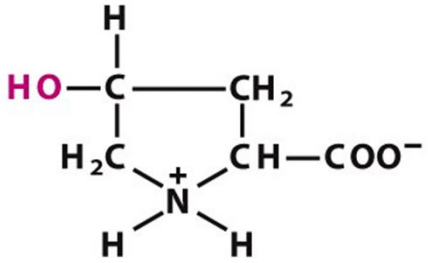
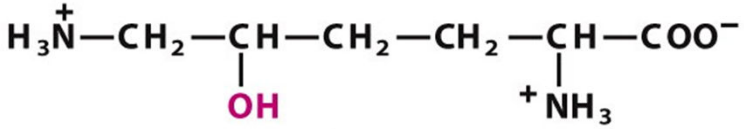
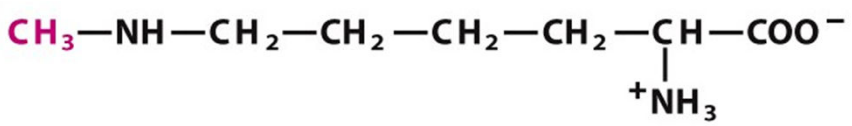
Nonstandard amino acids are found in modified proteins and as free metabolites



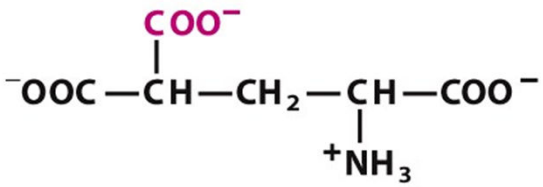
4-Hydroxyproline



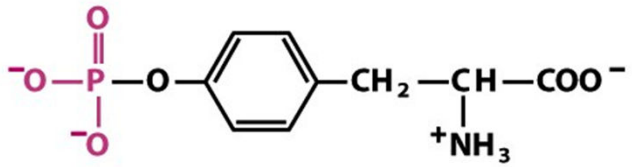
5-Hydroxylysine



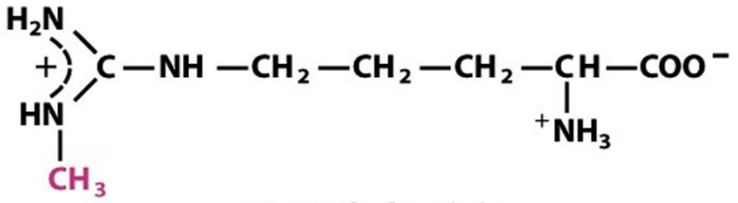
6-*N*-Methyllysine



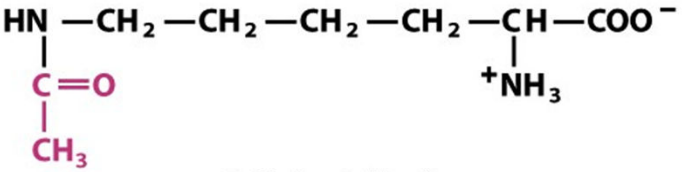
γ-Carboxyglutamate



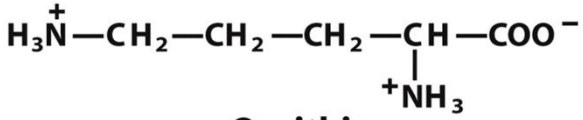
Phosphotyrosine



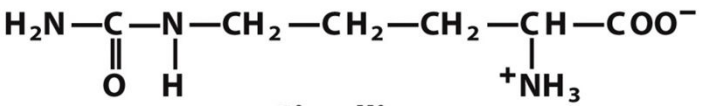
σ-*N*-Methylarginine



6-*N*-Acetyllysine

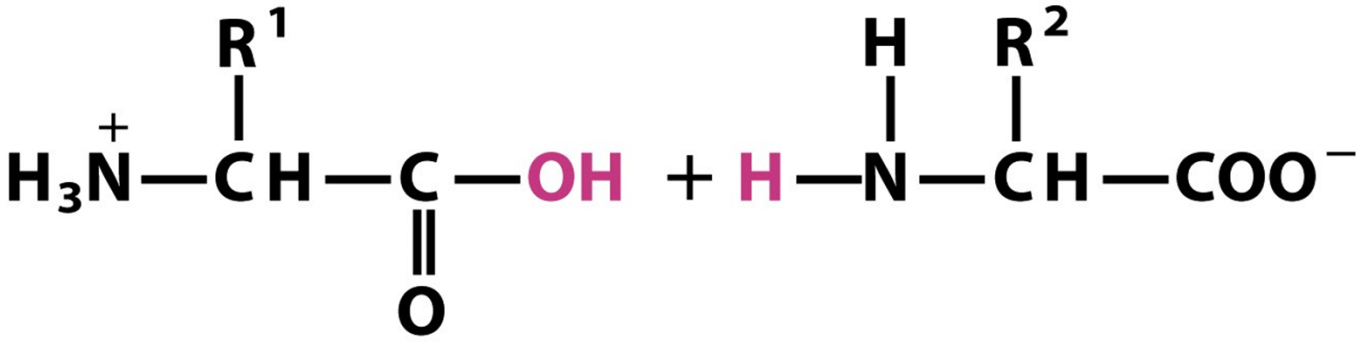


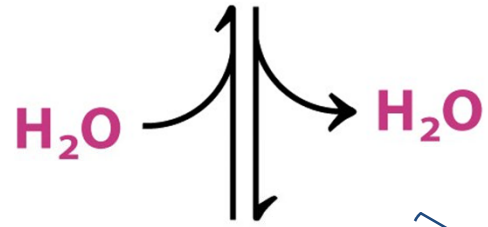
Ornithine



Citrulline

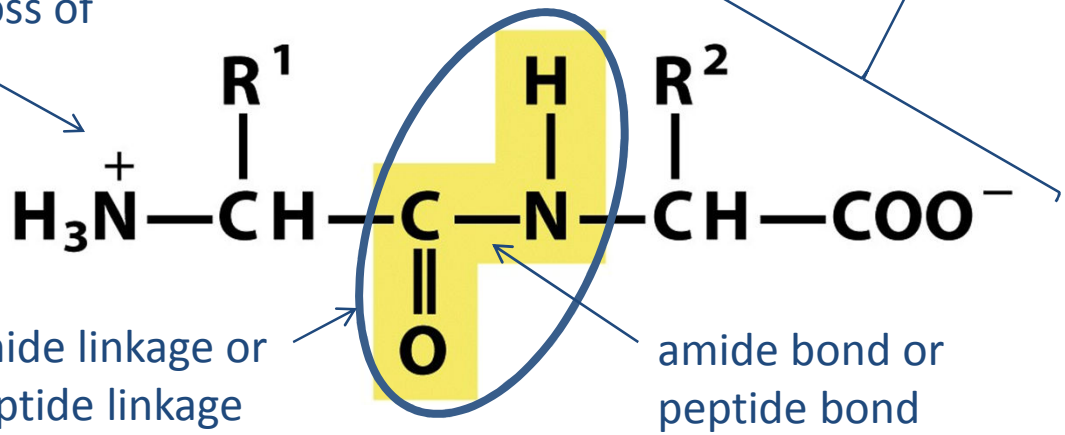
Amino acids connect via amide linkages (releasing water – a condensation reaction) to form peptides





In polymers formed through condensation, subunits are often called 'residues,' due to loss of water

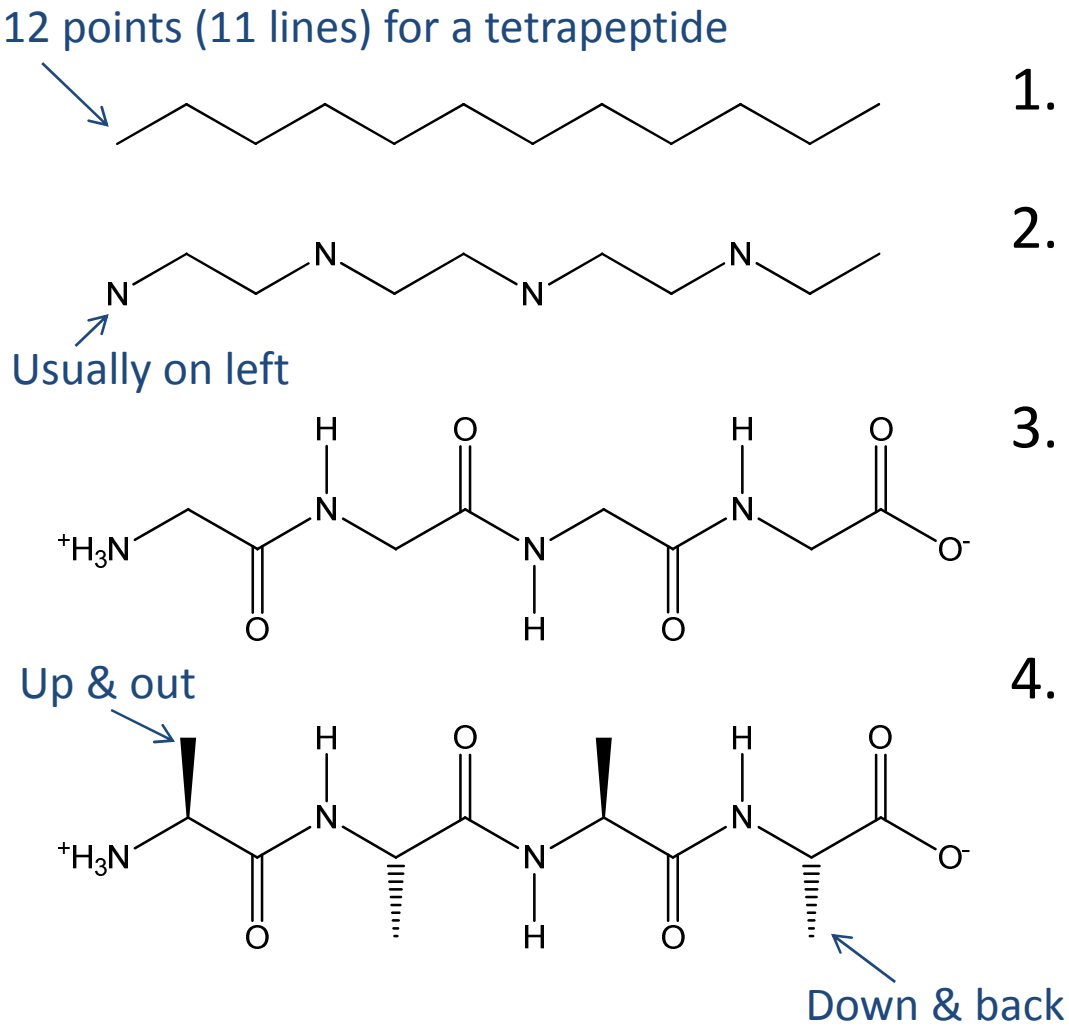
- dipeptide (2 aa's)
- oligopeptide (few aa's)



amide linkage or peptide linkage

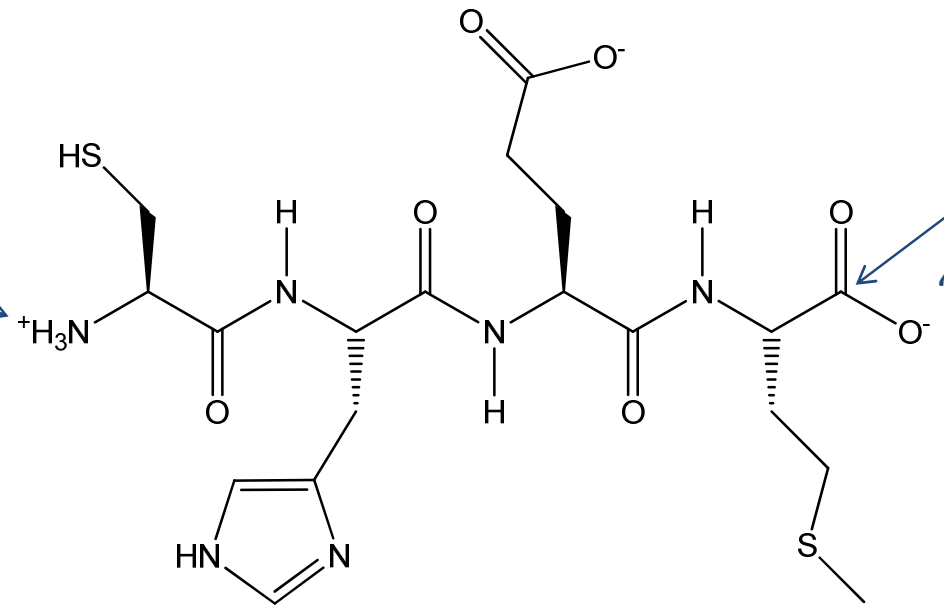
amide bond or peptide bond

How to draw a peptide with correct stereochemistry



1. Draw the backbone, 3 points per amino acid
2. Add in nitrogen: 1st point and every 3rd following
3. Add hydrogens and oxygens to complete the backbone
4. Add side chains: draw "up and out, down and back" for L-amino acids (opposite for D)

There are several ways to name a peptide's sequence, but all start from the N-terminus



"Amino terminus"
"N-terminus"
"Amino-terminal end"
"N-terminal end"
"N-term"

"Carboxyl terminus"
"C-terminus"
"Carboxyl-terminal end"
"C-terminal end"
"C-term"

1. Name amino acids as substituents of C-terminal amino acid: Cysteinyl-histidinyl-glutamyl-methionine (rare, except dipeptides)
2. Write three-letter abbreviations : Cys-His-Glu-Met (common)
3. Write one-letter abbreviations : CHEM (most common)

→ Names imply L stereochemistry; any D must be indicated (ex: Gly-D-Ala-Pro)

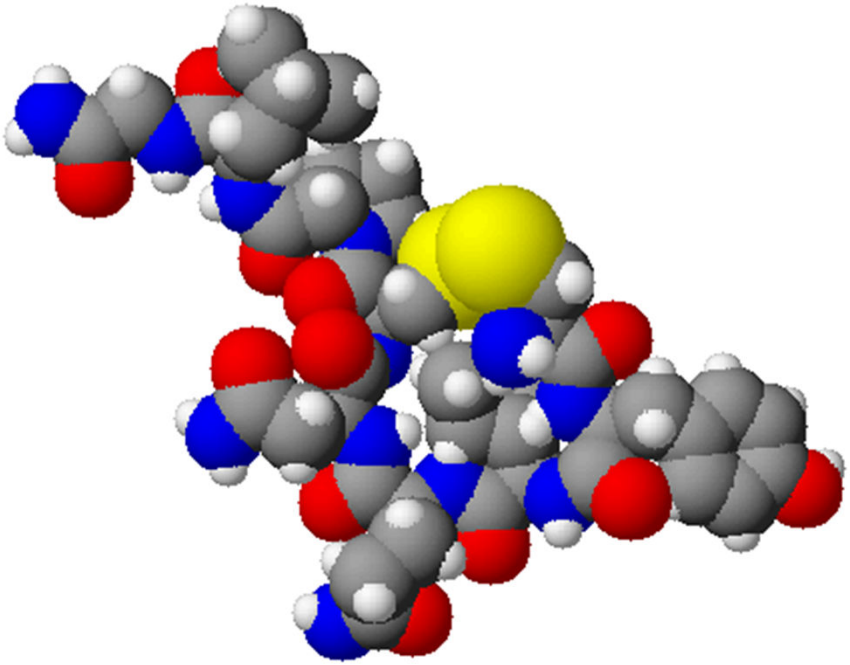
Small peptides are important in biochemistry

Peptide hormones

Ex: oxytocin (the love hormone)

Causes uterine contractions

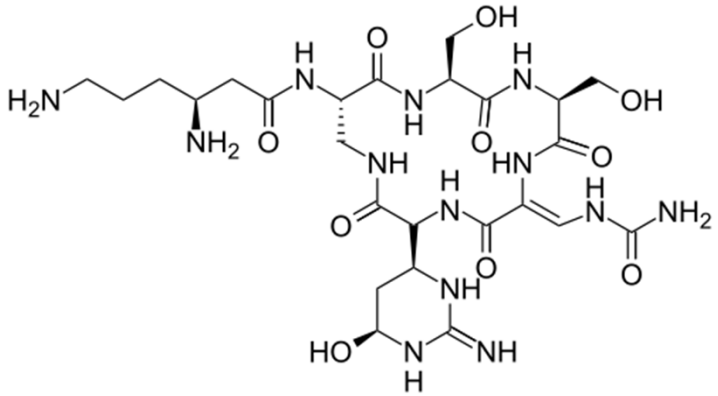
Important for forming connections



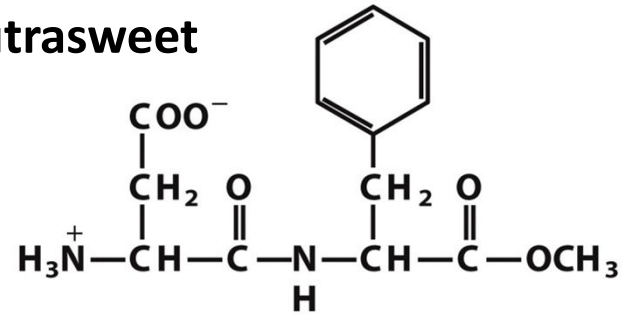
Cys-Tyr-Ile-Gln-Asn-Cys-Pro-Leu-Gly-NH_2

Peptide antibiotics

Ex: viomycin – Used in a drug cocktail against *M. tuberculosis*



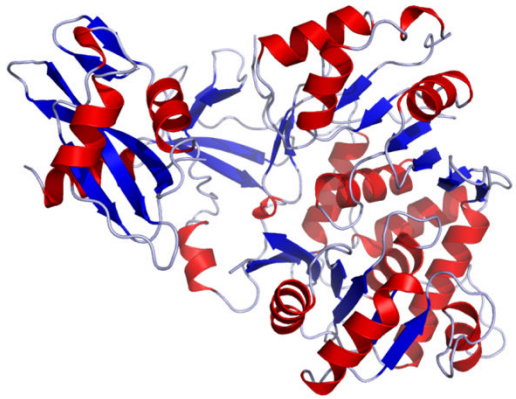
Nutrasweet



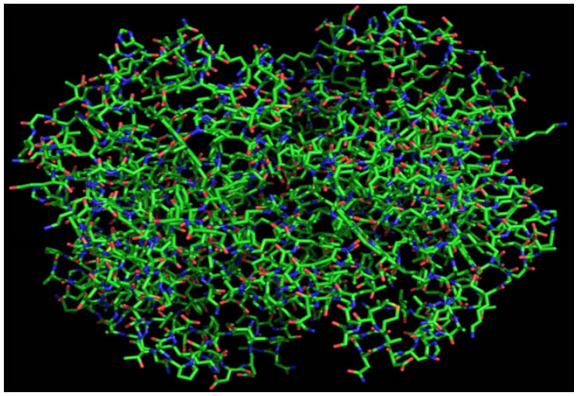
L-Aspartyl-L-phenylalanine methyl ester (aspartame)

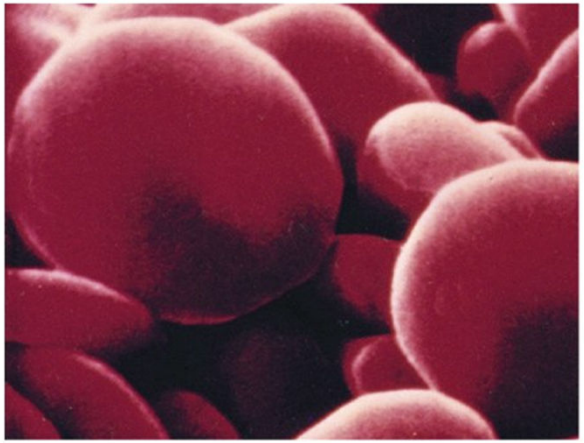
Proteins are essential components of all organisms and carry out a diversity of functions

Luciferase (an enzyme) – catalyzes a light-producing reaction in fireflies

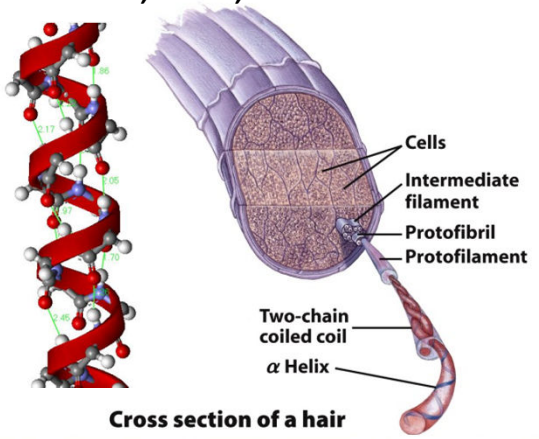


Hemoglobin – binds O_2 and carries it throughout the blood stream





α-Keratin – provides structure to animal horns, hooves, hair, and nails



Proteins vary in size and in number of chains

TABLE 3–2 Molecular Data on Some Proteins

	Molecular weight	Number of residues	Number of polypeptide chains
Cytochrome *c* (human)	13,000	104	1
Ribonuclease A (bovine pancreas)	13,700	124	1
Lysozyme (chicken egg white)	13,930	129	1
Myoglobin (equine heart)	16,890	153	1
Chymotrypsin (bovine pancreas)	21,600	241	3
Chymotrypsinogen (bovine)	22,000	245	1
Hemoglobin (human)	64,500	574	4
Serum albumin (human)	68,500	609	1
Hexokinase (yeast)	102,000	972	2
RNA polymerase (*E. coli*)	450,000	4,158	5
Apolipoprotein B (human)	513,000	4,536	1
Glutamine synthetase (*E. coli*)	619,000	5,628	12
Titin (human)	2,993,000	26,926	1

Proteins vary in composition

TABLE 3–3 Amino Acid Composition of Two Proteins

Amino acid	Number of residues per molecule of protein*: Bovine cytochrome c	%	Bovine chymotrypsinogen	%
Ala	6	6	22	9
Arg	2	2	4	2
Asn	5	5	15	6
Asp	3	3	8	3
Cys	2	2	10	4
Gln	3	3	10	4
Glu	9	9	5	2
Gly	14	13	23	9
His	3	3	2	1
Ile	6	6	10	4
Leu	6	6	19	8
Lys	18	17	14	6
Met	2	2	2	1
Phe	4	4	6	2
Pro	4	4	9	4
Ser	1	1	28	11
Thr	8	8	23	9
Trp	1	1	8	3
Tyr	4	4	4	2
Val	3	3	23	9
Total	104	100	245	100

← Varying proportions of amino acids

Variability in use of additional compounds ↓

TABLE 3–4 Conjugated Proteins

Class	Prosthetic group
Lipoproteins	Lipids
Glycoproteins	Carbohydrates
Phosphoproteins	Phosphate groups
Hemoproteins	Heme (iron porphyrin)
Flavoproteins	Flavin nucleotides
Metalloproteins	Iron
	Zinc
	Calcium
	Molybdenum
	Copper

Protein variability is theoretically limitless (although realistically limited)

```
 1    2    3    4    5    …
Met-Ala-Phe-Gly-Ala-Pro-Gly-Asp-Gln-His-…
20 x 20 x 20 x 20 x 20 x …
```

For a protein with 100 aa, number of possible aa sequences = $20^{100} \approx 10^{130}$
For comparison, there are ~10^{80} atoms in the (observable) universe!

Additional variability can come from:

- Variation in chain length
- Variation in number of chains
- Protein modifications
- Binding of prosthetic groups

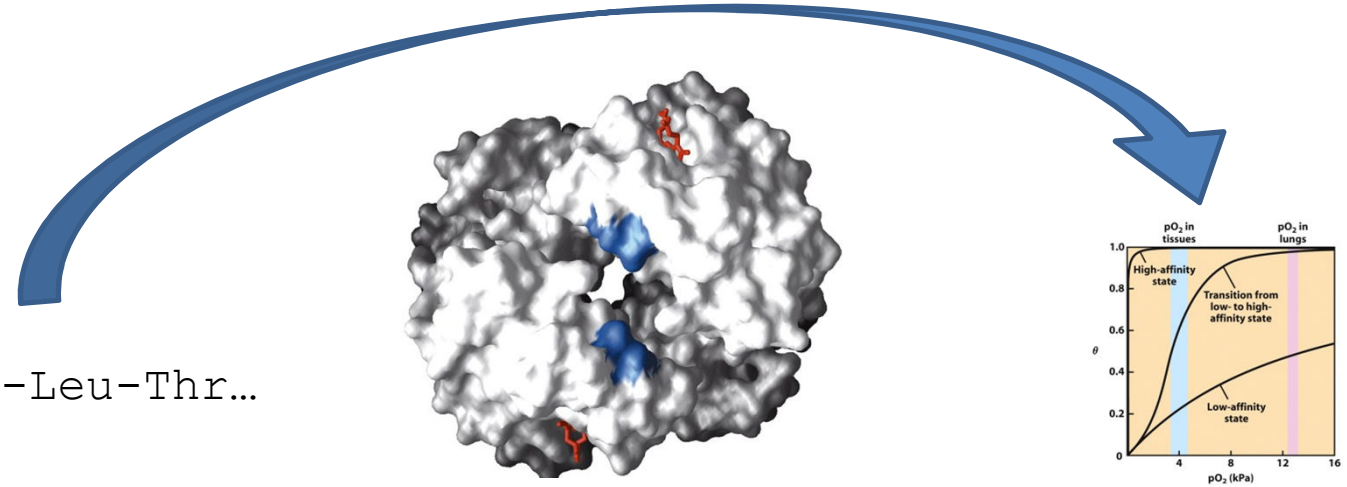
A protein is a folded, functional polypeptide (not just any polymer of amino acids)

Realistic limitations:

- Length limited by ability and fidelity of synthesis
- Parameters limited by functionality, usefulness – Does it fold? Does it provide a needed, useful function?
- Parameters of natural proteins are limited by evolution – Did nature find & keep it?

There are maybe 10^7 proteins on earth

A protein's function derives from its structure, and its structure is determined by its sequence.



Sequence → **Structure** (Conformation) (Fold) → **Function**

How?

The properties of the amino acids determine which can interact and how.

The connectivity (sequence) limits the possible interactions and directs the position of the polypeptide chain.

Non-covalent interactions and reversible bonds are important in the structure of proteins

Electrostatic interactions

Ionic interactions
(also called *ion pairs* or *salt bridges*)

Hydrogen bonds

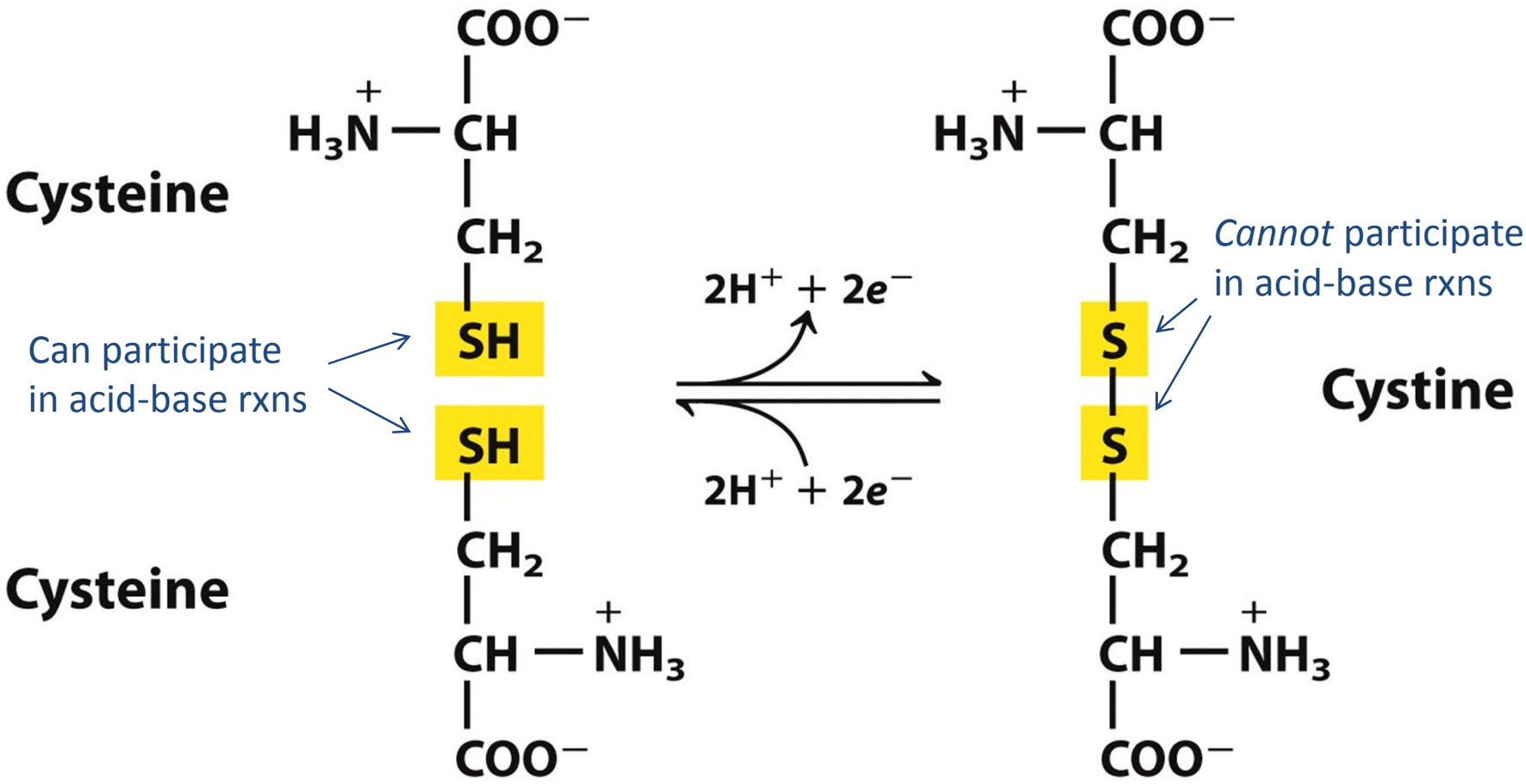
van der Waals forces
- Dipole-dipole interactions
- Dipole-induced dipole interactions
- London dispersion forces

Other interactions

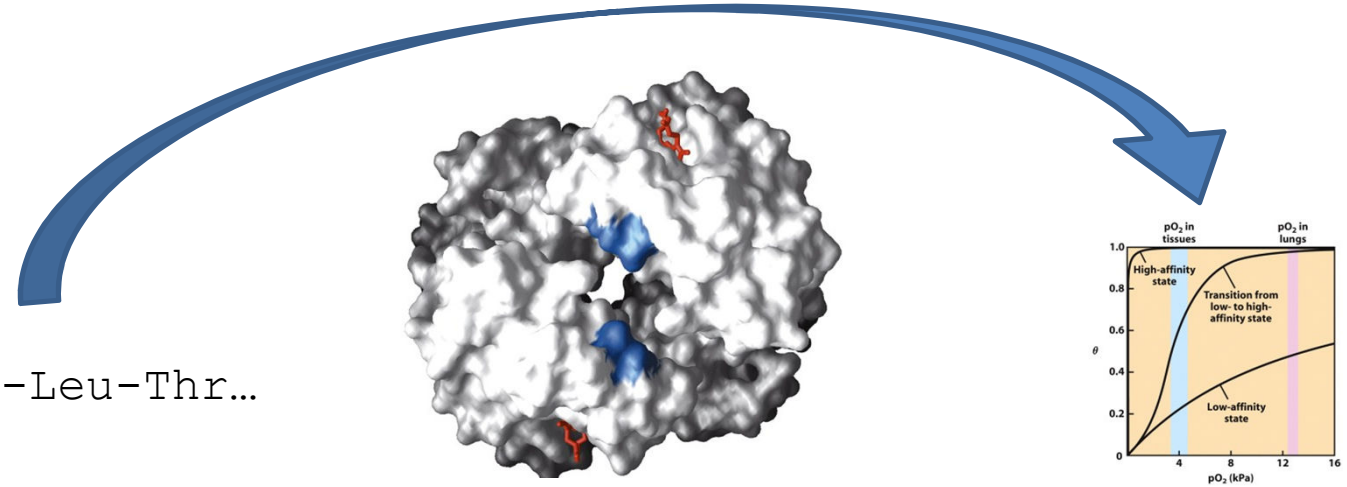
Hydrophobic effect

Disulfide bonds

Disulfide bonds form when two sulfhydryl groups are oxidized (give up electrons)



A protein's function derives from its structure, and its structure is determined by its sequence.



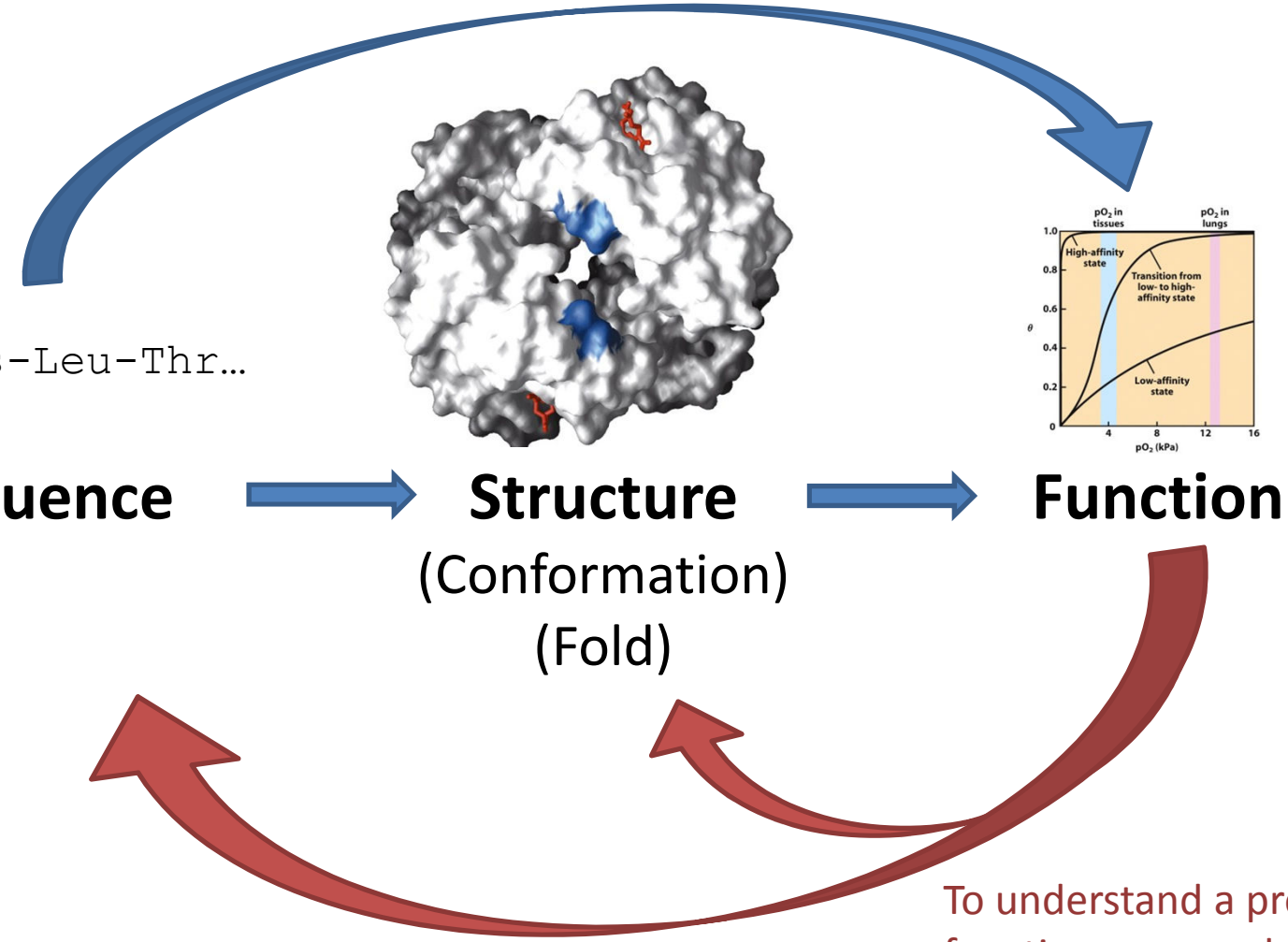
Sequence → **Structure** (Conformation) (Fold) → **Function**

How?

The properties of the amino acids determine which can interact and how.

The connectivity (sequence) limits the possible interactions and directs the position of the polypeptide chain.

A protein's function derives from its structure, and its structure is determined by its sequence.



To understand a protein's function, we need to know its structure (and sequence)