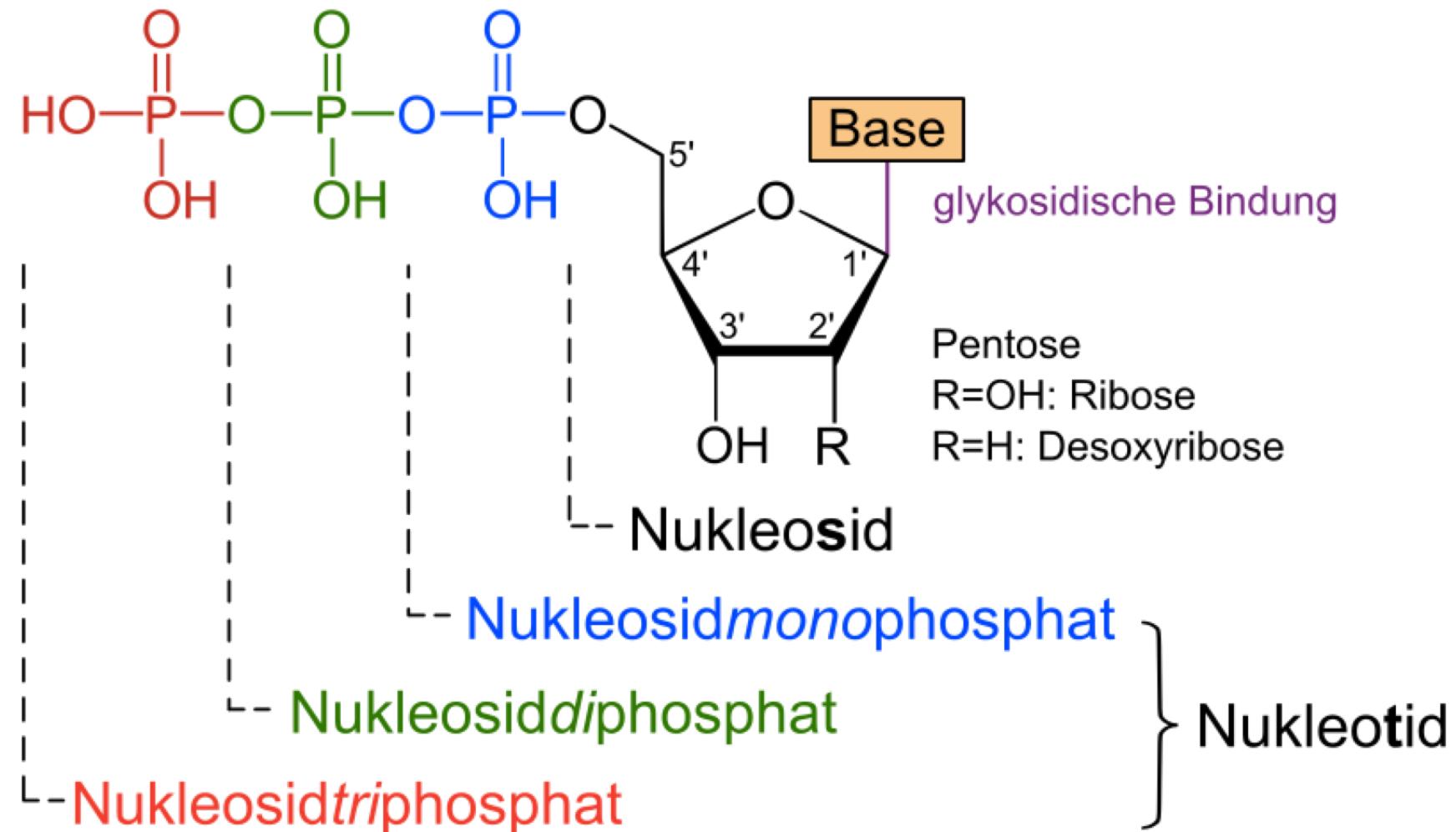
Chemical structure of nucleic acids

RNA and DNA

- **Polymers of nucleotides:** several similar monomers bound to each other
 - linear
 - non-periodic
- **DNA:** deoxyribonucleic acid — hereditary information
- **RNA:** ribonucleic acid — multiple (not entirely known) functions

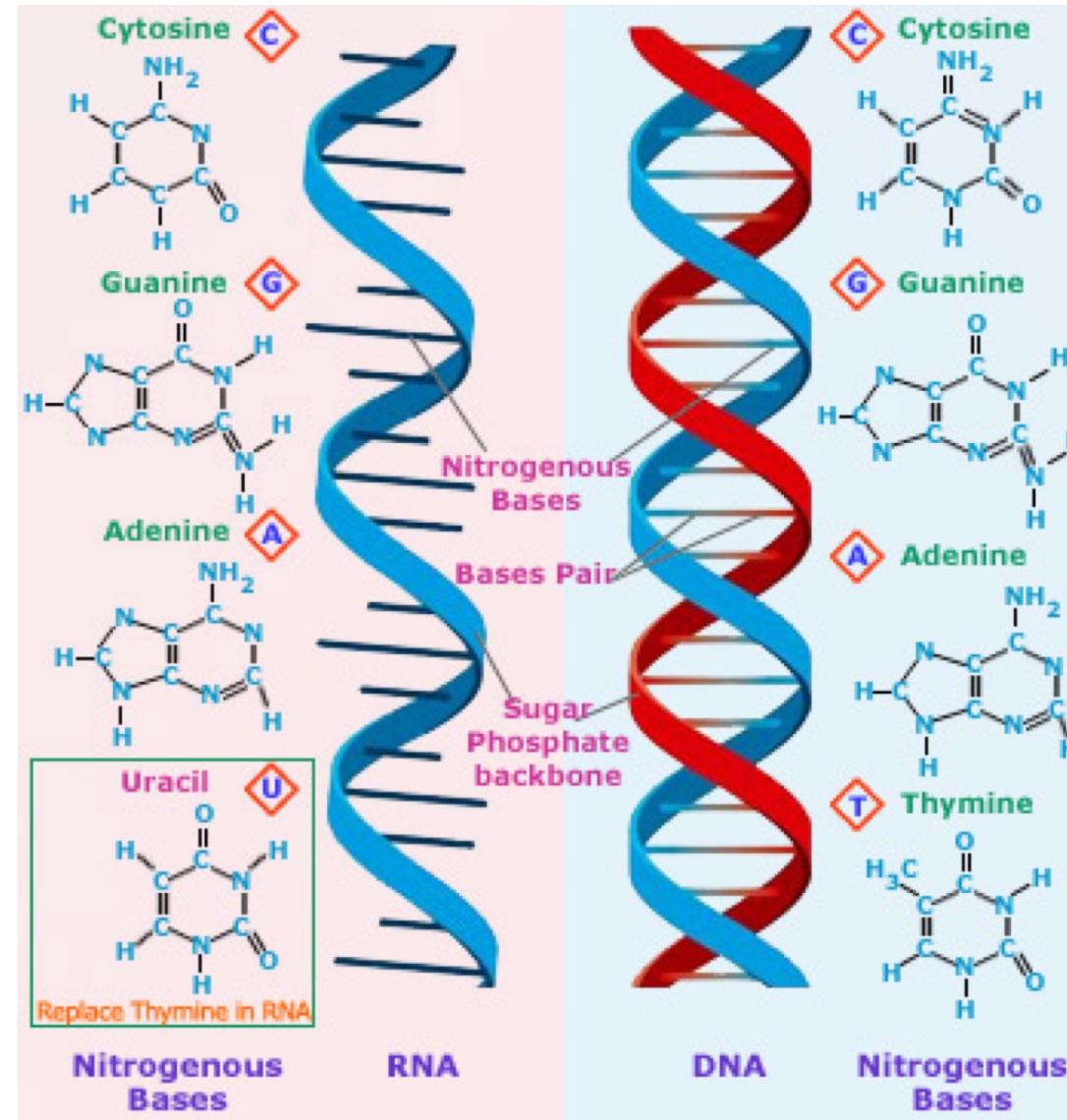


Nucleotide: nitrogenous base + sugar + phosphate



Bases in DNA and RNA

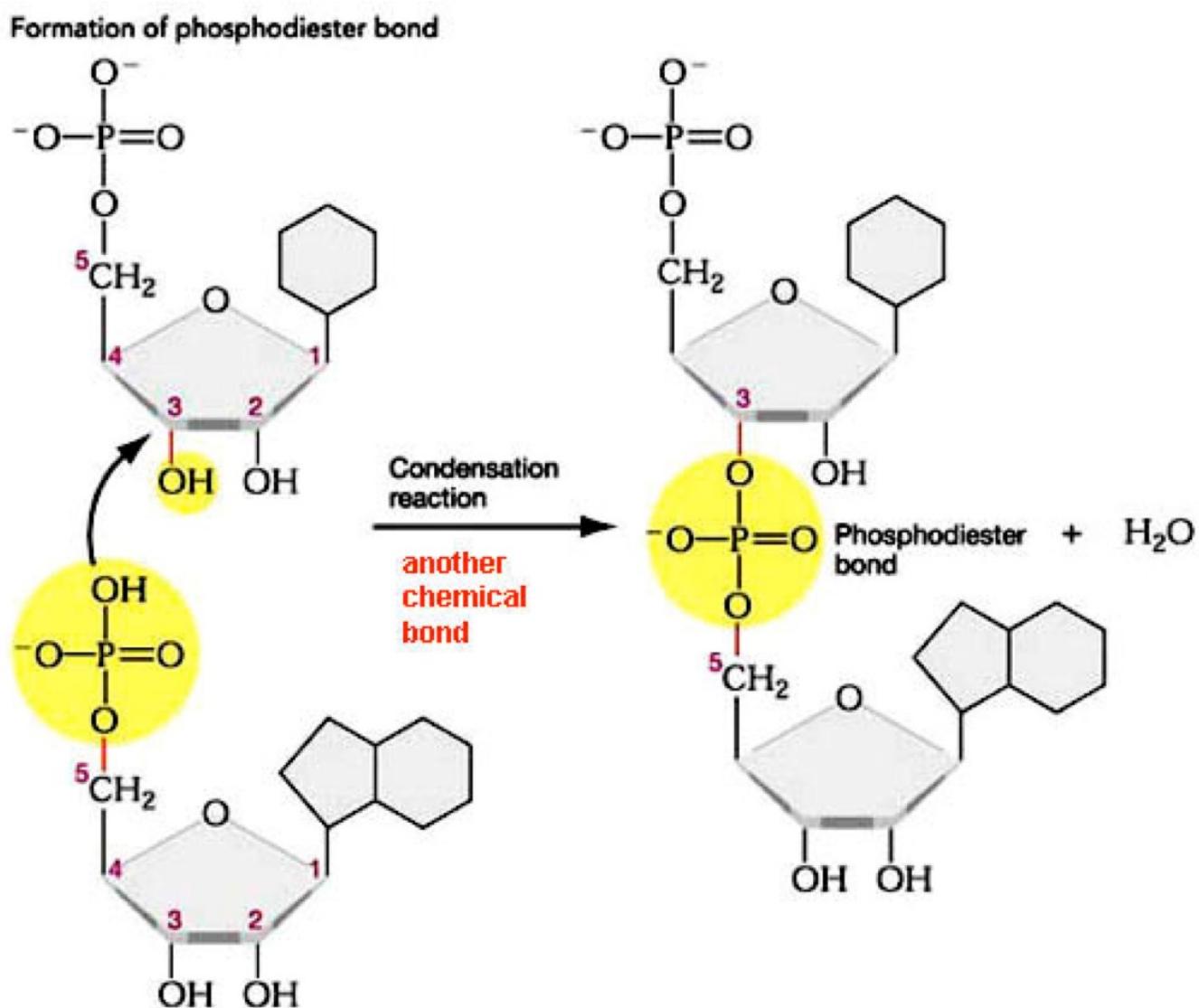
RNA:
Cytosine
Guanine
Adenine
Uracil



DNA:
Cytosine
Guanine
Adenine
Thymine

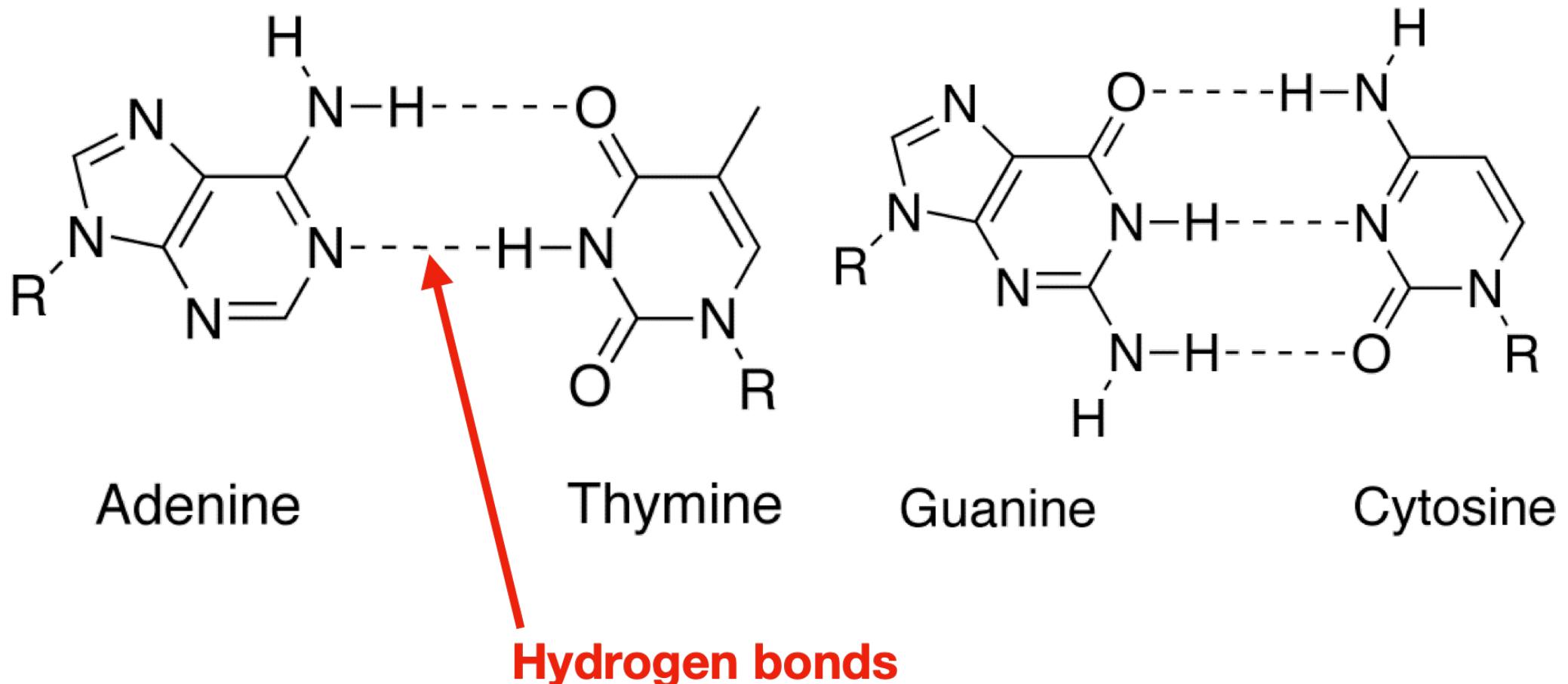
Phosphodiester bond

- binds nucleotides together to form DNA and RNA
- condensation (dehydration) reaction
- $5' \rightarrow 3'$ direction
- negatively charged



Base pairing

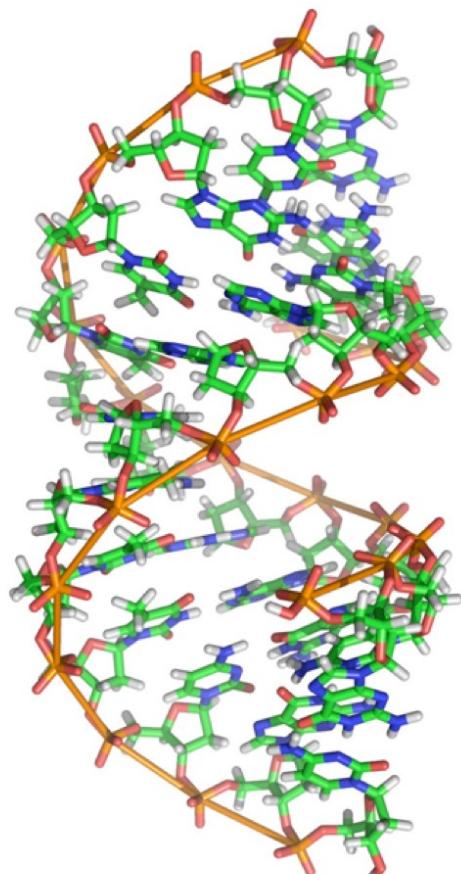
- Canonical (Watson-Crick): A-T(U), G-C



Forms of the DNA duplex (two-stranded structure)

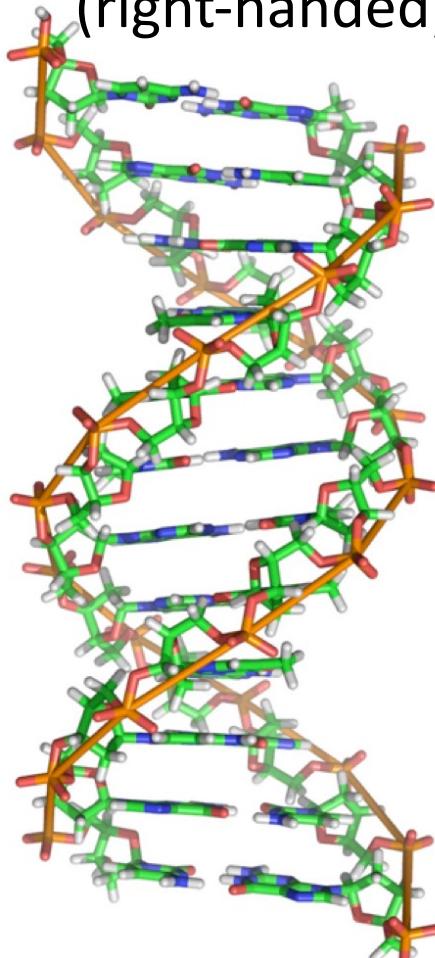
A-form

(right-handed)



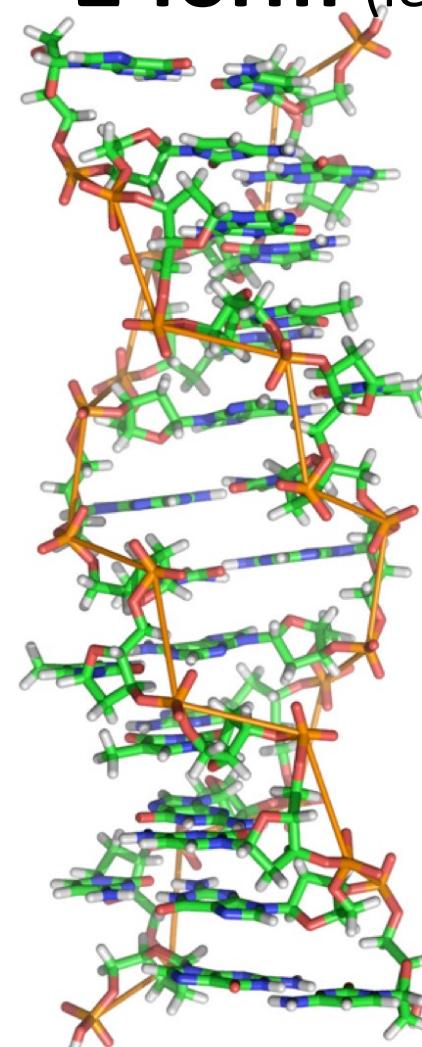
B-form

(right-handed)



Z-form

(left-handed)



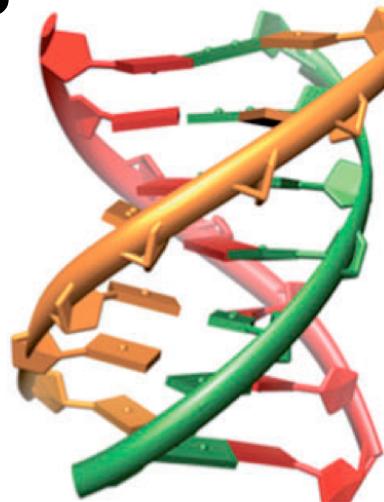
Functions of different forms of DNA duplex

- **B-form:** most common form of DNA
- **A-form:** occurs when DNA is dehydrated; adaptation to extreme desiccation or high temperature. Also more compact (can be important for packaging viral particles)
- **Z-form:** provides relief of torsional strain during transcription. Formation can be induced by supercoiling, salts, cations, specific proteins. Some sequence motifs are prone to form Z-DNA
- There are also C-DNA, E-DNA, P-DNA, S-DNA, etc., but they don't occur naturally

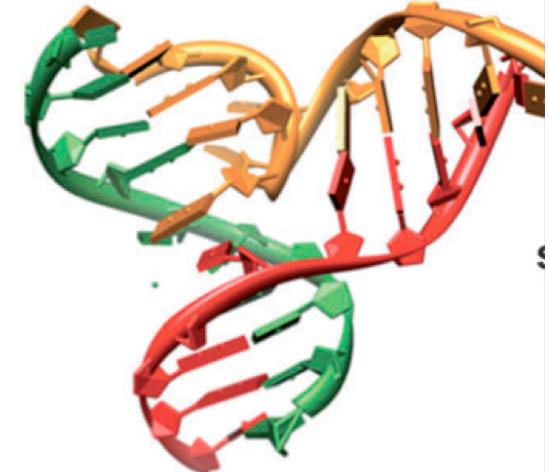
Higher-order DNA structures

- Can be implicated in gene regulation
- Can form during replication, transcription, reparation processes

three-stranded DNA:

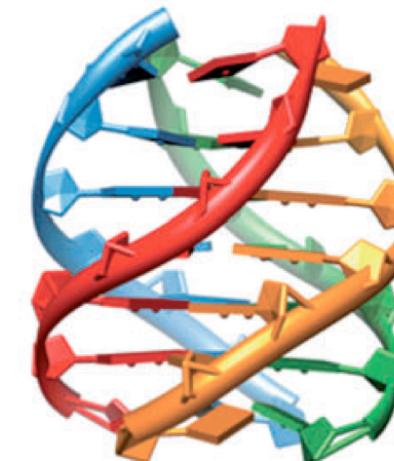


Triplex DNA



Three-way junction DNA

four-stranded DNA:



G-quadruplex DNA

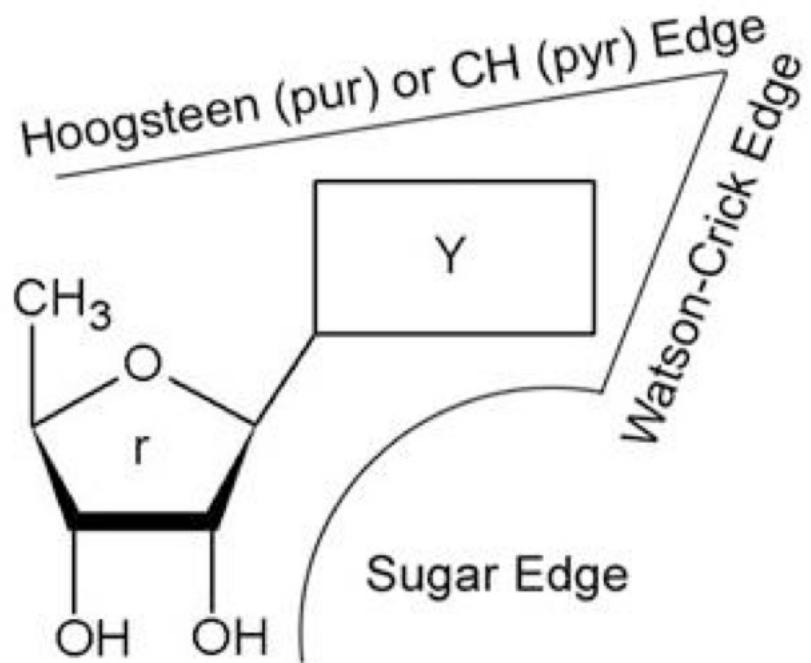


Four-way junction DNA

Non-canonical base pairing

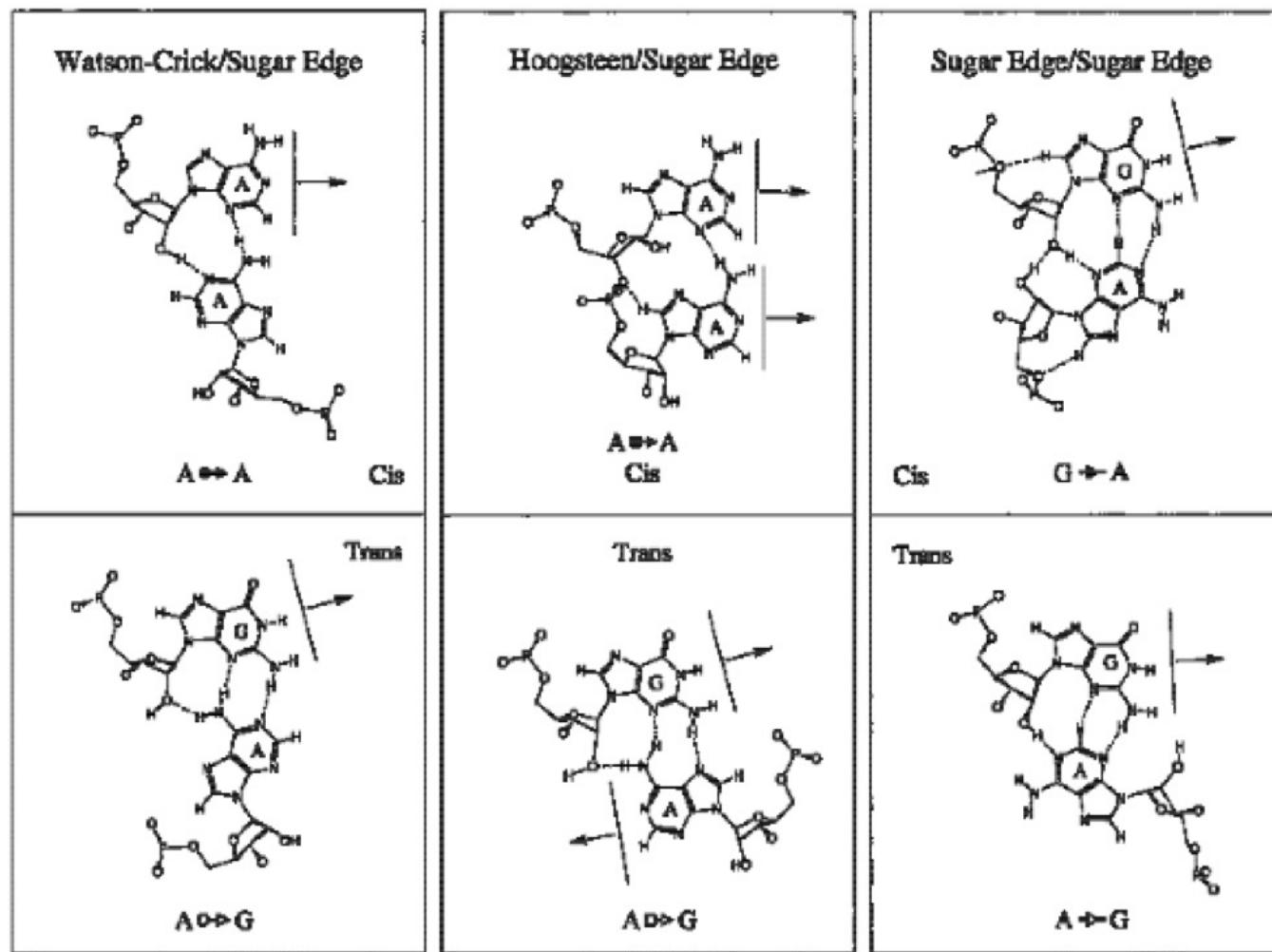
Almost all combinations are possible

Other edges of a nucleotide:

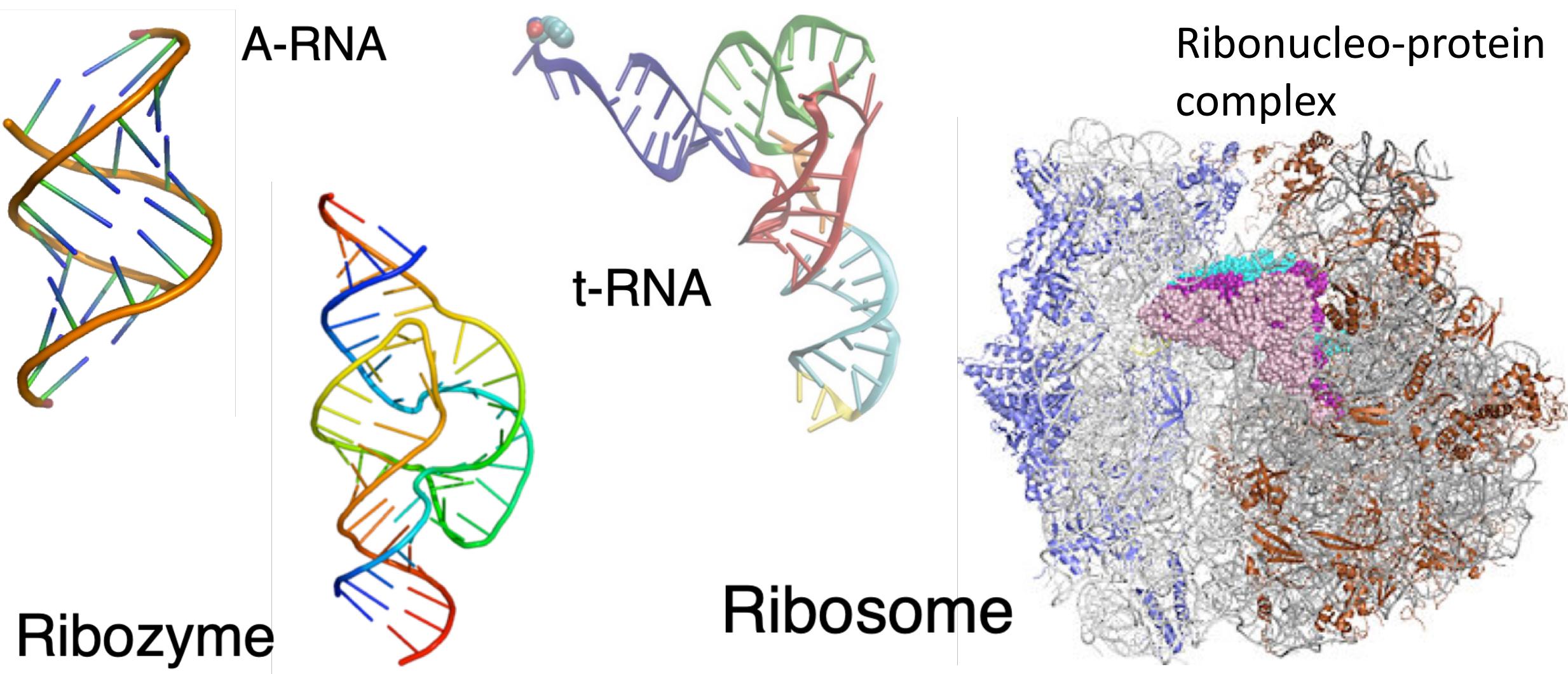


05.01.22

Structural E



RNA duplex and other 3D structures



Ribozyme

A-RNA

t-RNA

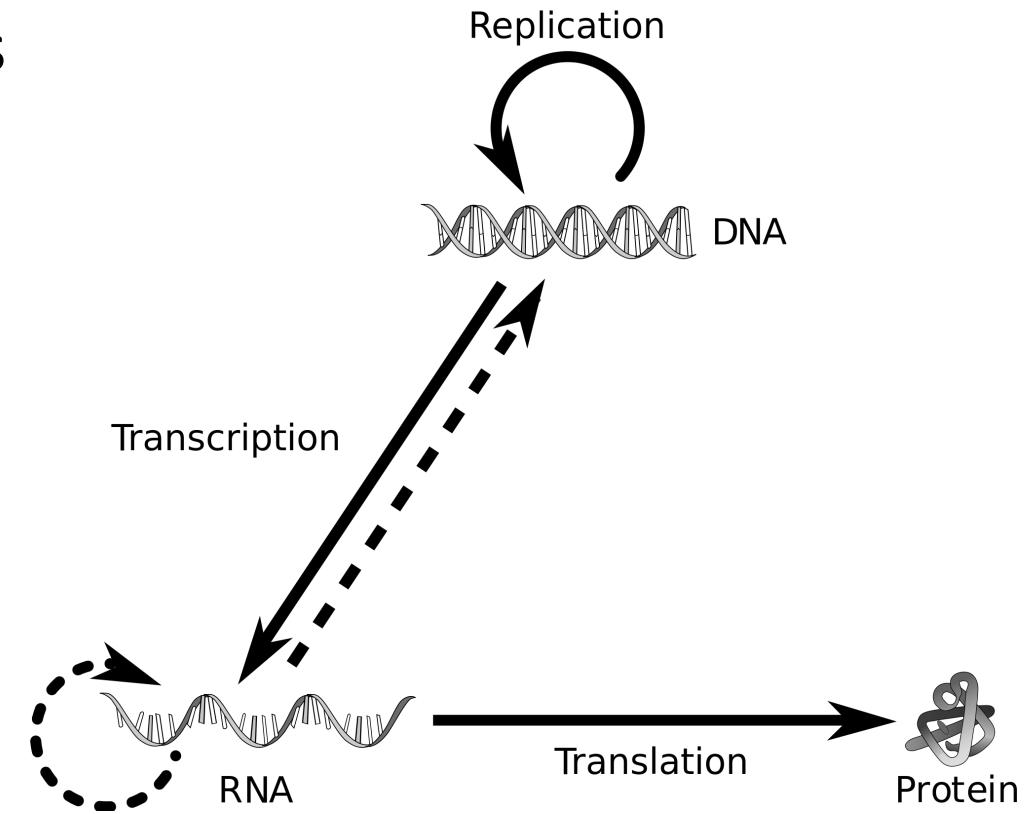
Ribosome

Ribonucleo-protein
complex

Chemical structure of proteins

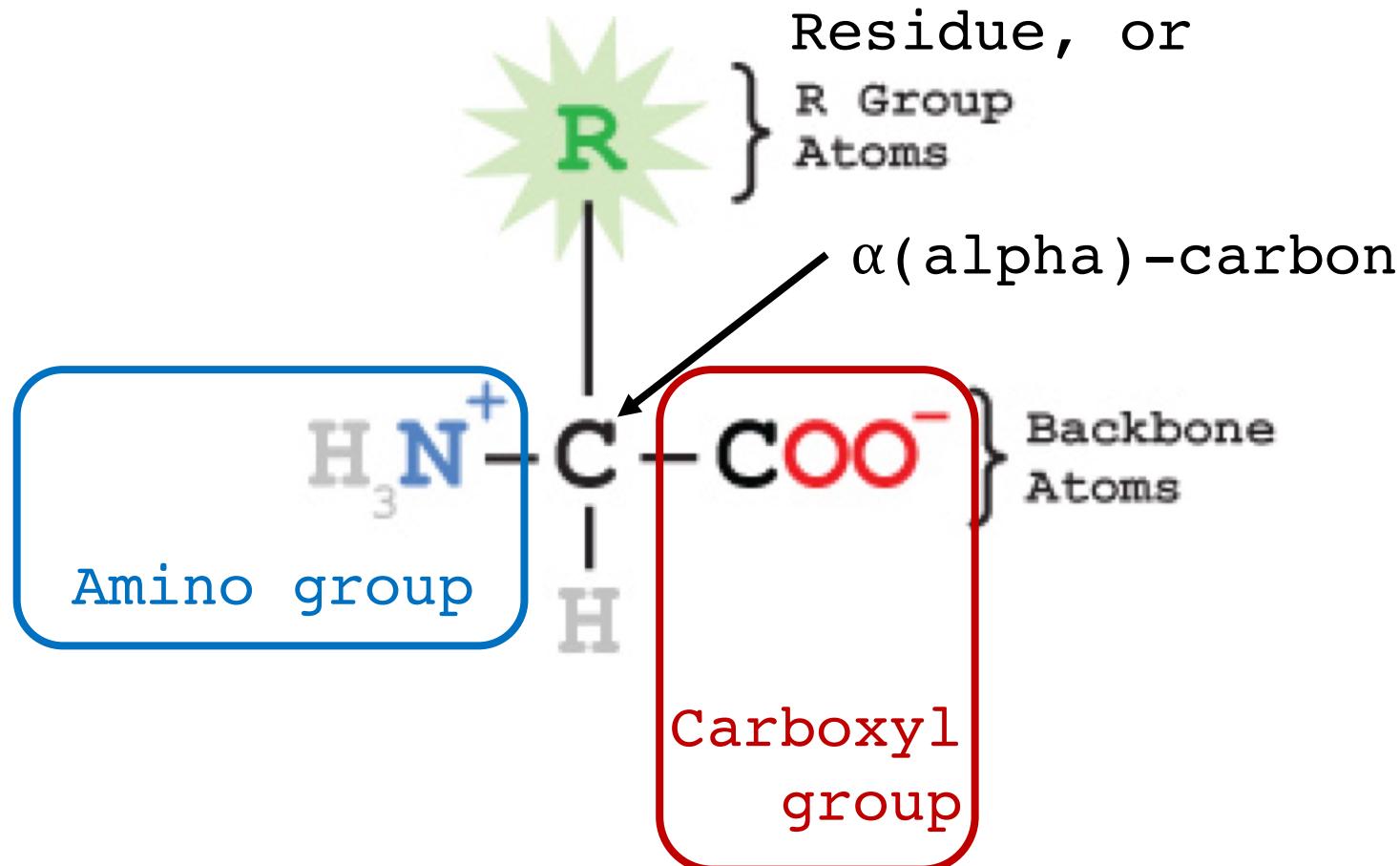
Proteins

- Like DNA/RNA, linear aperiodic polymers
- Chemically different
- Structure and function of any protein is dictated by its sequence of amino acids
- The sequence of amino acids is encoded in the sequence of nucleic acids in the genome



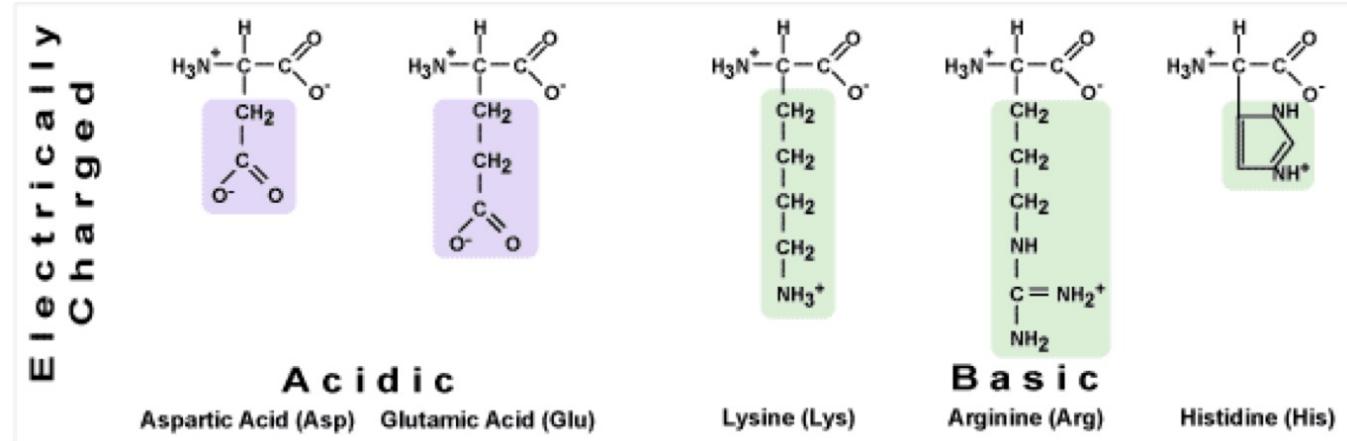
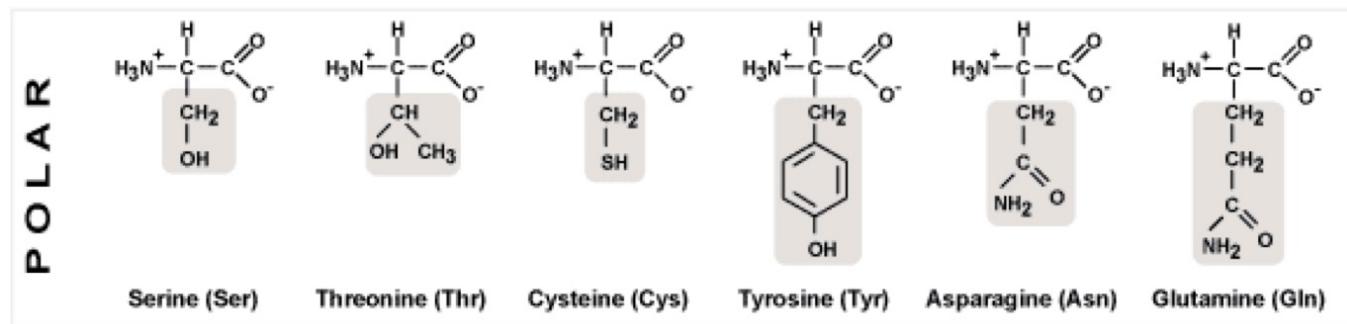
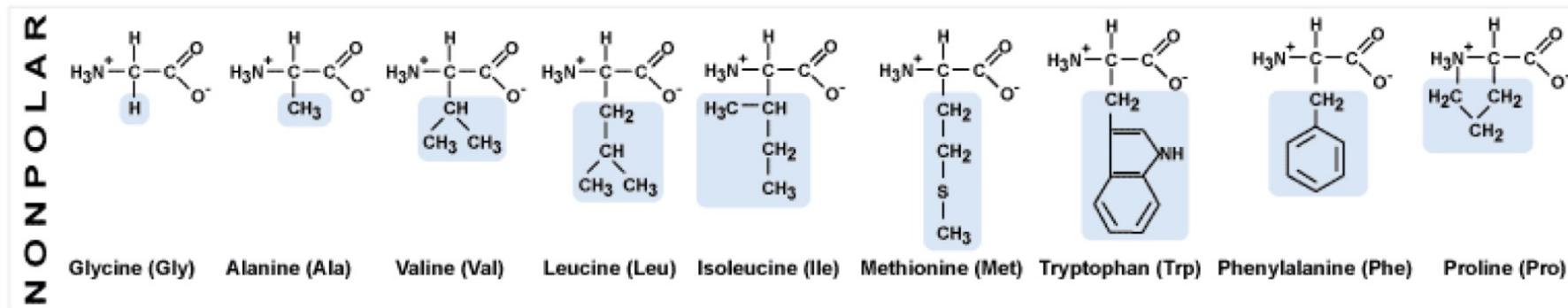
--The central dogma

Amino acid

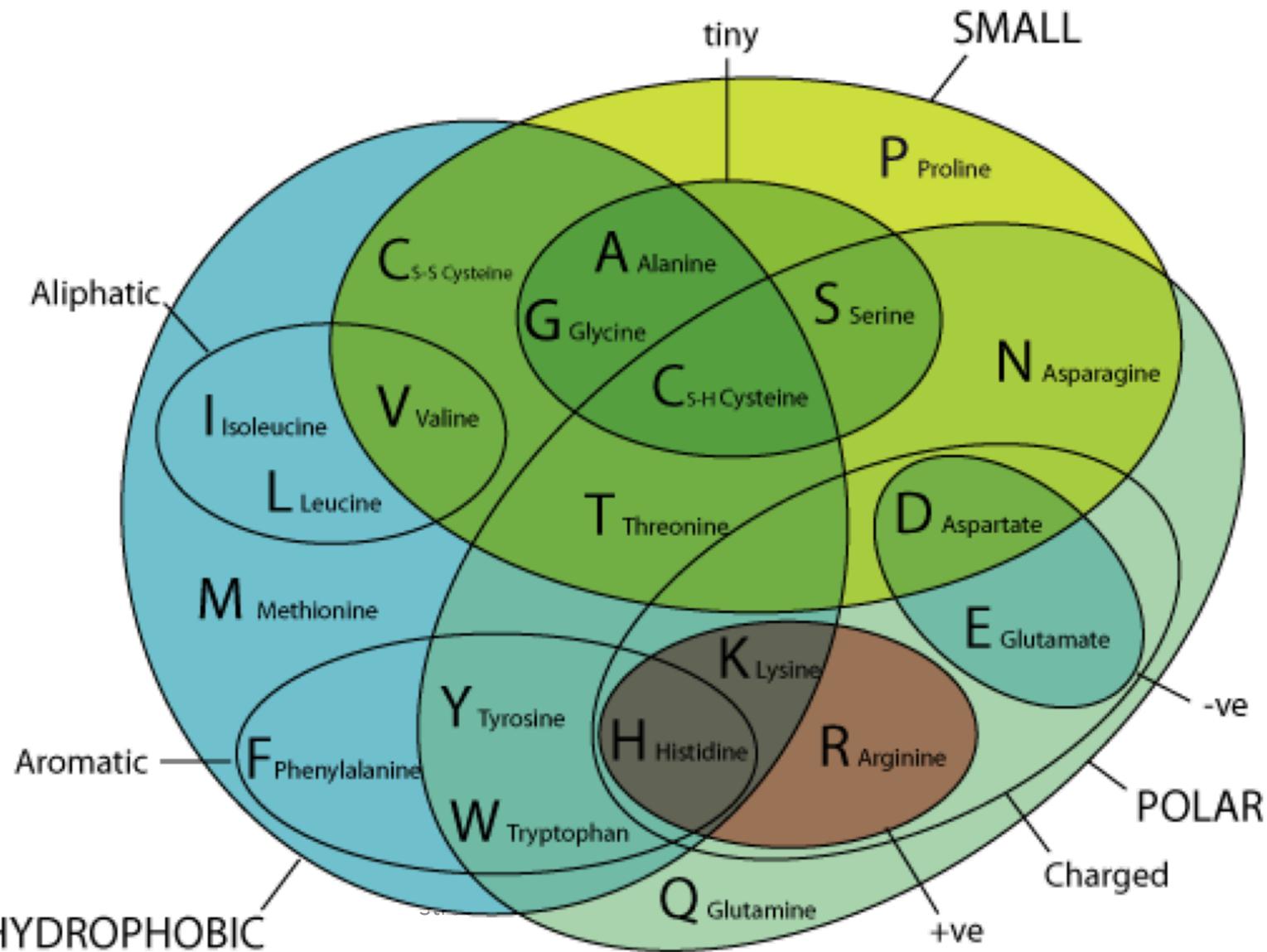


In proteins: 20 naturally occurring amino acids (+modifications)

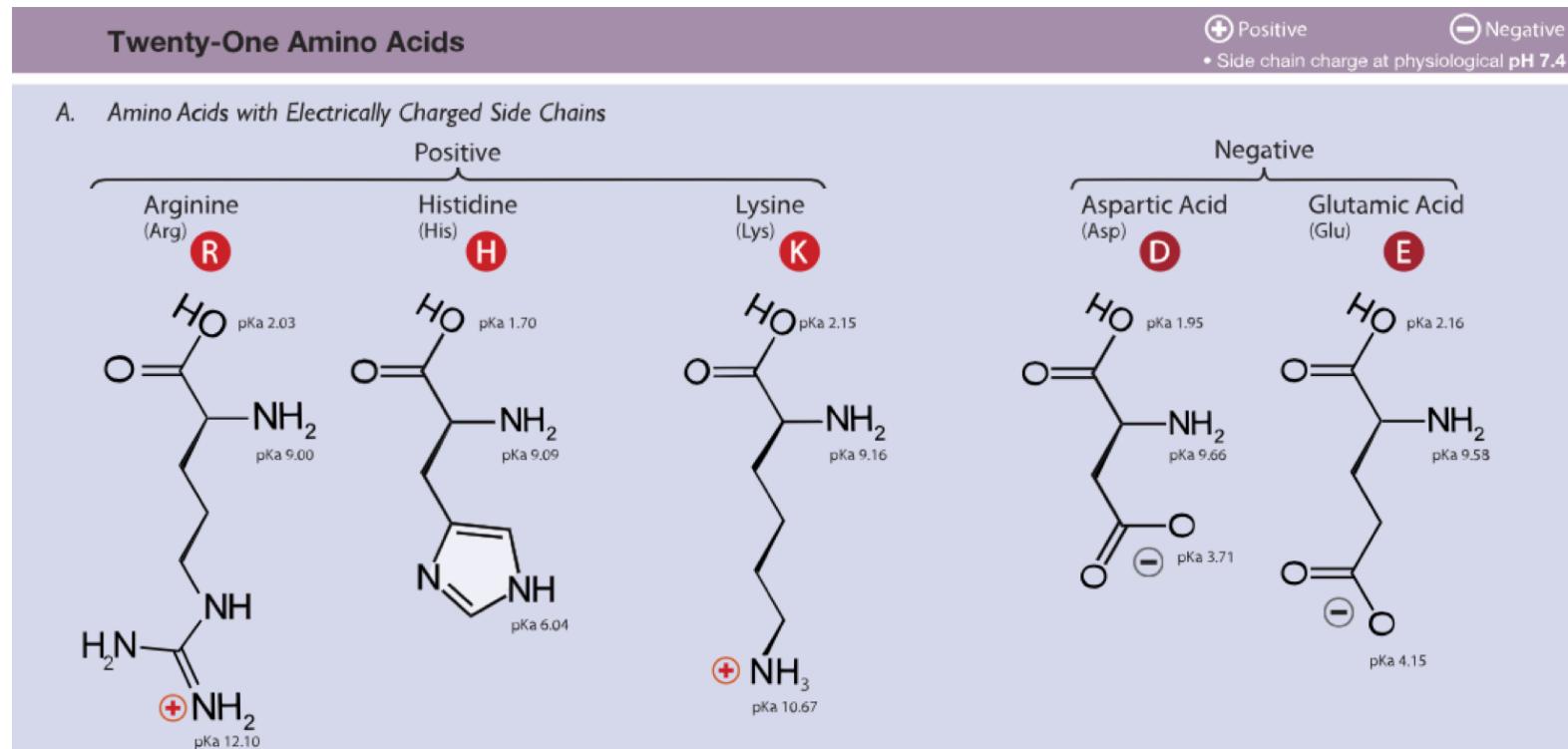
Amino acids have different chemical properties



Amino acids have different chemical properties



Chemically similar amino acids: charged



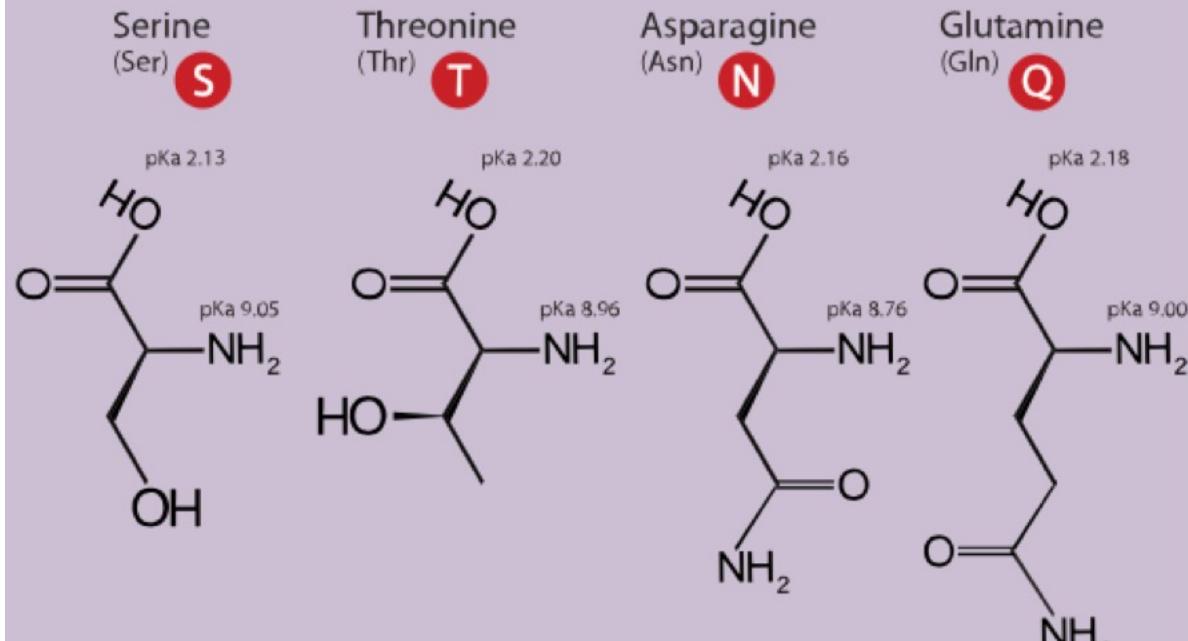
Positively charged

Negatively charged

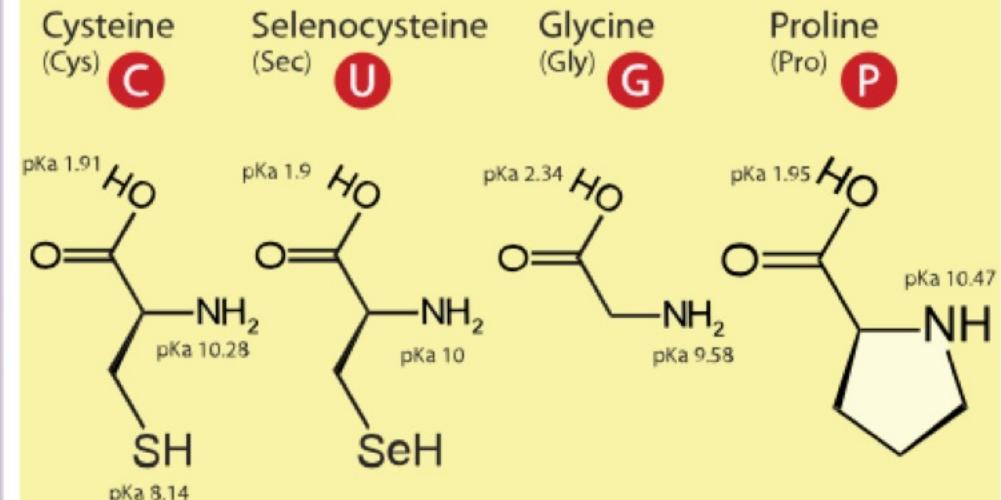
hydrophilic

Chemically similar amino acids: polar

B. Amino Acids with Polar Uncharged Side Chains

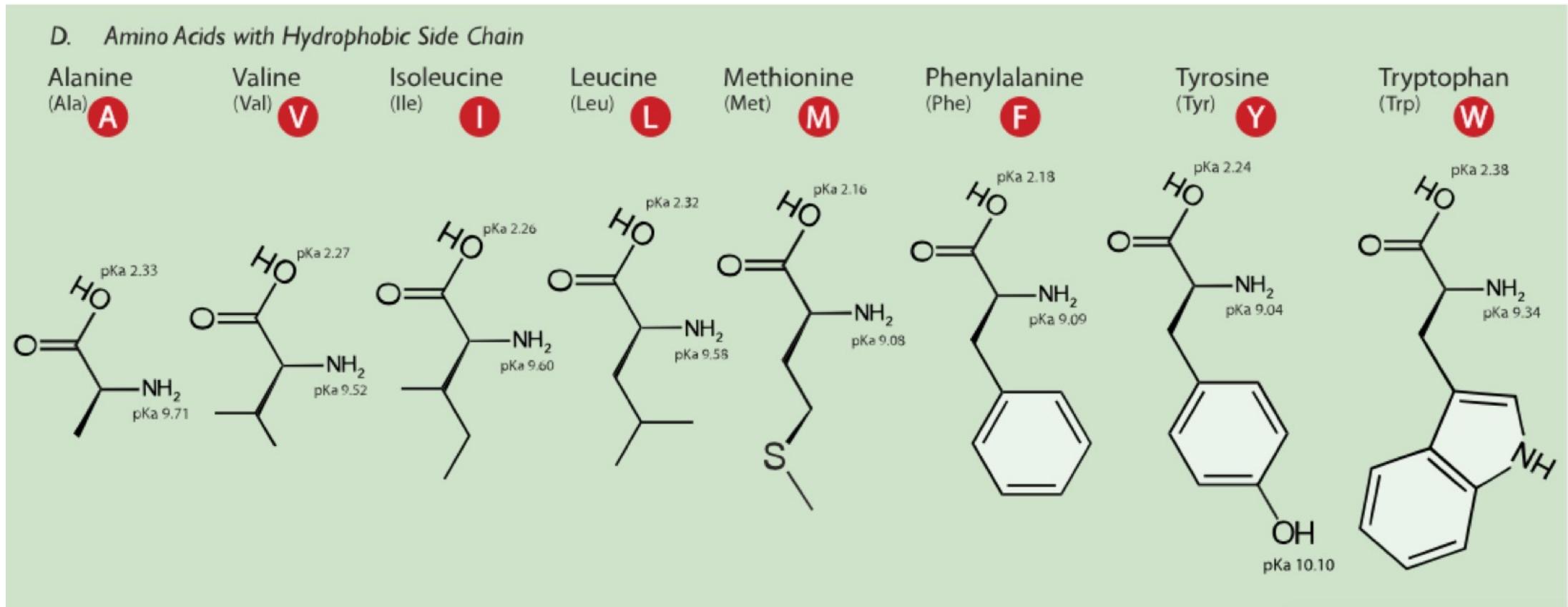


C. Special Cases



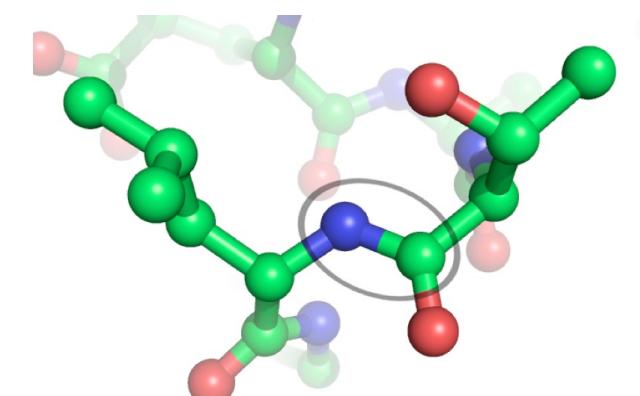
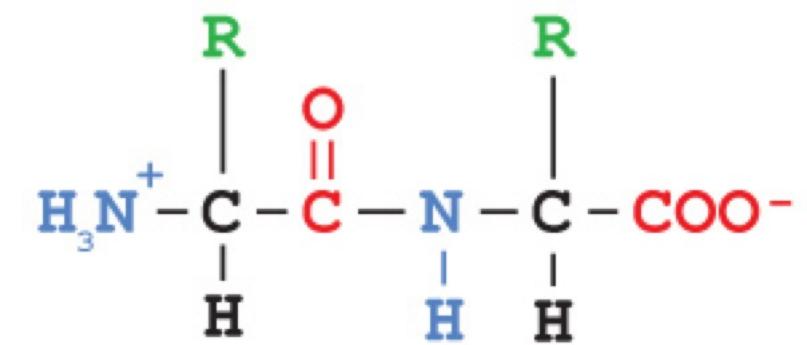
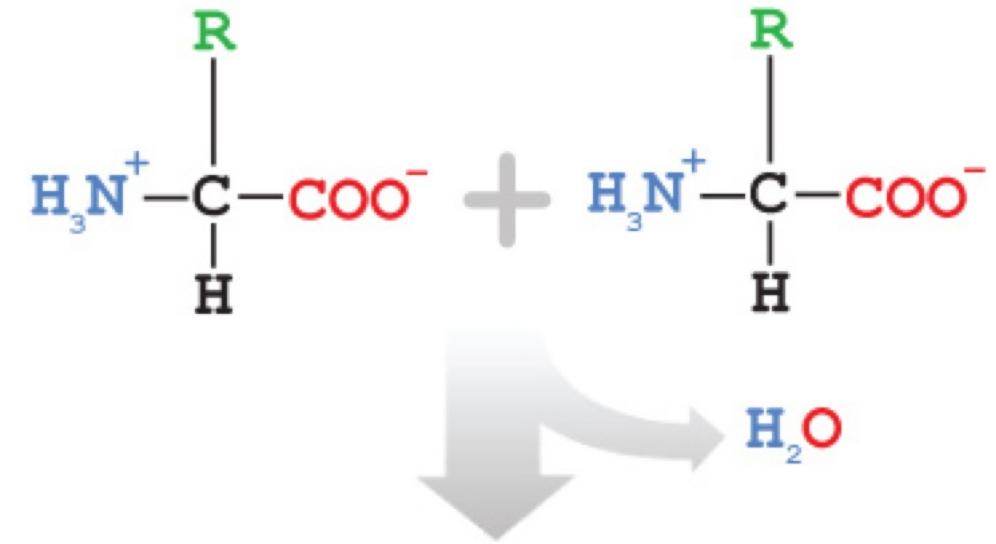
Also hydrophilic

Chemically similar amino acids: hydrophobic

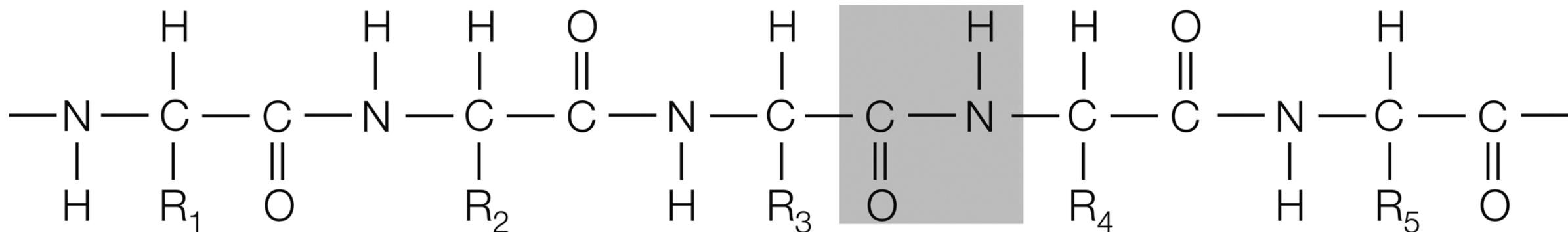


Peptide bond

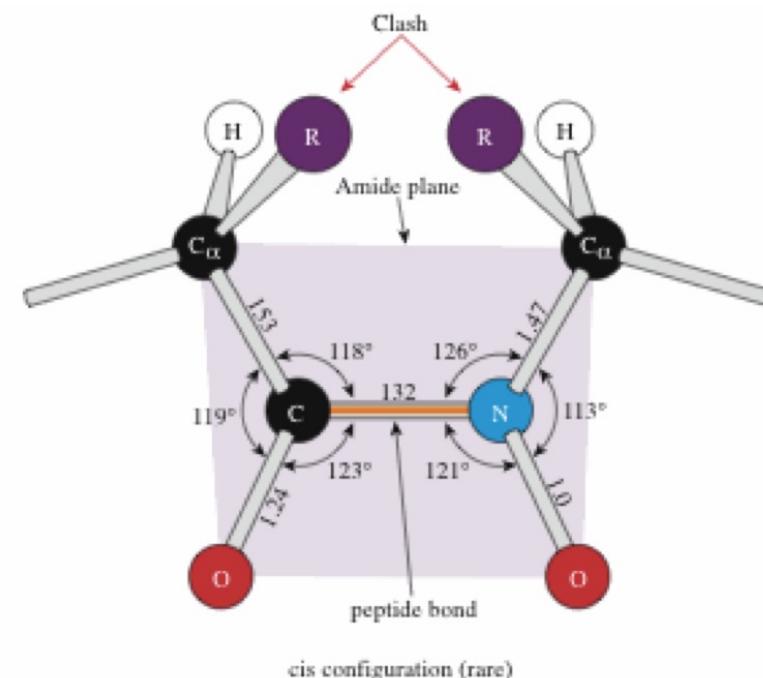
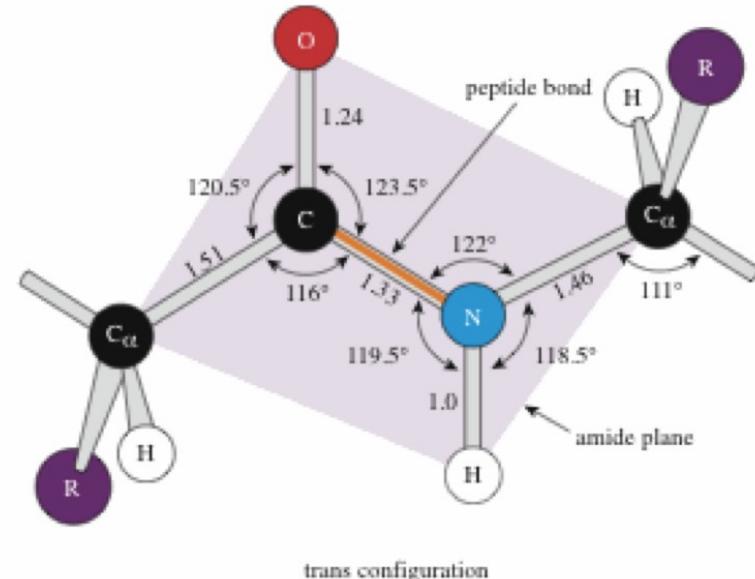
- binds amino acids to form a protein (**polypeptide**)
- condensation (dehydration) reaction
- between carboxyl and amino group of two amino acids
- Direction: N- to C-terminus



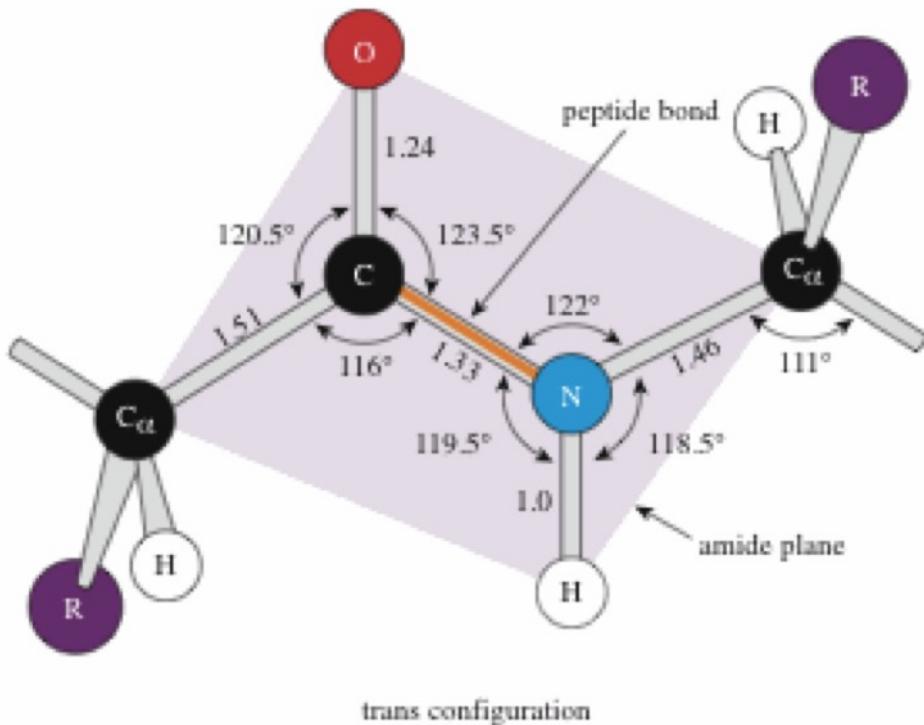
Peptide bond



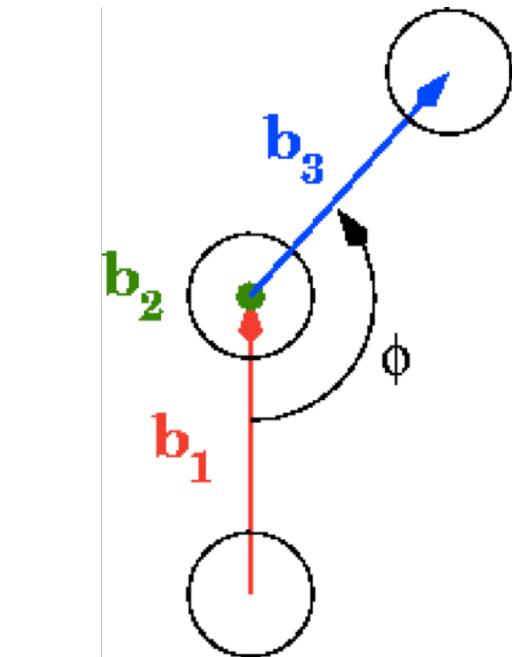
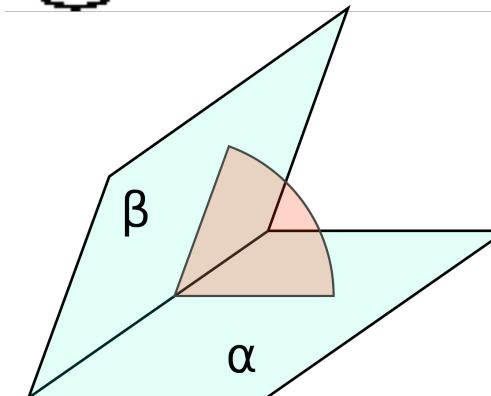
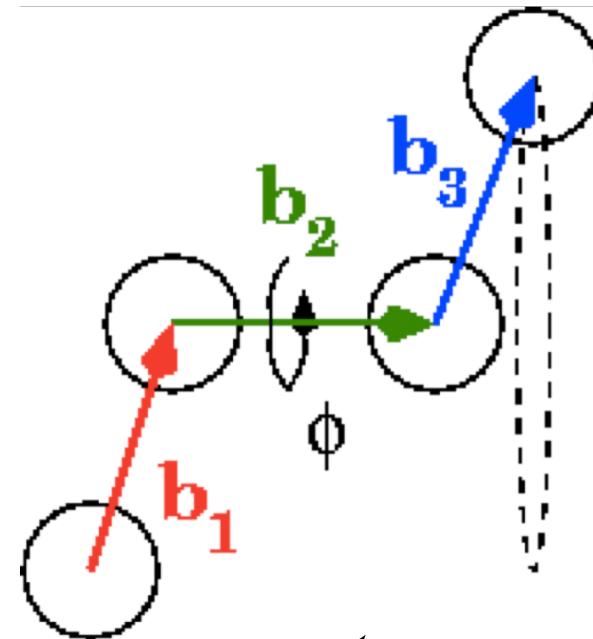
Planar!



Dihedral angles (a.k.a. torsion angles)



trans configuration



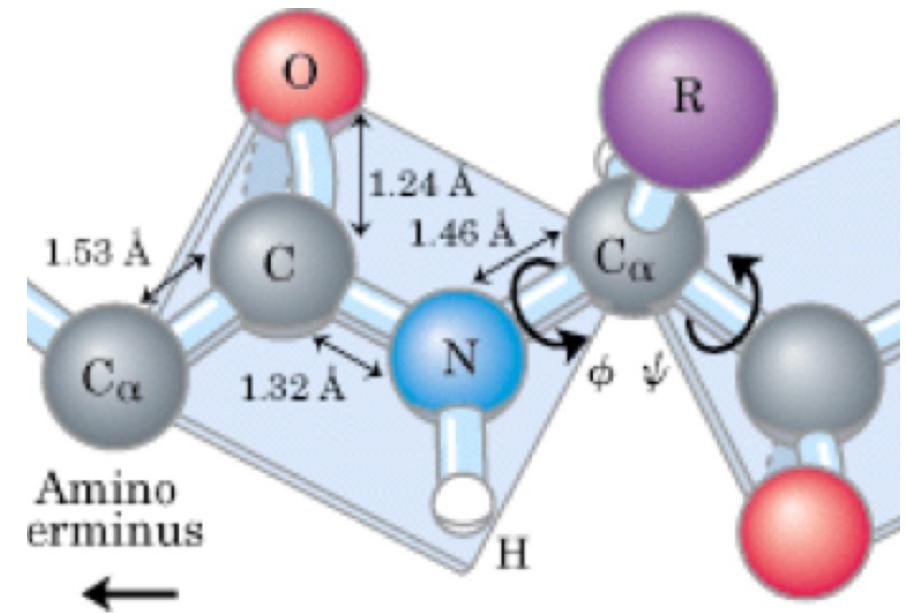
Dihedral angle:
general
mathematical
term

Dihedral angles (a.k.a. torsion angles)

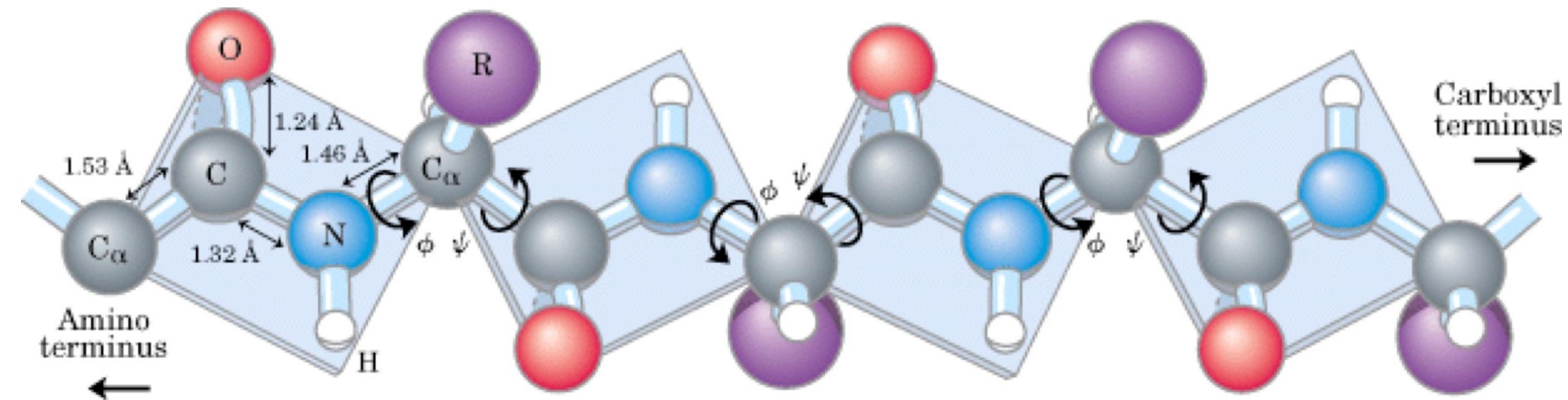
- Peptide bond is planar => rotation possible around N-C_α (ϕ) and C_α-C (ψ) bonds
- Not all angle values allowed due to steric constraints imposed by the residues



Psi- and phi-type figurines,
1450-1100 BC,
Mycenaean Greece

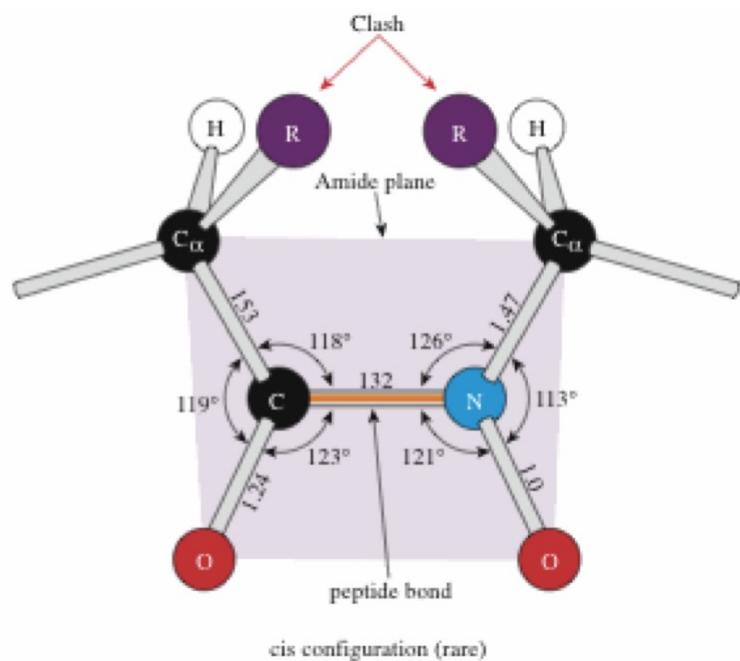


Dihedral angles (a.k.a. torsion angles)

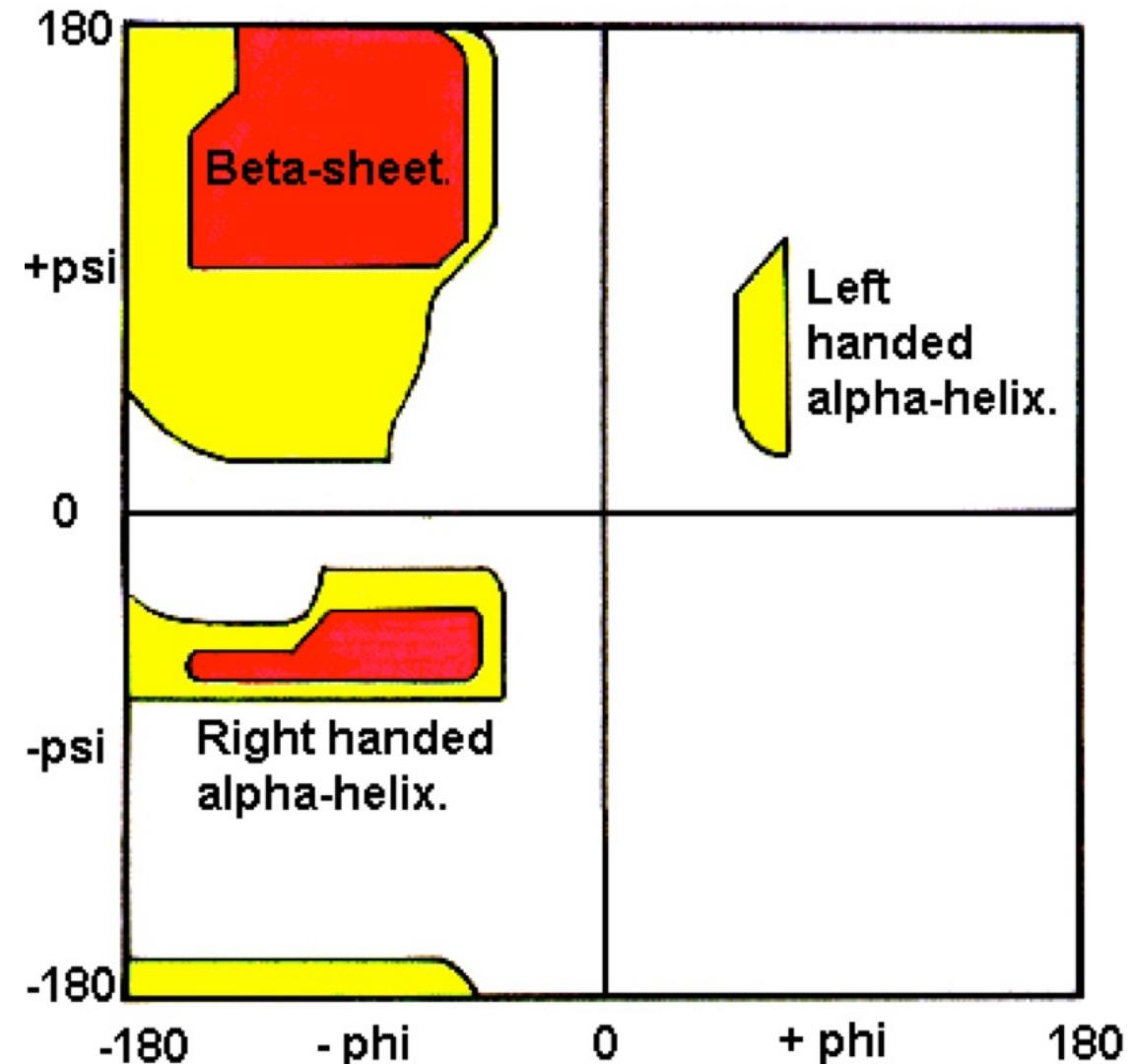


Ramachandran plot

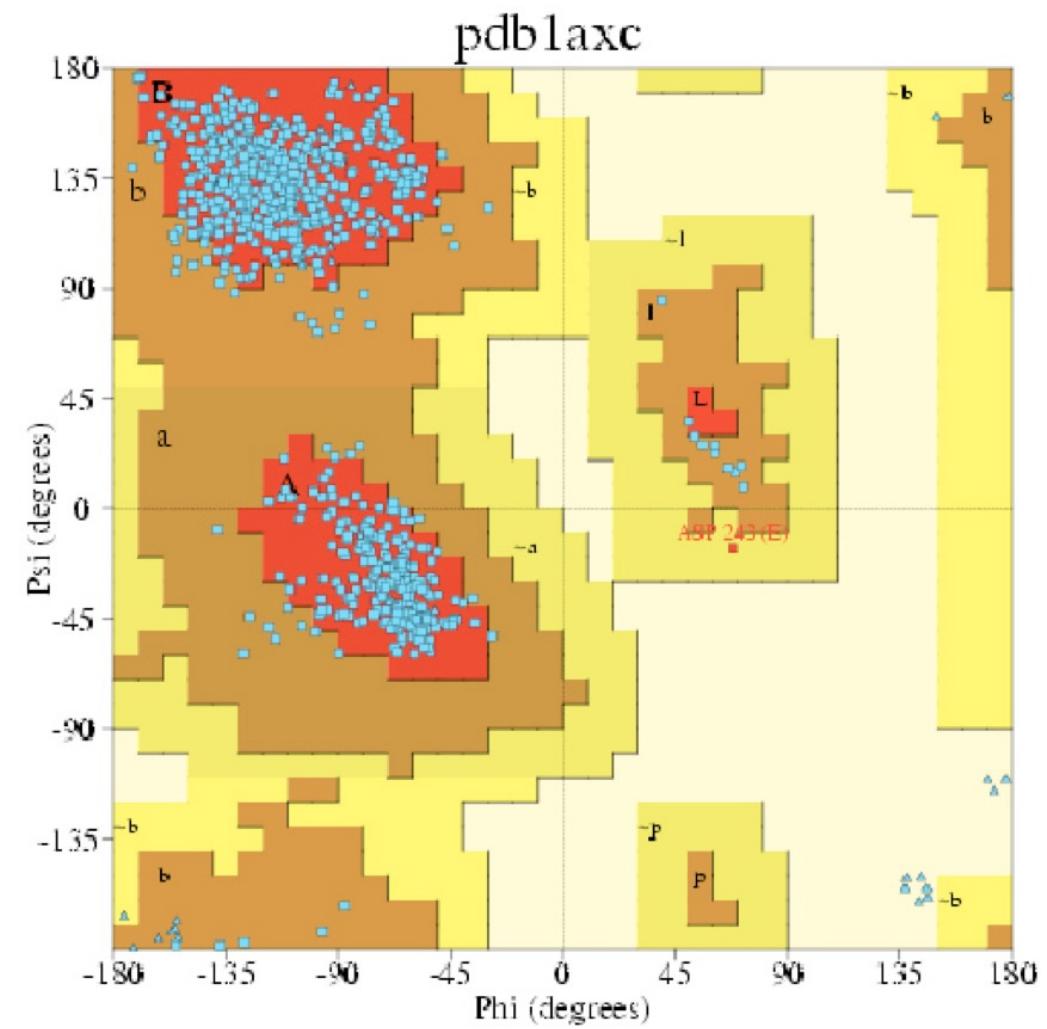
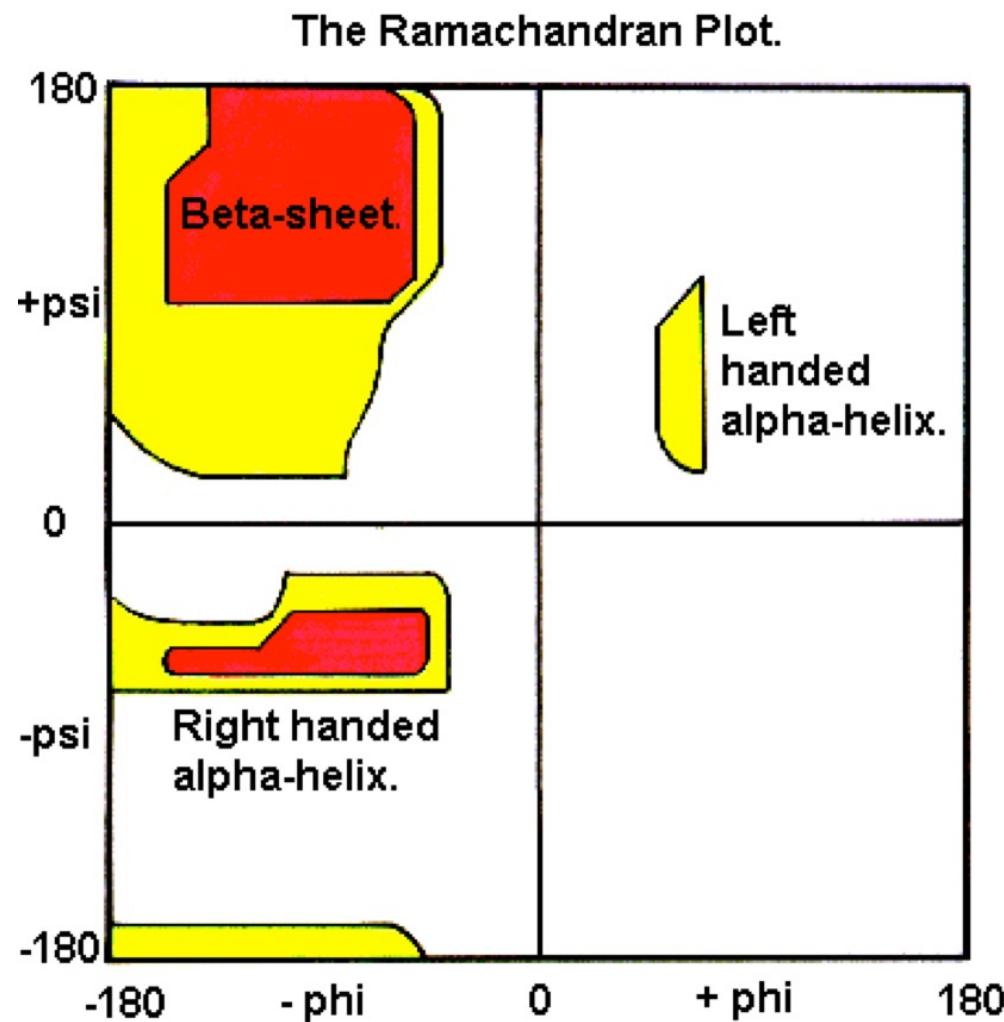
- Plots φ vs. ψ angles
- Amino acid residues and secondary structure elements impose constraints



The Ramachandran Plot.



Ramachandran plot



3D structure of nucleic acids vs. proteins

	DNA	RNA	Proteins
	Linear aperiodic polymers		
Monomer	Nucleotide		Amino acid
Bond	Phosphodiester		Peptide
Function	Storage of genetic information	Various (one of most important: information transfer)	Various (most catalysts in cells are proteins)
Chemical properties of monomers	...not so much important for function	...may be important for function	...very important for function
3D structure	Mostly double-stranded helix	Various, depending on function	Complex and unique

Possible exam questions

- What chemical atoms can appear in biological macromolecules?
- What types of chemical bond can appear in biological macromolecules?
- What nitrogenous bases occur in DNA and RNA?
- What are the important parts of an amino acid?
- How many amino acids occur in proteins?
- How is the peptide bond formed?
- What are the dihedral angles in protein chains?
- Why are the values of dihedral angles constrained?