

BT 305: Interfaces of Biology with

1. Computer Science
2. Nanotechnology
3. Drug Discovery
4. Materials Science
5. Physics
6. Chemistry

The Target

The Drug

The Interaction

Why are these animals having unique identities?

Why are biological machines so robust?

What is a Disease, a Drug ?

What is the basis of Drug Discovery



GENOME

↓ **Transcription**

TRANSCRIPTOME

RNA copies of the active protein-coding genes

↓ **Translation**

PROTEOME

The cell's repertoire of proteins

Proteins !!!

- Proteins are biological macromolecules.
- Proteins make up about 15% of the mass of the average person
- Smaller protein molecules play a vital role in keeping our body working properly.
- Hormones, antibodies , and enzymes are all proteins !!!!!
- Proteins are essential parts of organisms and participate in every process within cells

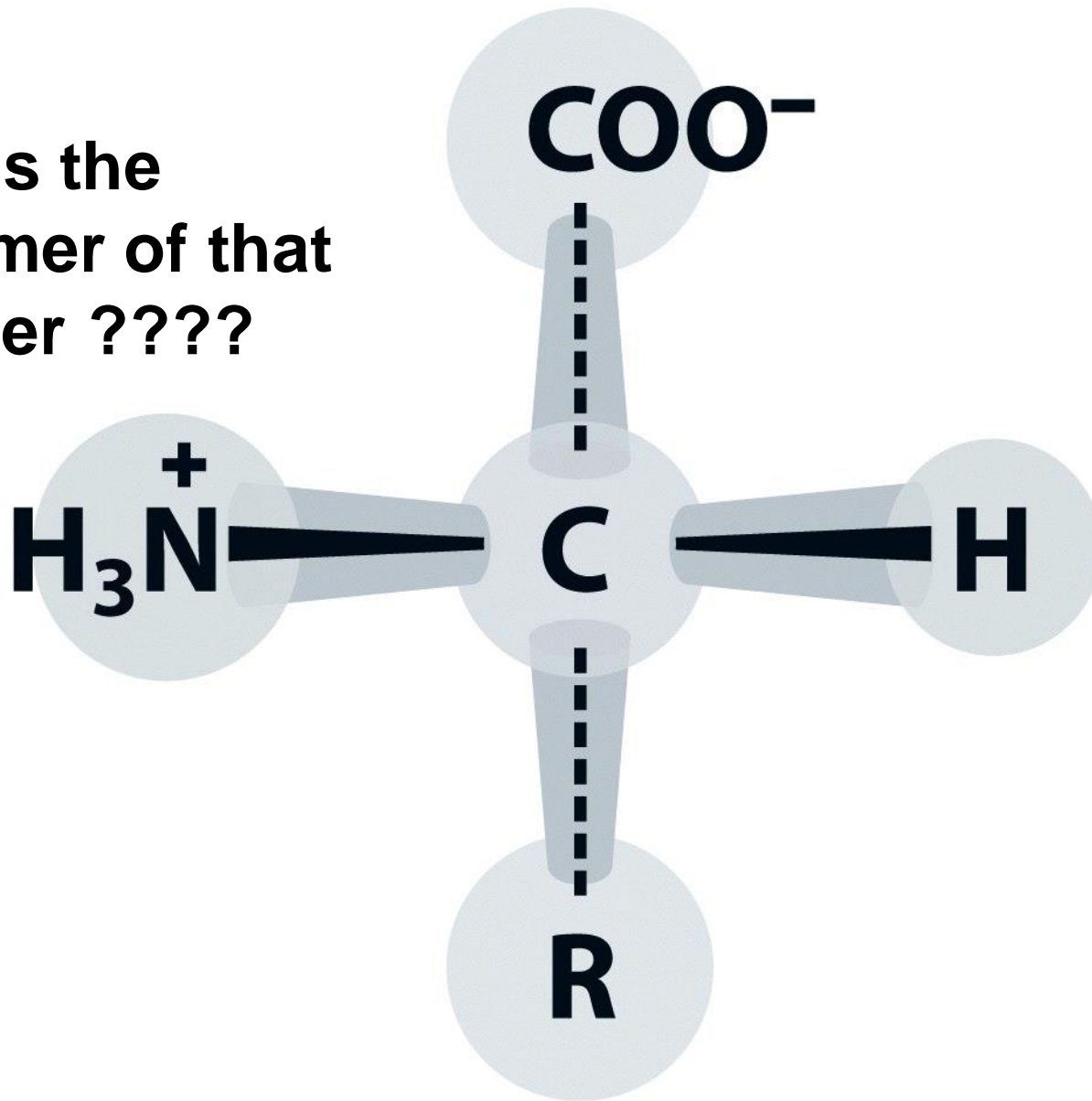


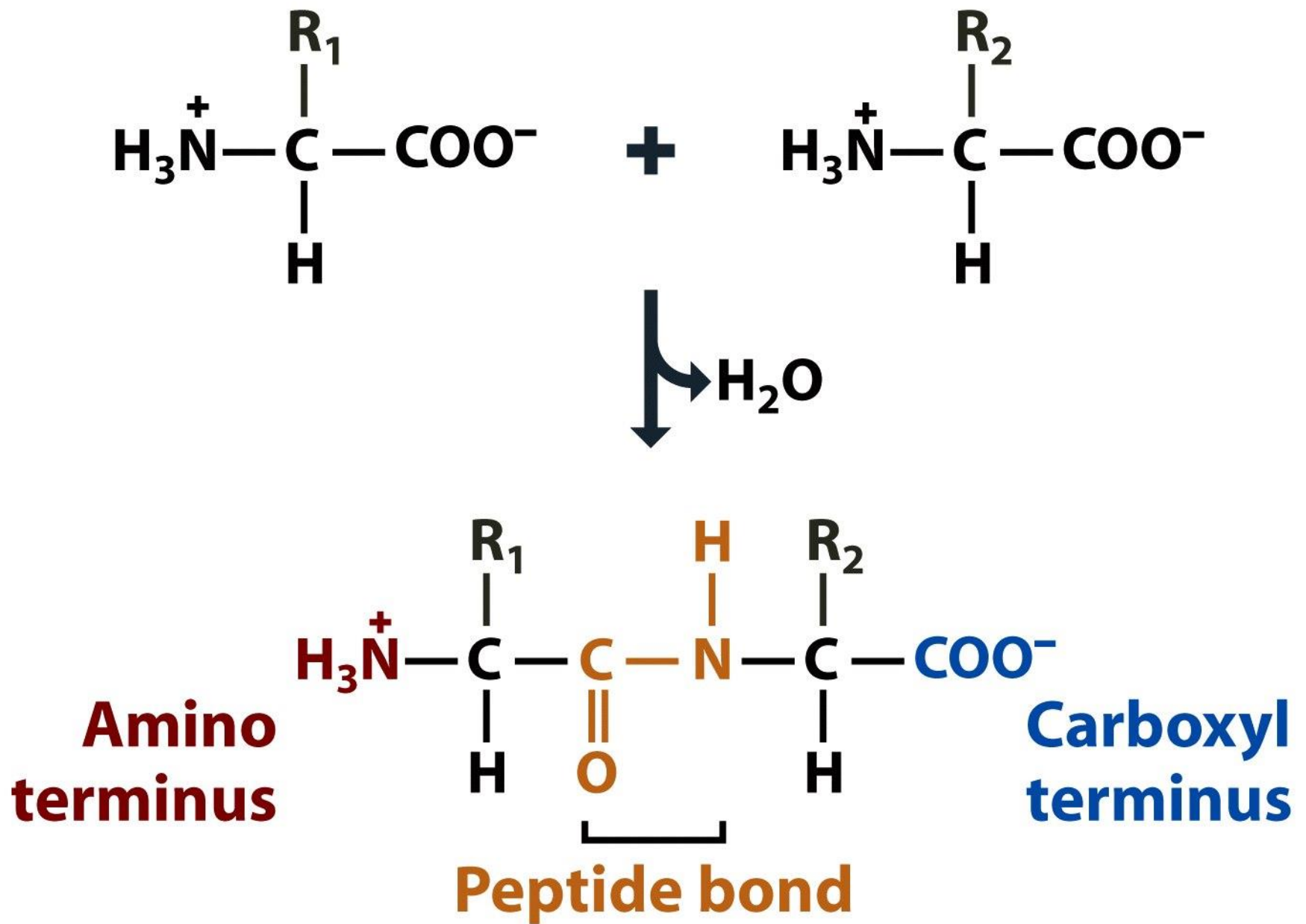
Myoglobin

Protein is a polymer

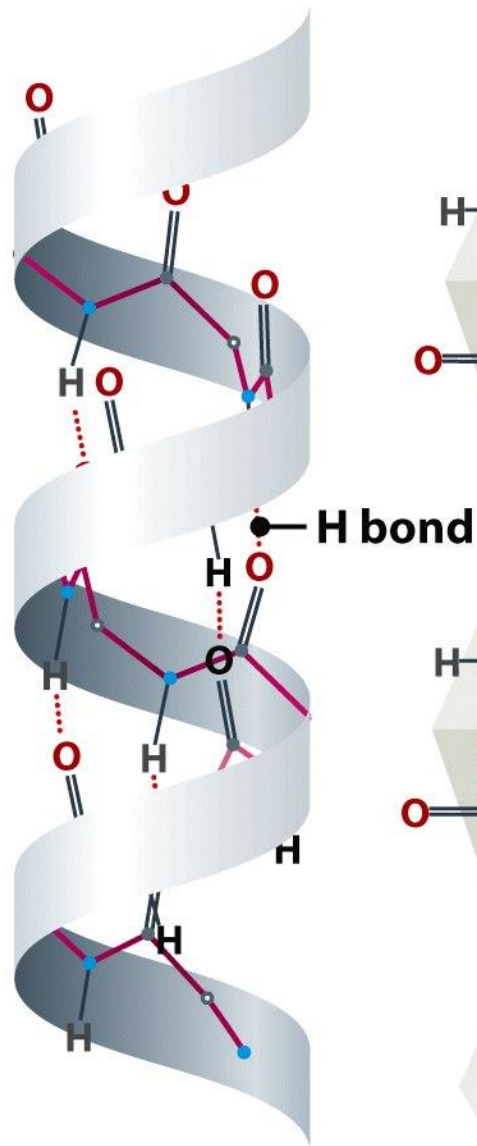
A HETERO POLYMER

**Then:
What is the
Monomer of that
Polymer ?????**

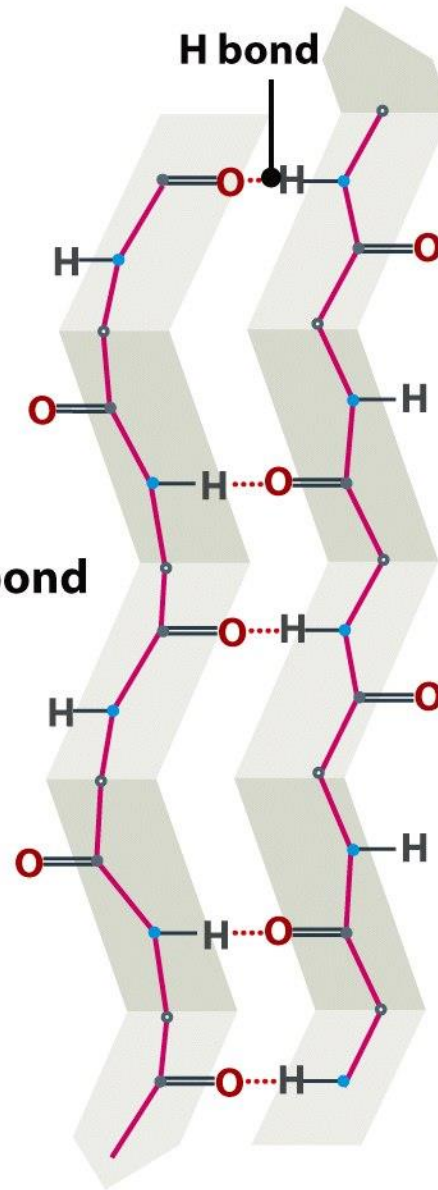




(A) α -helix



(B) β -sheet



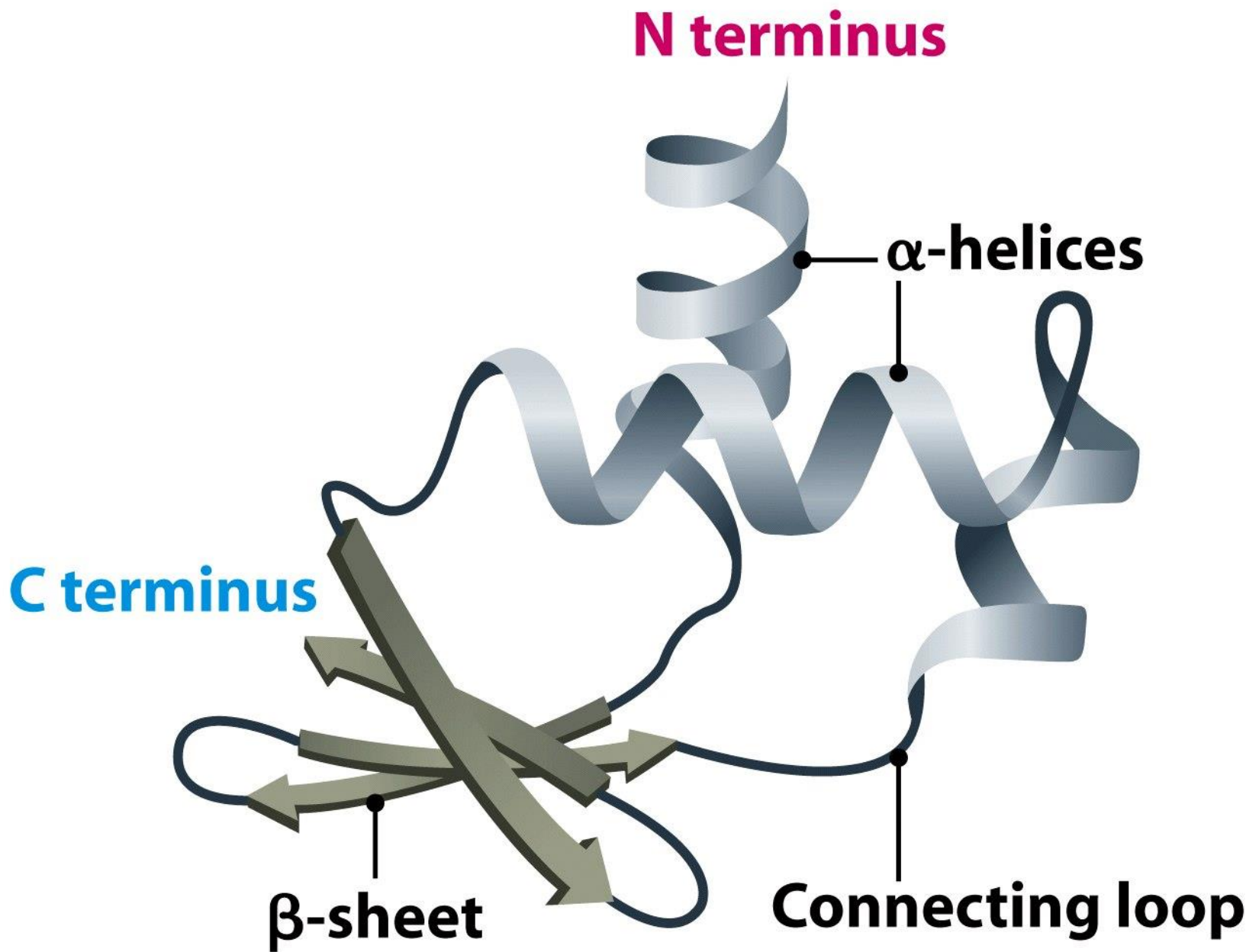
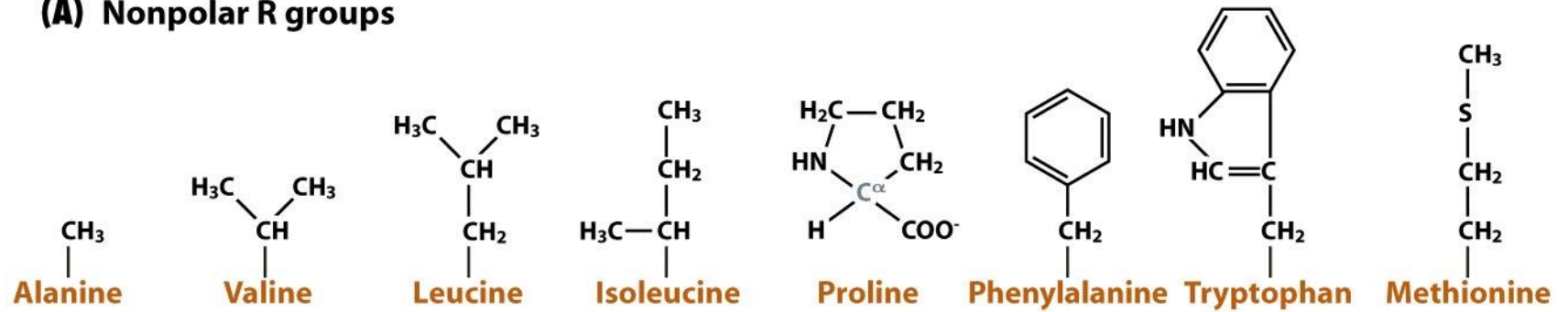
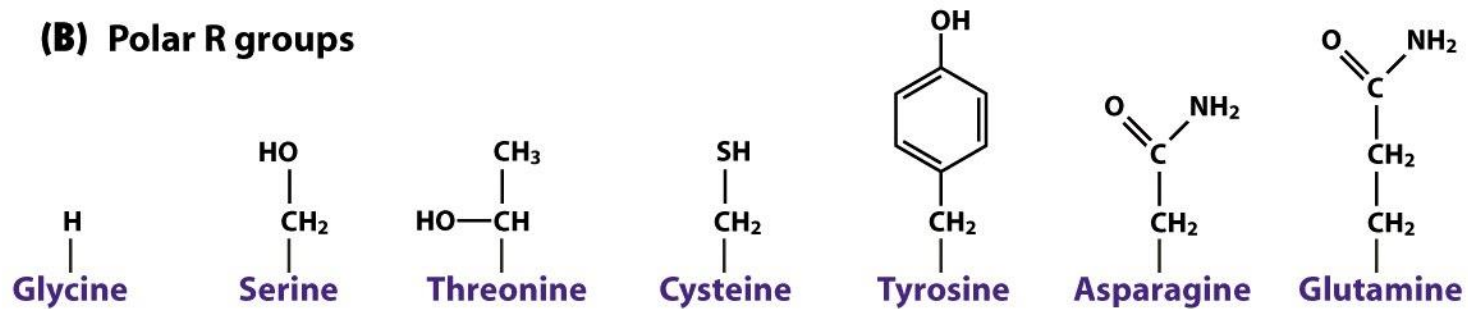


Figure 1.17 *Genomes 3* (© Garland Science 2007)

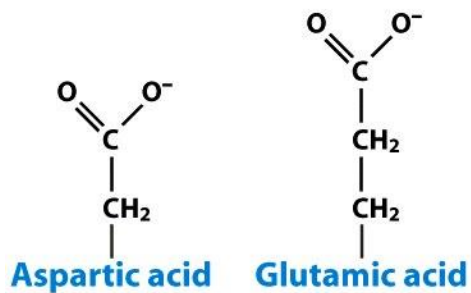
(A) Nonpolar R groups



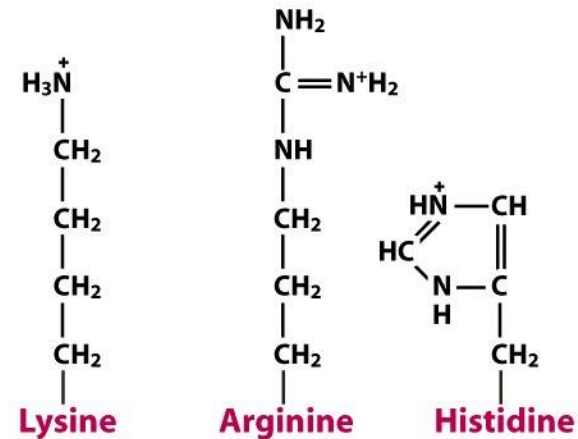
(B) Polar R groups



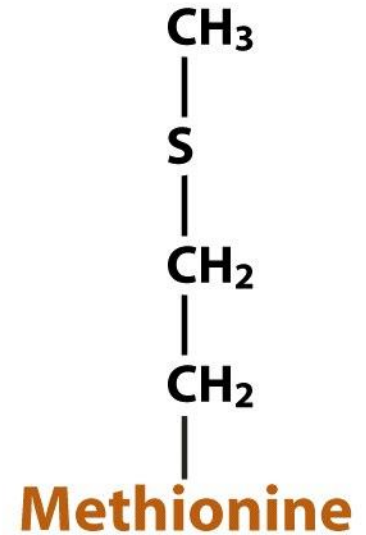
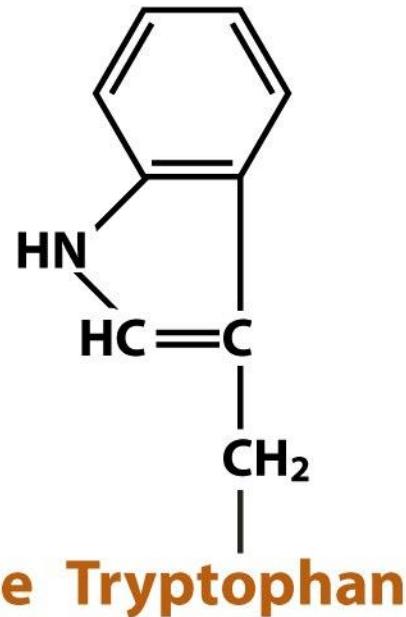
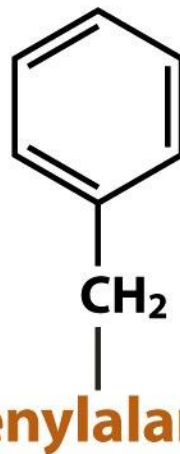
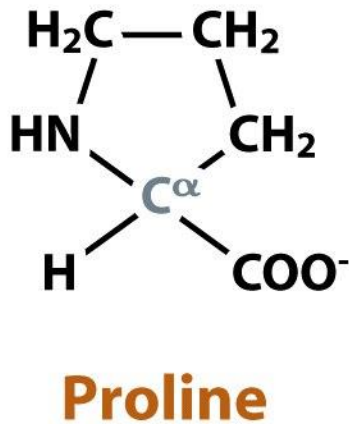
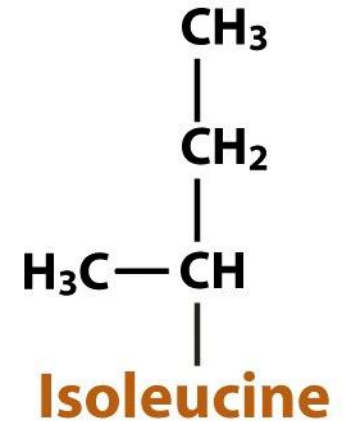
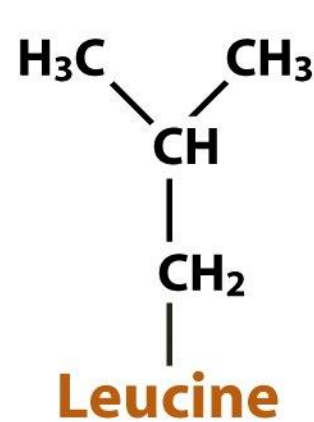
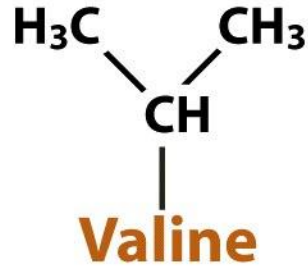
(C) Negatively charged R groups



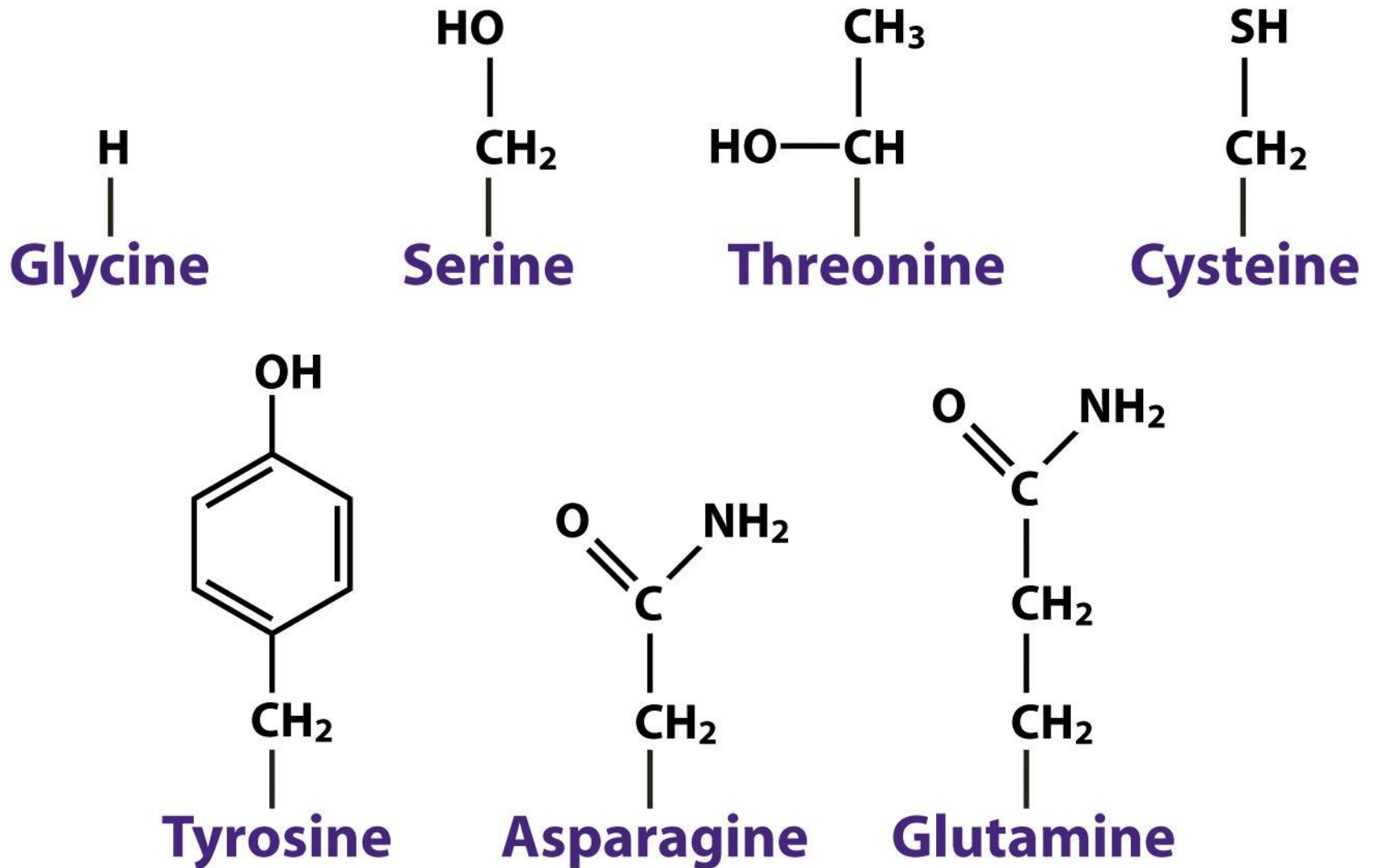
(D) Positively charged R groups



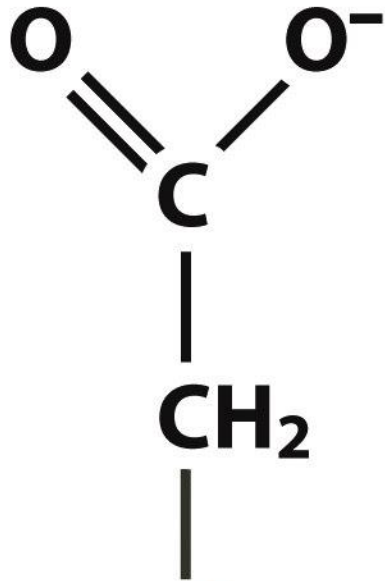
Nonpolar R groups



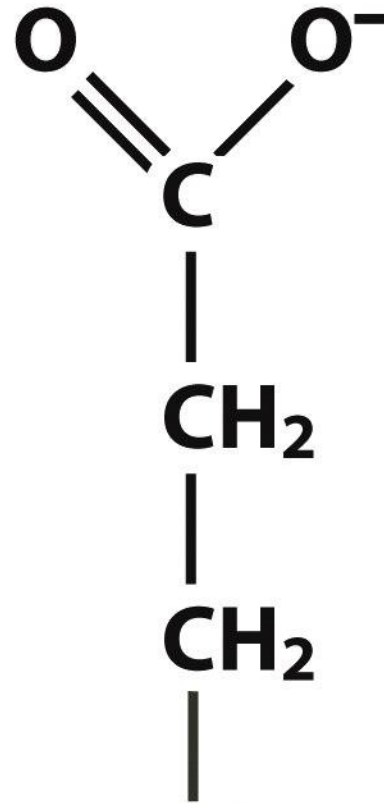
Polar R groups



Negatively charged R groups

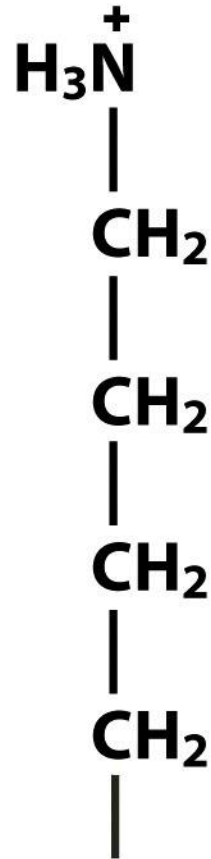


Aspartic acid

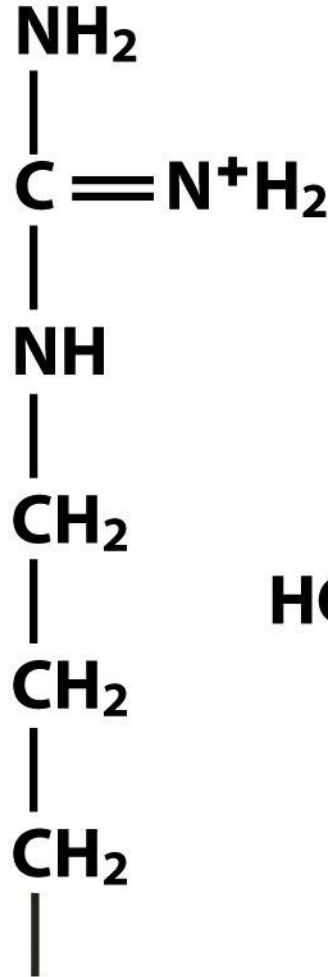


Glutamic acid

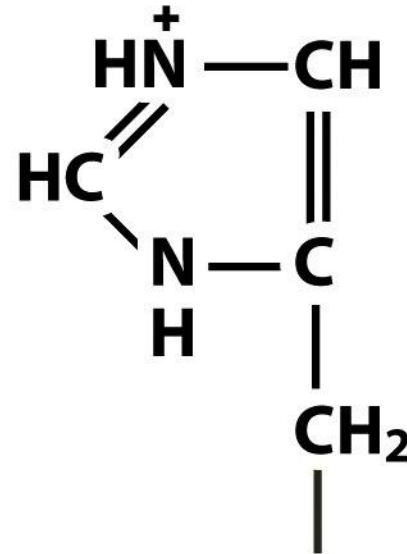
Positively charged R groups



Lysine



Arginine



Histidine

Table 1.2 Amino acid abbreviations

Amino acid	Abbreviation	
	<i>Three-letter</i>	<i>One-letter</i>
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

What is a Protein ?????

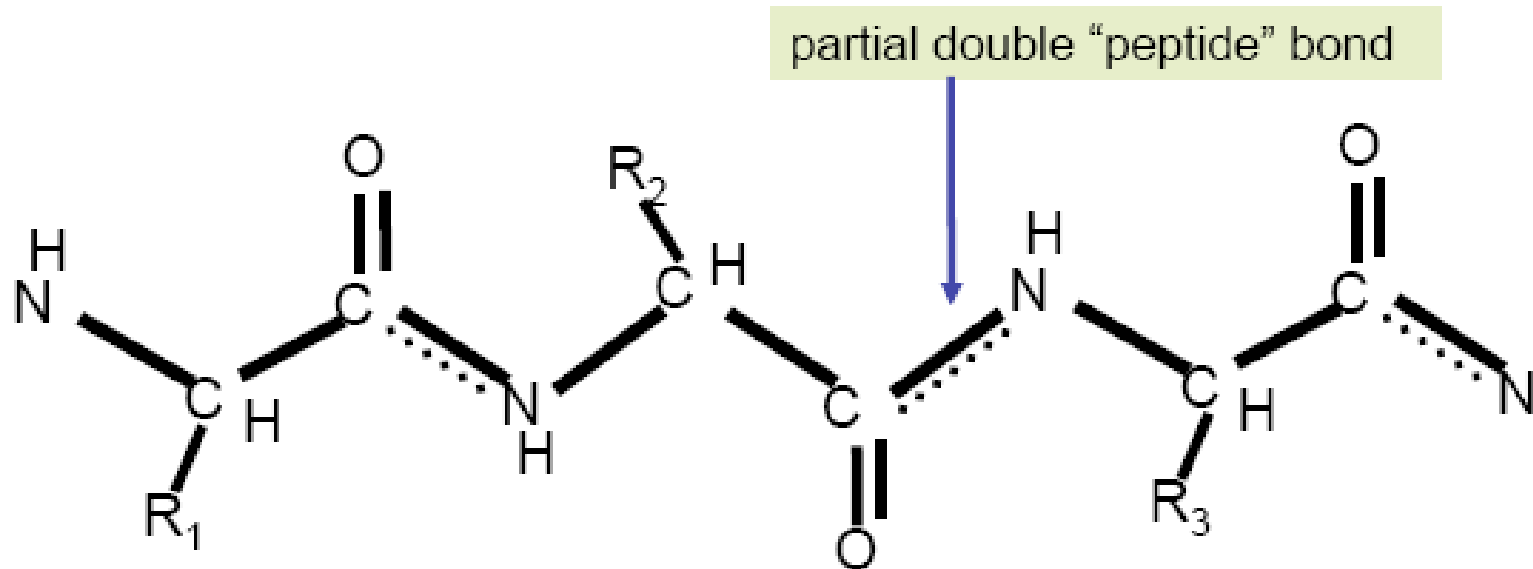
Informatics Viewpoint ..Two underlying principles

1. **Evolution** -- *proteins sequences are selected for function.*
2. **Energy** -- *biochemical systems seek the lowest free energy state.*

**MSAIQASWPSGTECIAKYNFHGTAEQDLPFC
KGDVLTIVAVTKDPNWKAKNKGREGIIPA
NYVQKREGV**

In theory this is all the information you need to make a 3D protein. Maximum information compression. Add water, add physical forces, add time (1 sec will do) and you will get a unique (within reason) 3D structure.

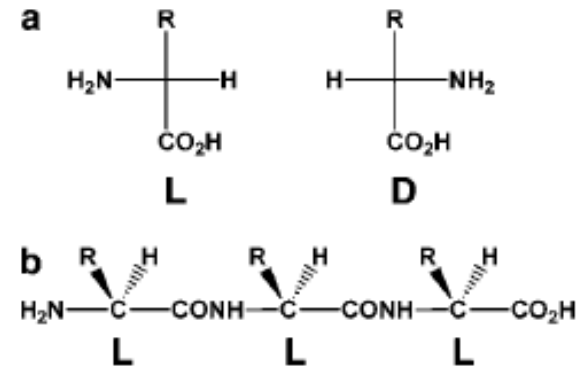
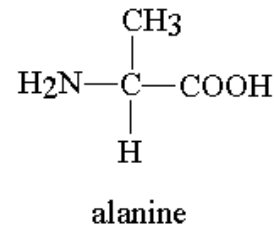
the backbone atoms



Backbone atom names are "N", "C-alpha" and "C" (or N,CA,C). Oxygen "O" is also considered a backbone atom, though strictly speaking it is a one-atom sidechain. All atoms in all amino acids have conventional names. Address them by name or they will not answer.

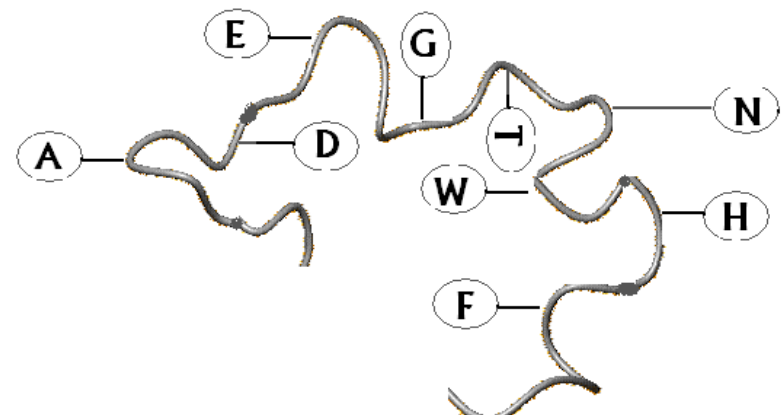
Protein is a (Hetero) polymer

Typical monomer

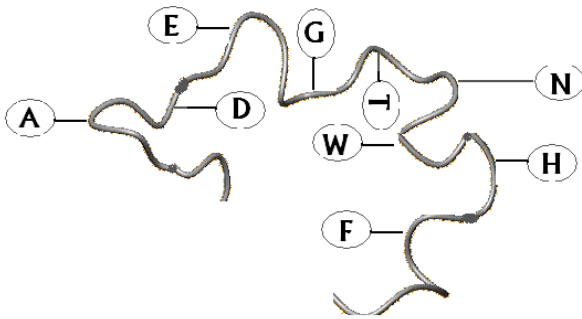


Amino acid sequence:

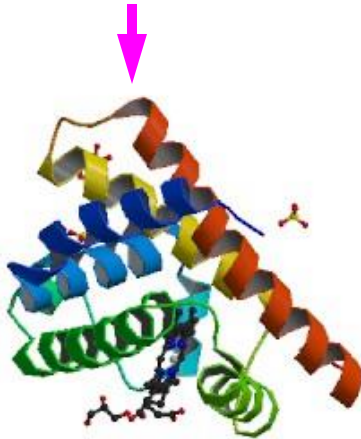
GLSDGEWQQVLNVWGKVEADIAGHGQEVLI
RLFTGHPETLEKFDKFKHLKTEAEMKASEDL
KKHGTVVLTA LGGILKKKGHHEAELKPLAQS
HATKHKIPIKYLEFISDAIIHVLHSHKHPGDFGA
DAQGAMTKALELFRNDIAAKYKELGFQGG



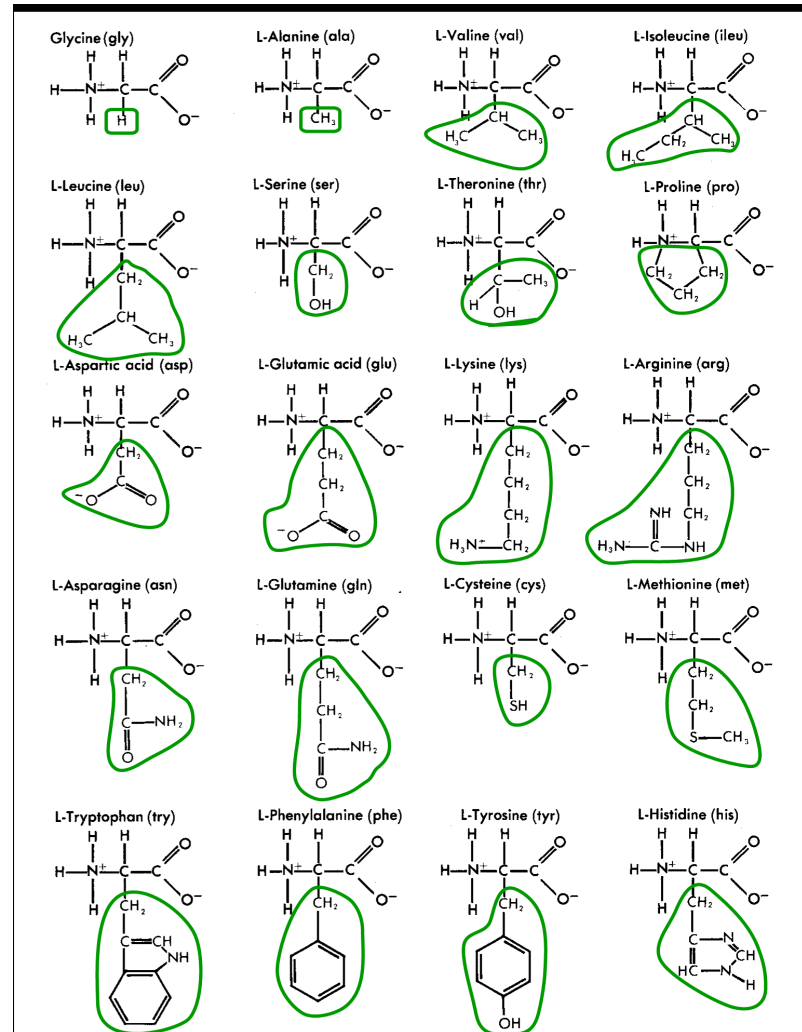
Folding of Protein to its prescribed structure



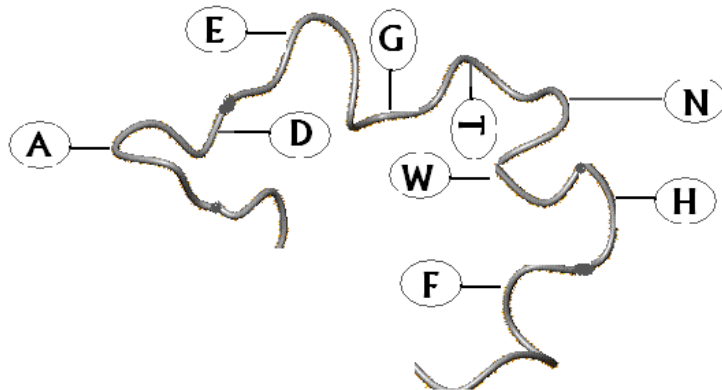
Non - Functional form



Functional form

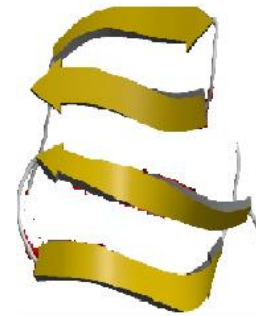


Structural Hierarchy in Proteins

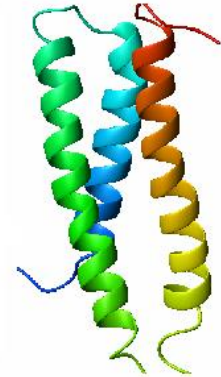


Primary

Sequence: ADEGTNWHF

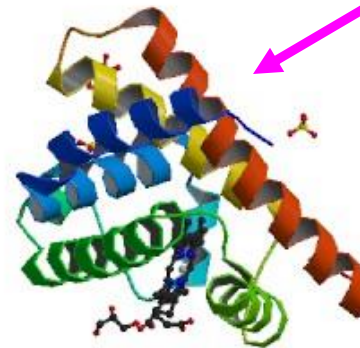
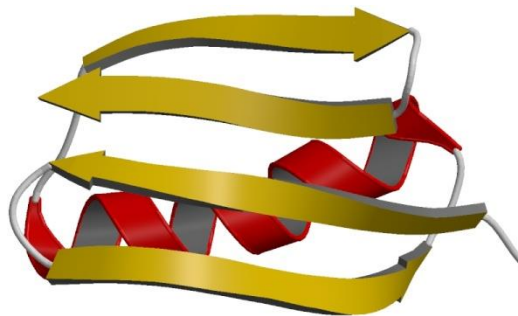


Sheet



Helix

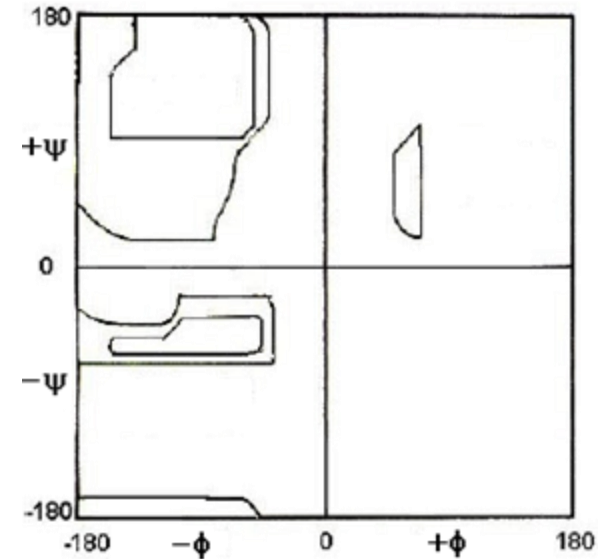
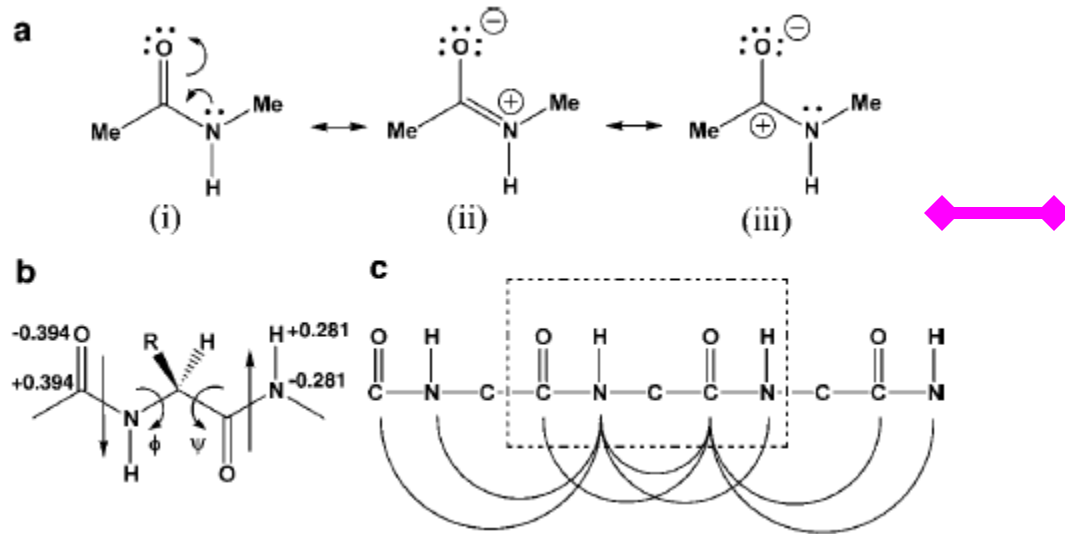
Secondary



Tertiary

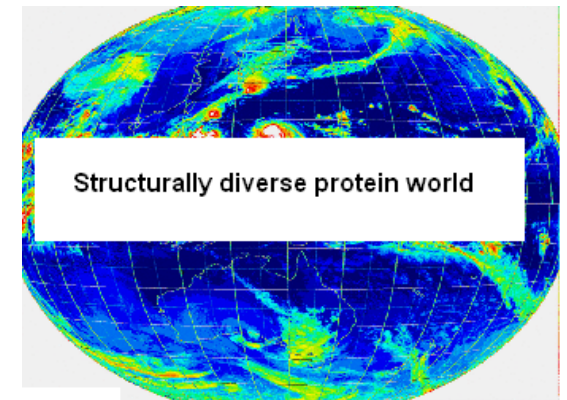
Local organization restricted (largely) to two - α helices and β sheets

The planar peptide bond

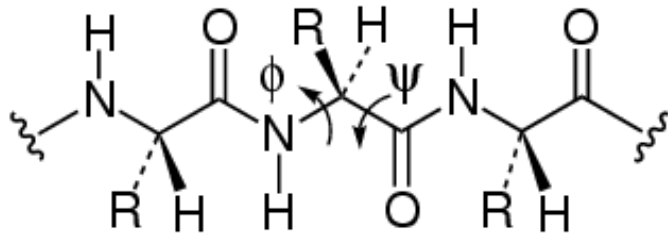
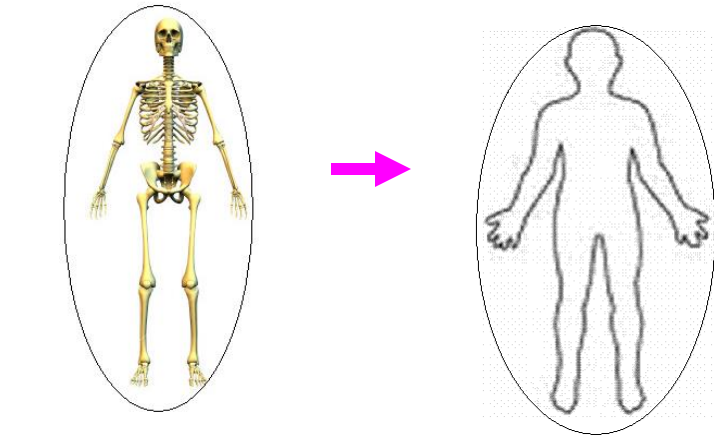


Conformationally constrained chain due to the planarity of the peptide bond

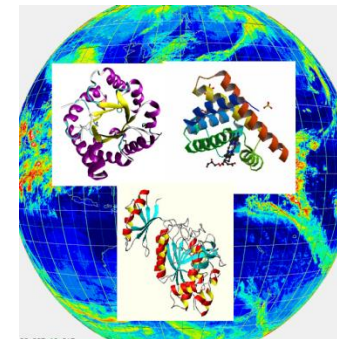
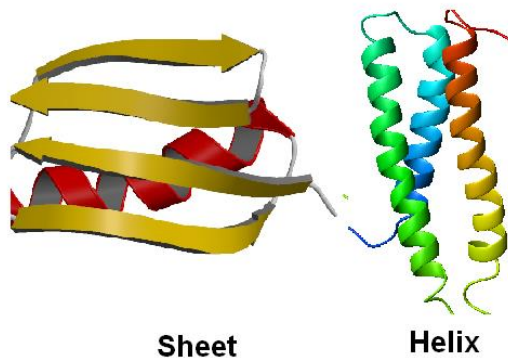
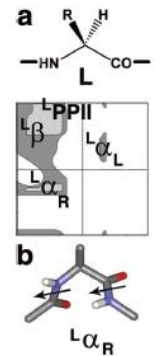
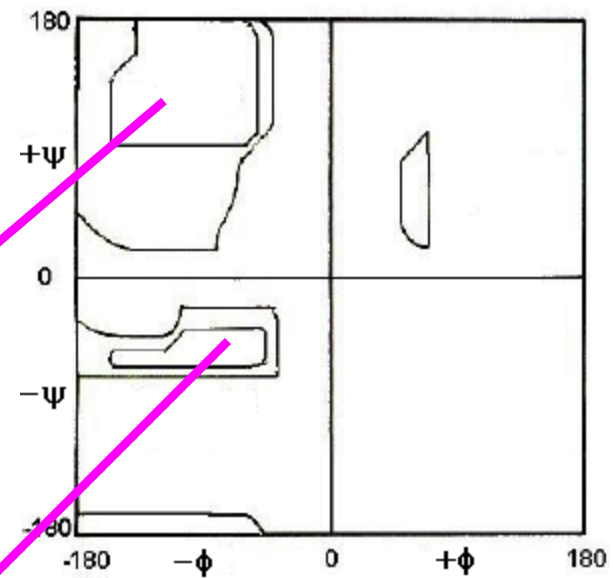
Conformational variables: ϕ, ψ dihedral angles



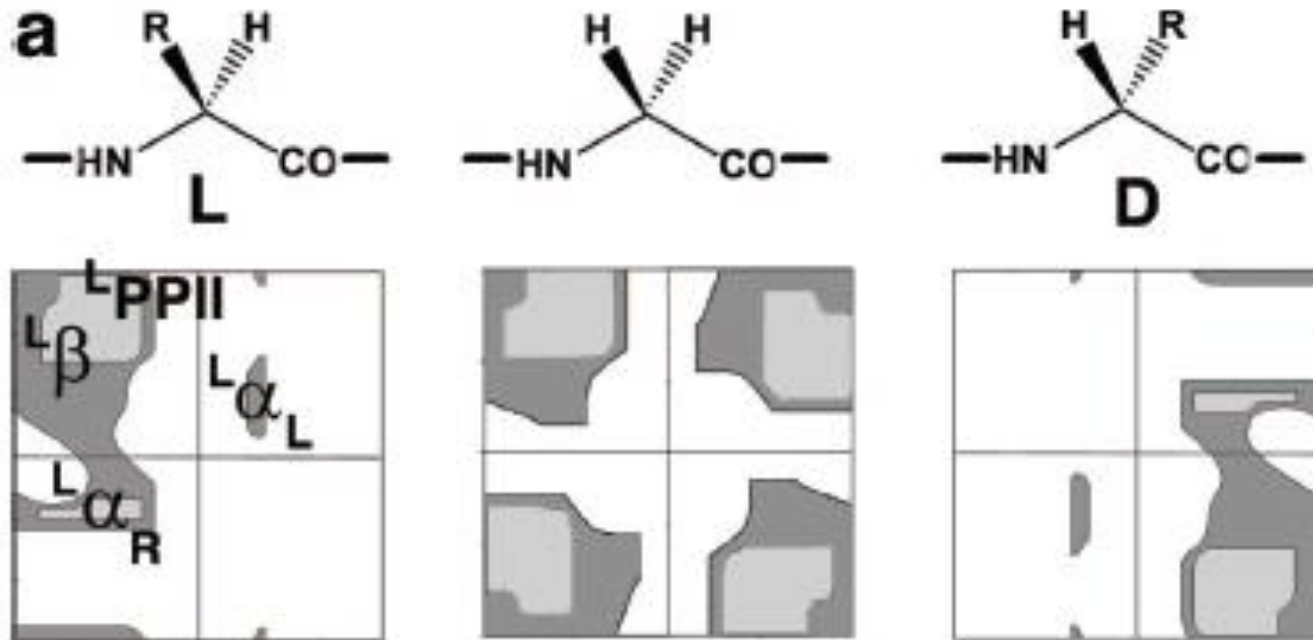
Skeletal Framework of a protein



Dihedral angle map
Ramachandran Map

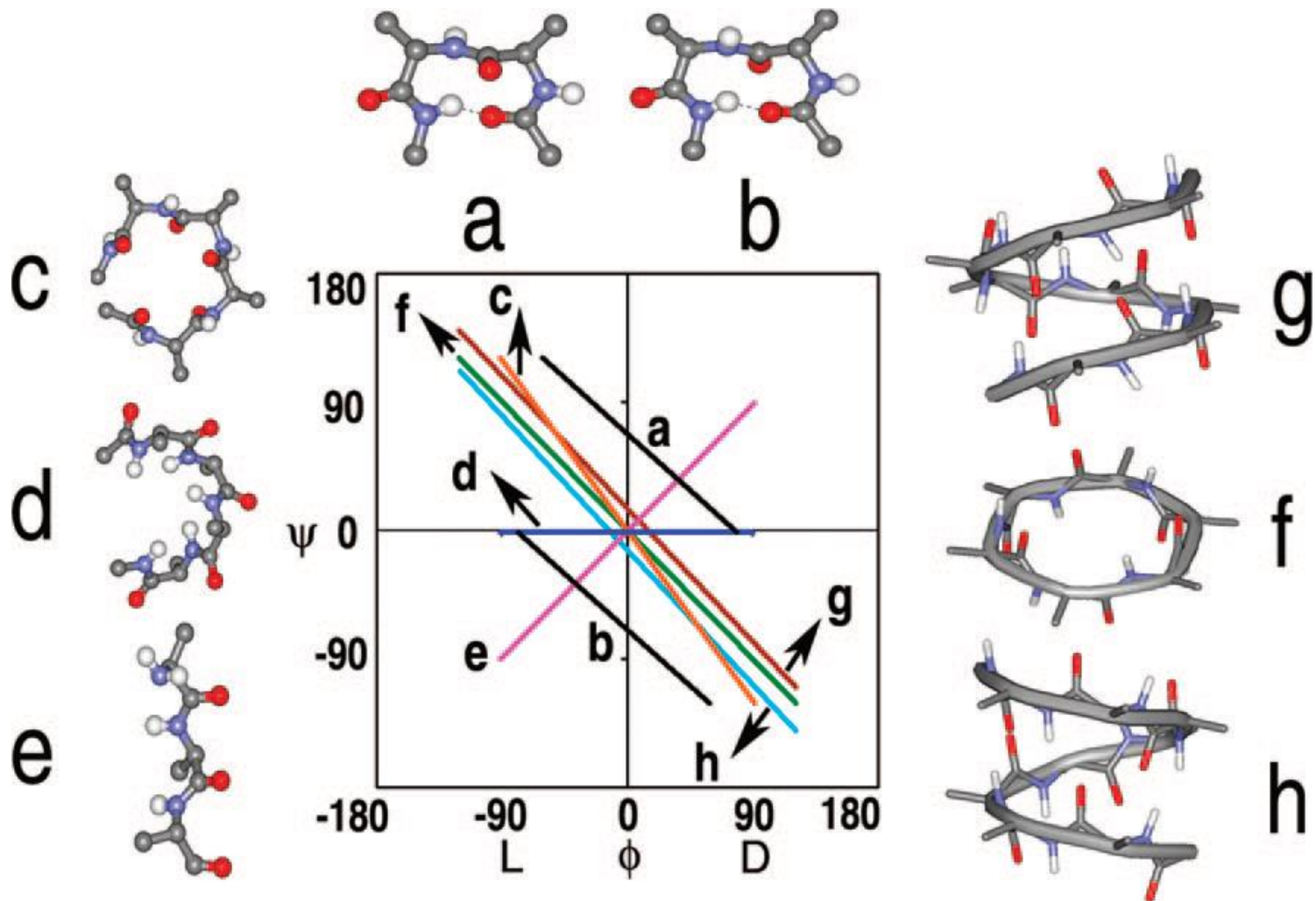


L and D amino acids



- L and D convention for amino acid configuration refers not to the optical activity of the amino acid
- Refers to the optical activity of the isomer of glyceraldehyde from which that amino acid can theoretically be synthesized
- (D-glyceraldehyde is dextrorotary; L-glyceraldehyde is levorotary).

Random Walk across Ramachandran Map



The Scope

2^{100} ($\sim 10^{30}$) diastereomeric possibilities for a 100 residue protein

(Poly L diastereomer is the only one known!)

20^{100} sequence variants for each diastereomer.

Total search space of $2^{100} \times 20^{100}$ variants

