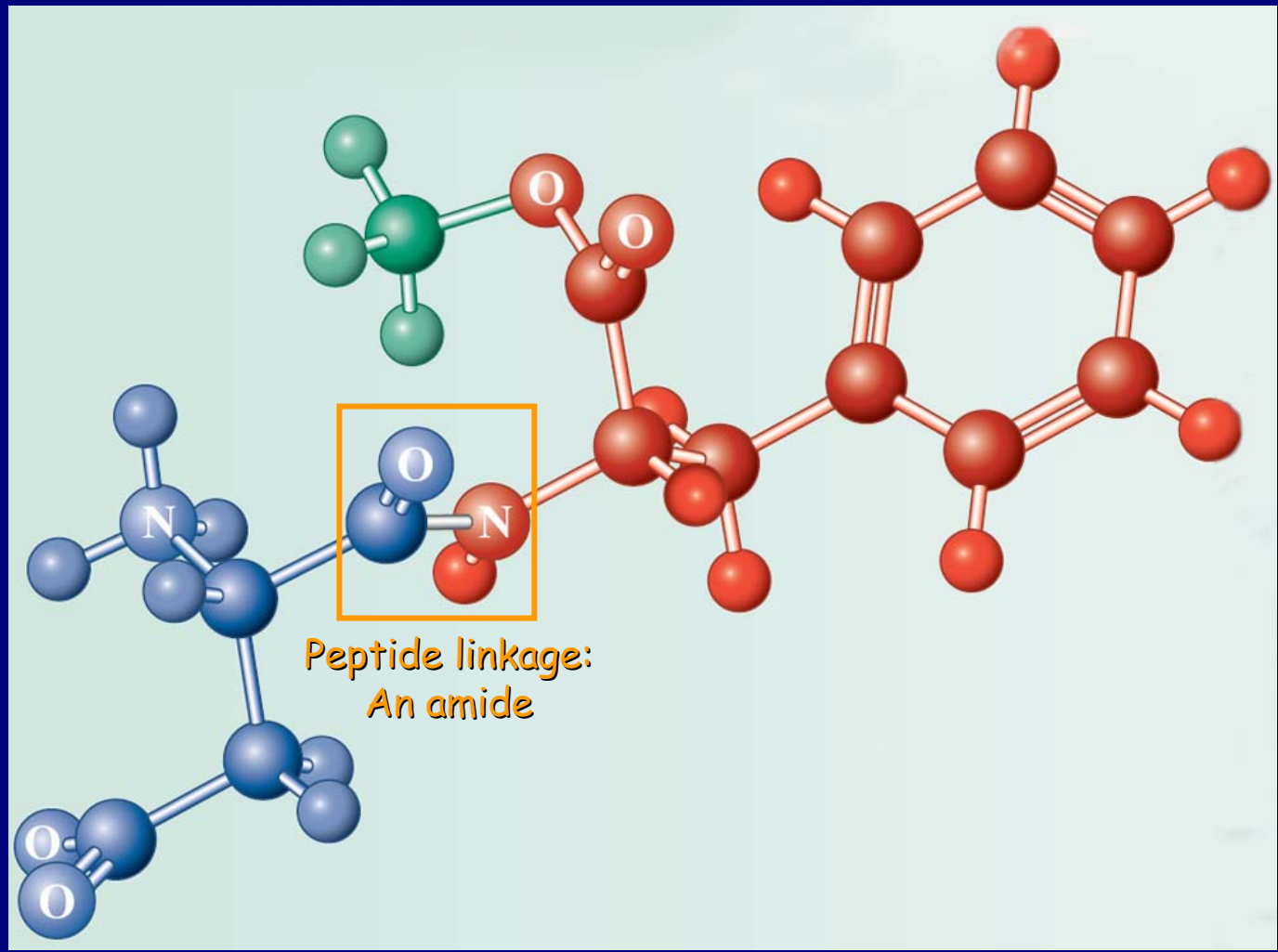
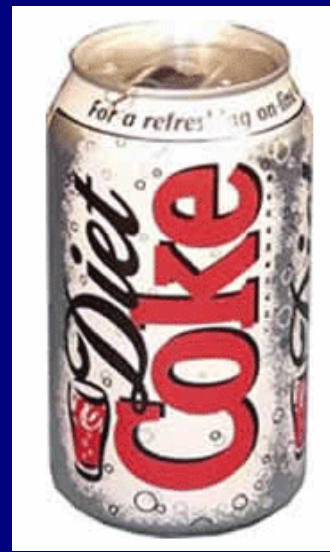
Chapter 26: Amino Acids, Peptides, Proteins, and Nucleic Acids

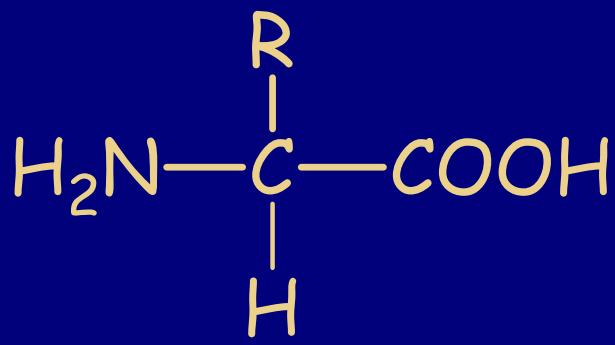




Aspartame, a dipeptide

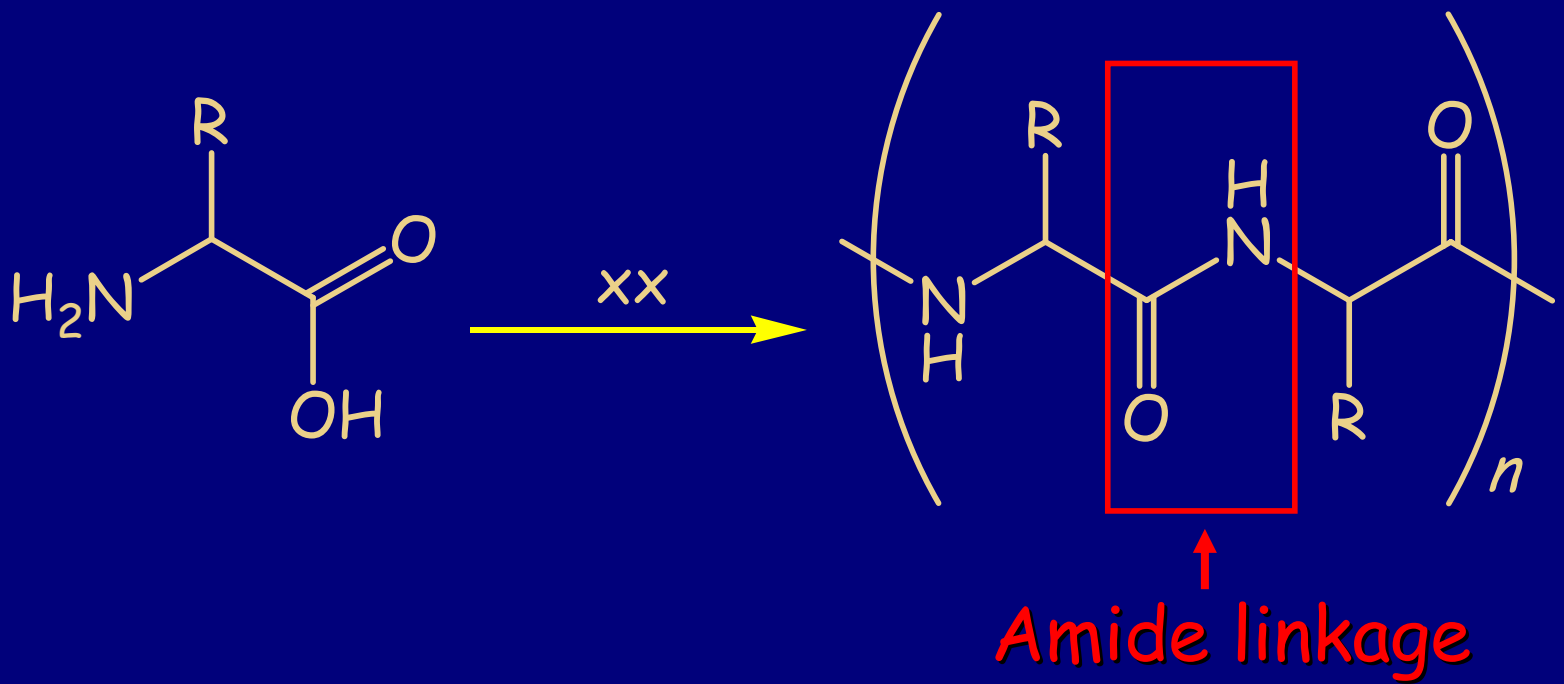
Amino Acids

In nature, the most common representatives are 2-amino acids (α-amino acids) with the general formula $RCH(NH_2)COOH$.



R = alkyl, acyl, amino, hydroxy, mercapto, sulfide, carboxy, guanidino, or imidazolyl groups

Amino acids give rise to polyamides: Polypeptides (proteins, enzymes)



For proteins, $n \geq 8000$ and MW > 1,000,000. Proteins are crucial for transport (O_2, hemoglobin), energy storage, catalysis, control of reactions, template for RNA/DNA action, antibodies, etc.

- More than 500 natural amino acids known
- Among the 20 main amino acids, 8 cannot be synthesized by the body (essential amino acids)

Names: We use common names. α-Stereocenter usually *S* (or, old nomenclature, L, from L-glyceraldehyde)

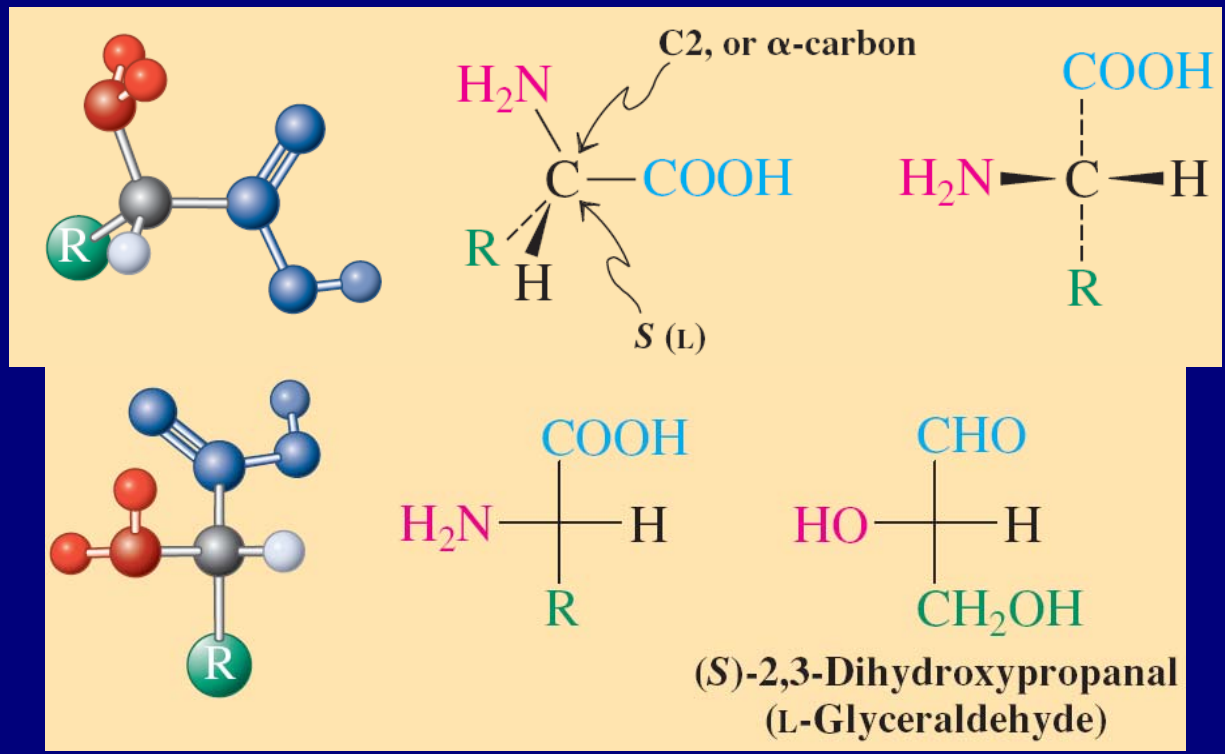
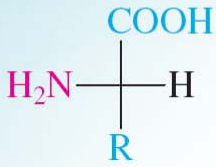
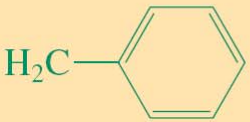
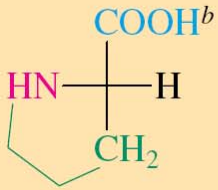
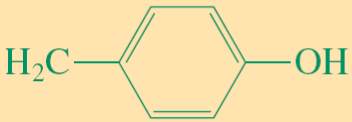
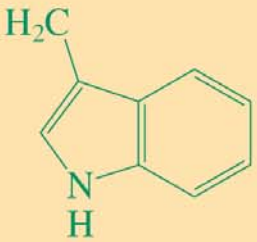
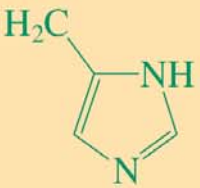


TABLE 26-1 Natural (2*S*)-Amino Acids



R	Name	Three-letter code	One-letter code	pK_a of α-COOH	pK_a of α-$^+NH_3$	pK_a of acidic function in R	Isoelectric point, p*I*
H	Glycine	Gly	G	2.3	9.6	—	6.0
Alkyl group							
CH_3	Alanine	Ala	A	2.3	9.7	—	6.0
$CH(CH_3)_2$	Valine[a]	Val	V	2.3	9.6	—	6.0
$CH_2CH(CH_3)_2$	Leucine[a]	Leu	L	2.4	9.6	—	6.0
$CHCH_2CH_3$ (*S*) CH_3	Isoleucine[a]	Ile	I	2.4	9.6	—	6.0
	Phenylalanine[a]	Phe	F	1.8	9.1	—	5.5
	Proline	Pro	P	2.0	10.6	—	6.3

Hydroxy containing							
CH_2OH	Serine	Ser	S	2.2	9.2	—	5.7
$CHOH$ (*R*) CH_3	Threonine[a]	Thr	T	2.1	9.1	—	5.6
	Tyrosine	Tyr	Y	2.2	9.1	10.1	5.7
Amino containing							
CH_2CNH_2 .	Asparagine	Asn	N	2.0	8.8	—	5.4
$CH_2CH_2CNH_2$	Glutamine	Gln	Q	2.2	9.1	—	5.7
$(CH_2)_4NH_2$	Lysine[a]	Lys	K	2.2	9.0	10.5[c]	9.7
$(CH_2)_3NHCNH_2$	Arginine[a]	Arg	R	2.2	9.0	12.5[c]	10.8

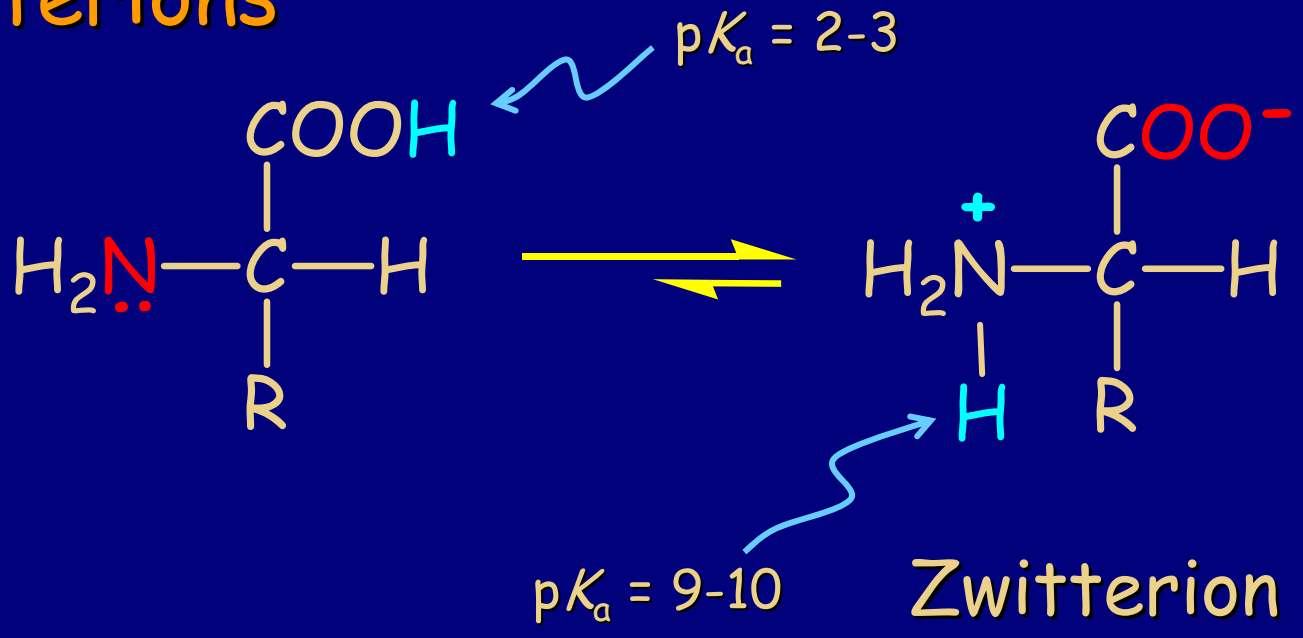
R	Name	Three-letter code	One-letter code	pK_a of α-COOH	pK_a of α-$^+NH_3$	pK_a of acidic function in R	Isoelectric point, pI
Amino containing *(continued)*							
	Tryptophan[a]	Trp	W	2.8	9.4	—	5.9
	Histidine[a]	His	H	1.8	9.2	6.1[c]	7.6
Mercapto or sulfide containing							
CH_2SH	Cysteine[d]	Cys	C	2.0	10.3	8.2	5.1
$CH_2CH_2SCH_3$	Methionine[a]	Met	M	2.3	9.2	—	5.7
Carboxy containing							
CH_2COOH	Aspartic acid	Asp	D	1.9	9.6	3.7	2.8
CH_2CH_2COOH	Glutamic acid	Glu	E	2.2	9.7	4.3	3.2

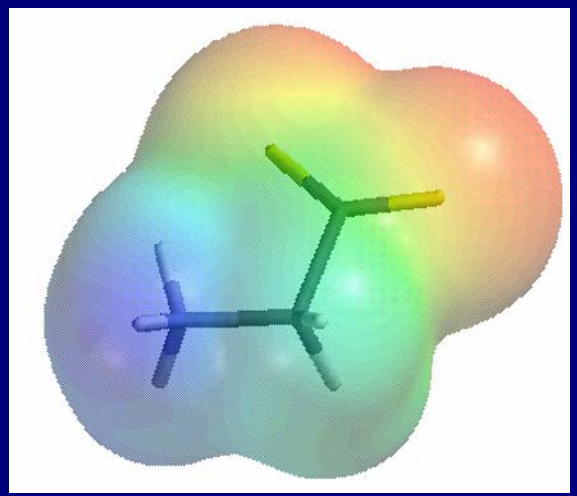
[a]Essential amino acids. [b]Entire structure. [c]pK_a of conjugate acid. [d]The stereocenter is *R* because the CH_2SH substituent has higher priority than the COOH group.

Used in reversible disulfide bridging

Monosodium glutamate

Amino acids are acidic and basic: Exist as zwitterions





$H_3\overset{+}{N}CH_2COO^-$

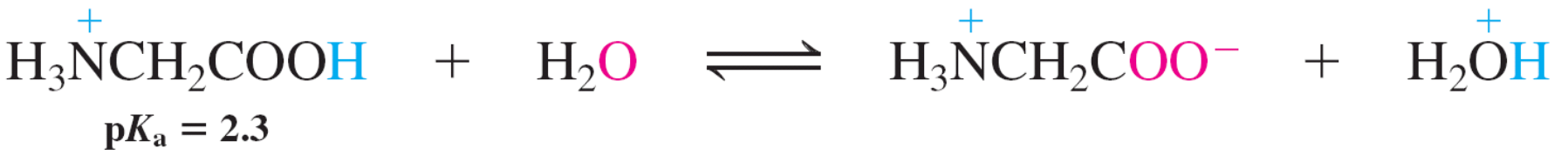
Glycine as a zwitterion

Structure depends on pH:

Predominates at pH < 1

Predominates at pH ~ 6

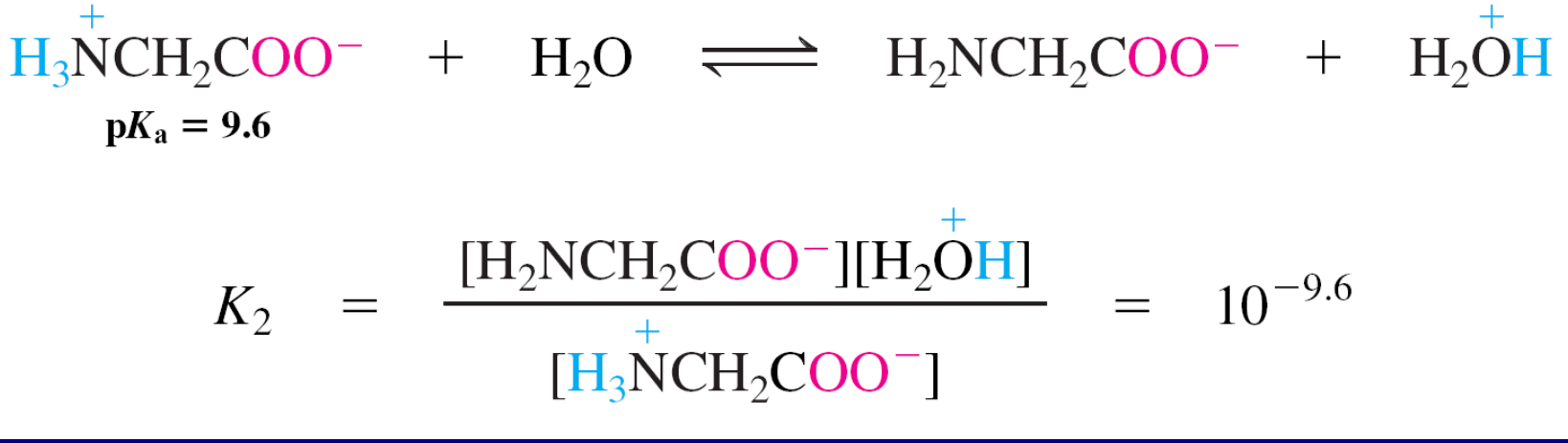
Predominates at pH > 13



$pK_a = 2.3$

$$K_1 = \frac{[H_3\overset{+}{N}CH_2COO^-][H_2O\overset{+}{H}]}{[H_3\overset{+}{N}CH_2COOH]} = 10^{-2.3}$$

Compare pK_a CH_3COOH = 4.74; the -NH_3^+ group acidifies



Cf. pK_a $CH_3NH_3^+ = 10.62$; $-CO_2^-$ also acidifies, by induction

Isoelectric Point: When solution is charge neutral

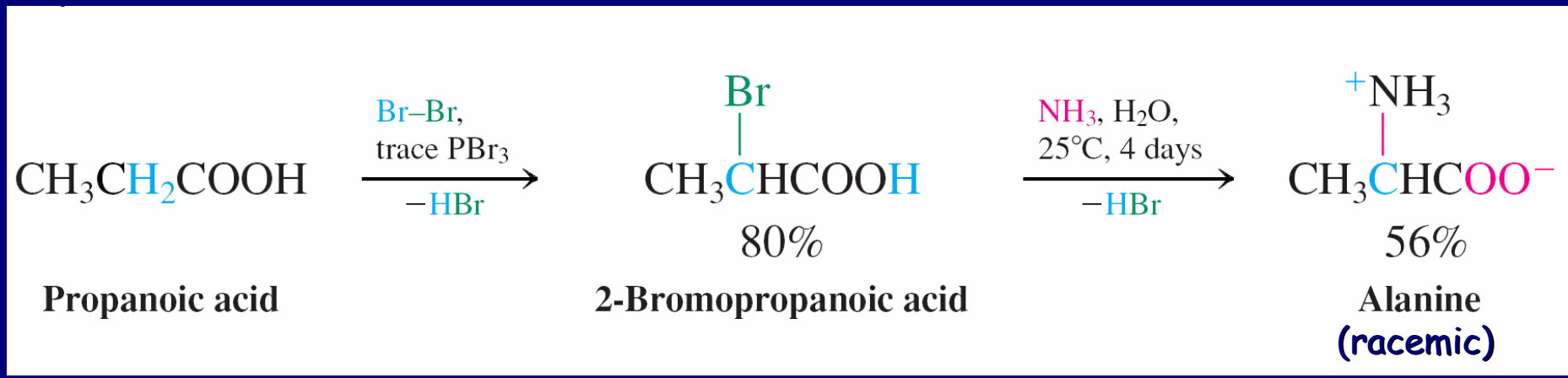
$$[H_3\overset{+}{N}CHCOOH] = [H_2NCHCOO^-]$$

(R) (R)

At this point pH = $pI = \frac{pK_{a\text{-}COOH} + pK_{a\text{-}\overset{+}{N}H_2H}}{2}$

Synthesis of Amino Acids

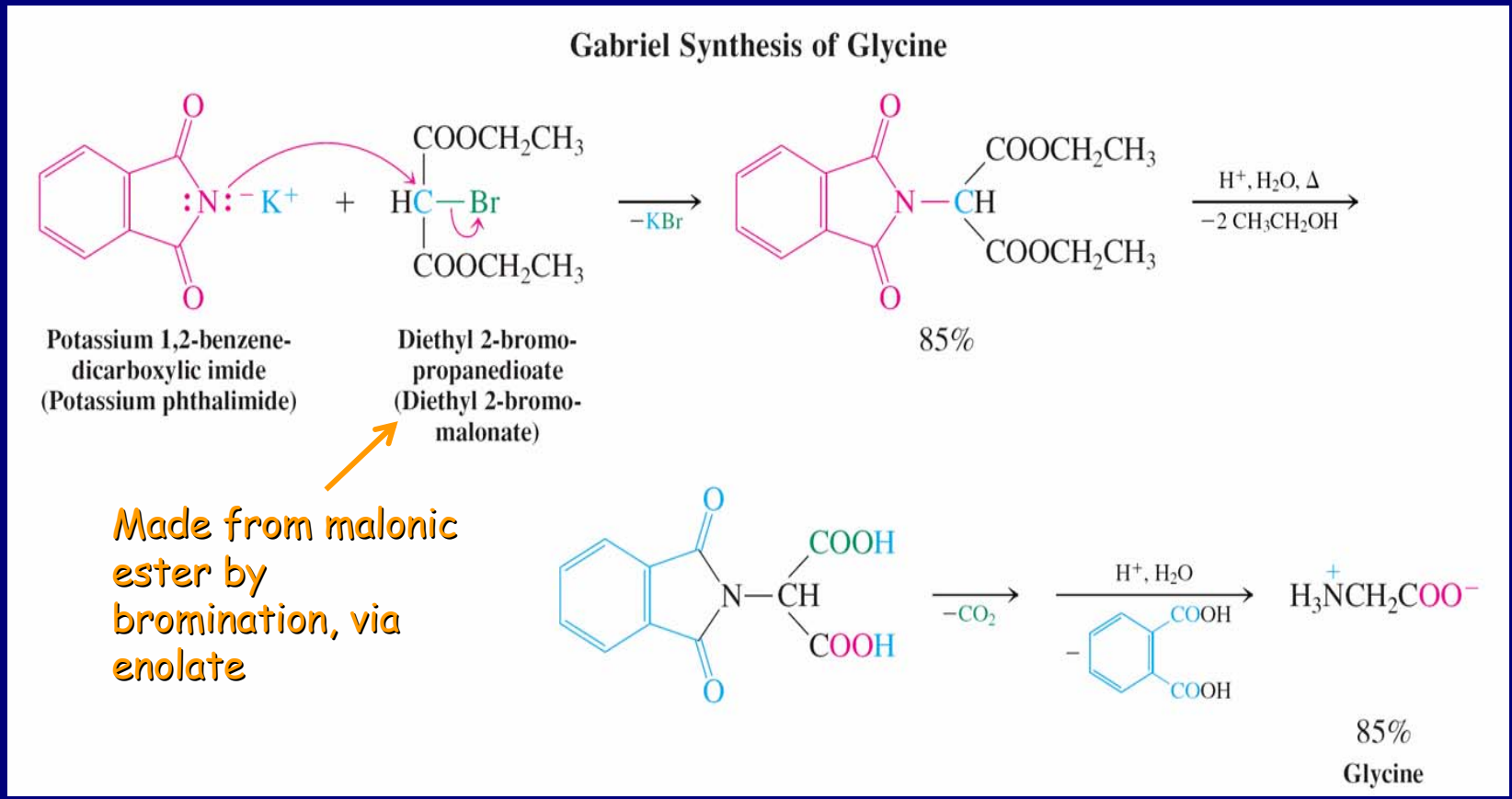
1. Hell-Volhard-Zelinsky, then Amination



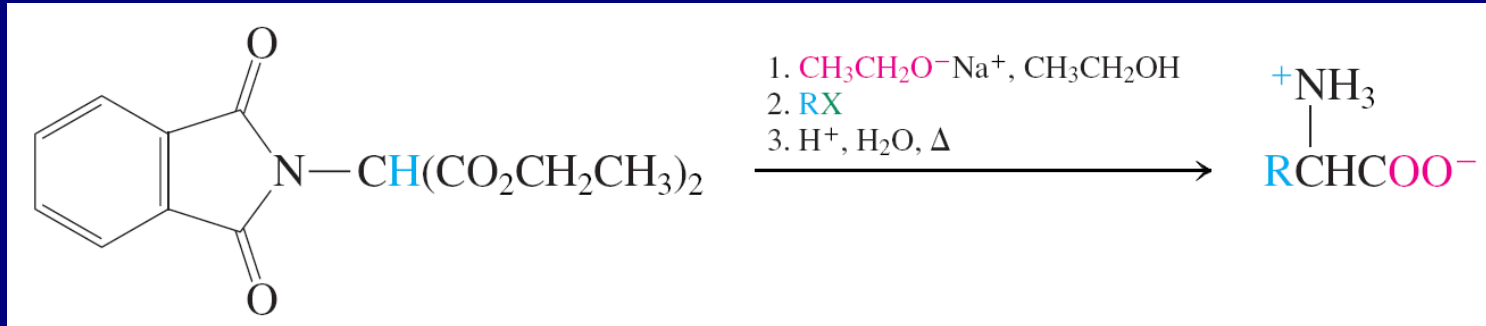
Yields not great

Better: Gabriel Synthesis ($RX \rightarrow RNH_2$)

2. Gabriel Synthesis

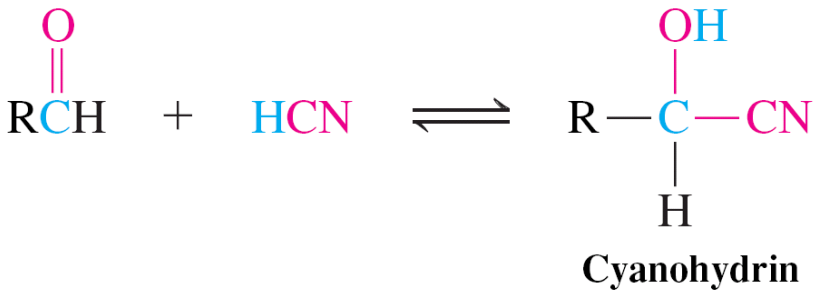


General:



3. Strecker Synthesis

Recall:

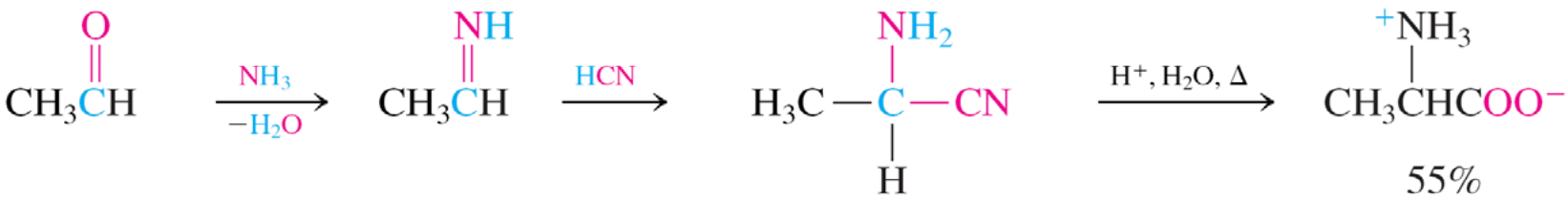


Cyanohydrin

Adolf Strecker
(1822-1871)

When HCN is used in the presence of ammonia, e.g. NH_4CN or $NH_4Cl/NaCN$, we get the corresponding amino cyanide:

Strecker Synthesis of Alanine

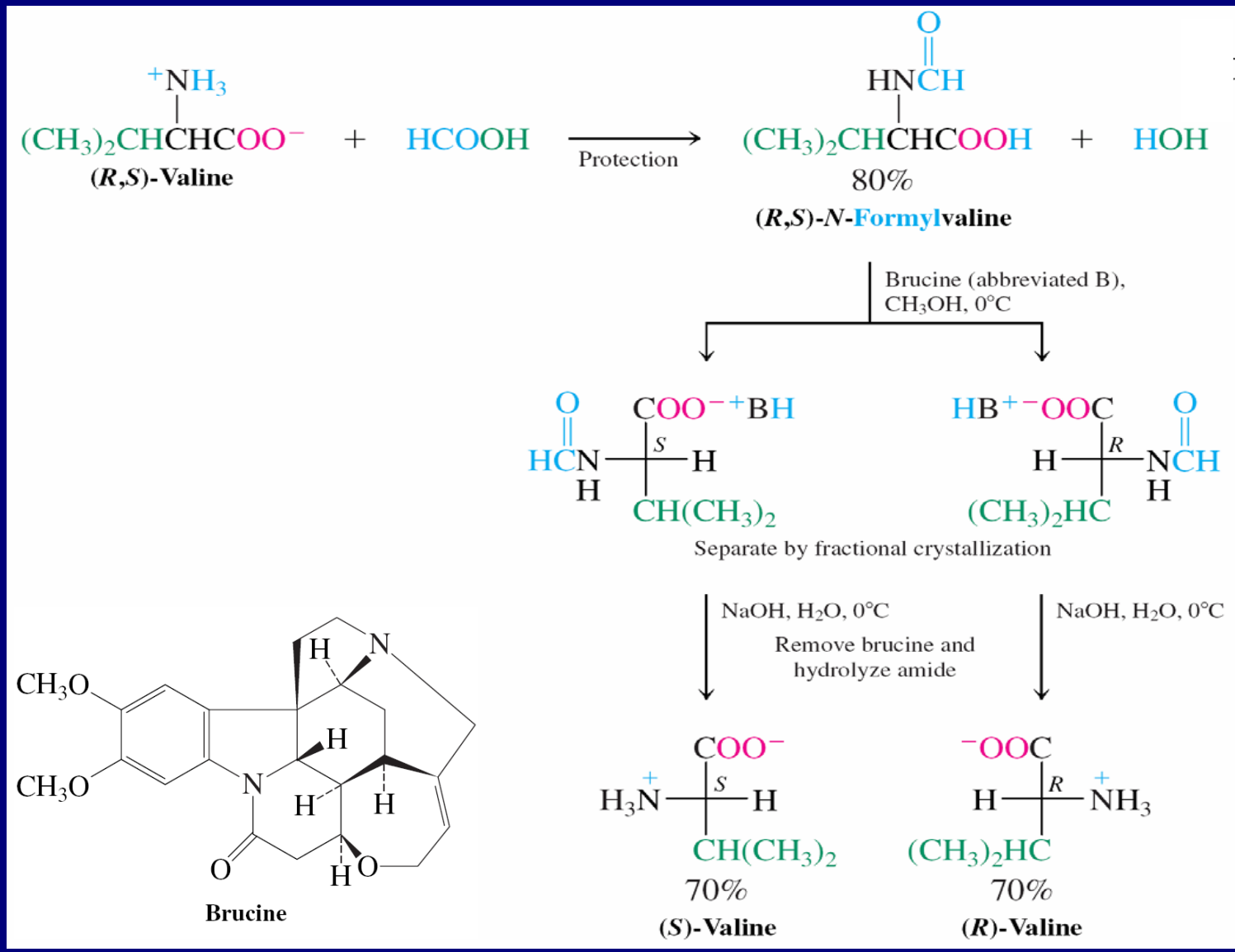


Acetaldehyde — Imine — 2-Aminopropanenitrile — Alanine

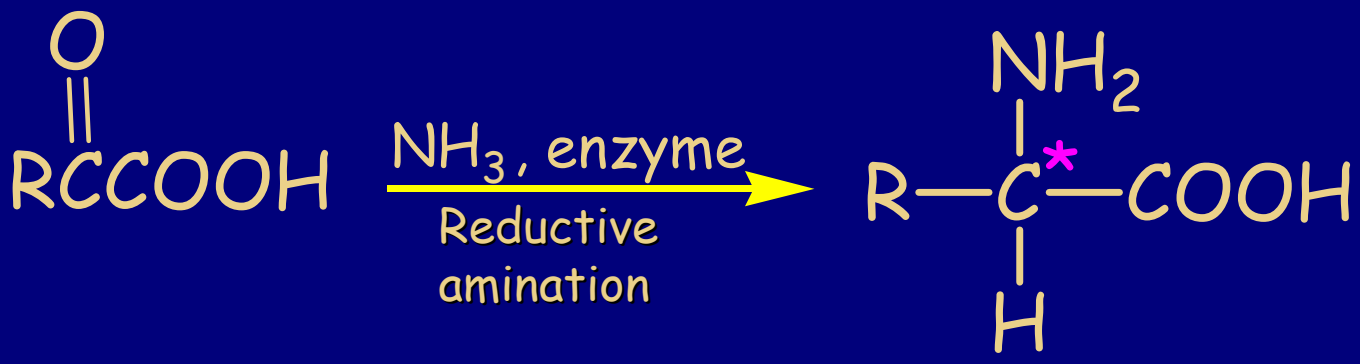
Segovia Strecker

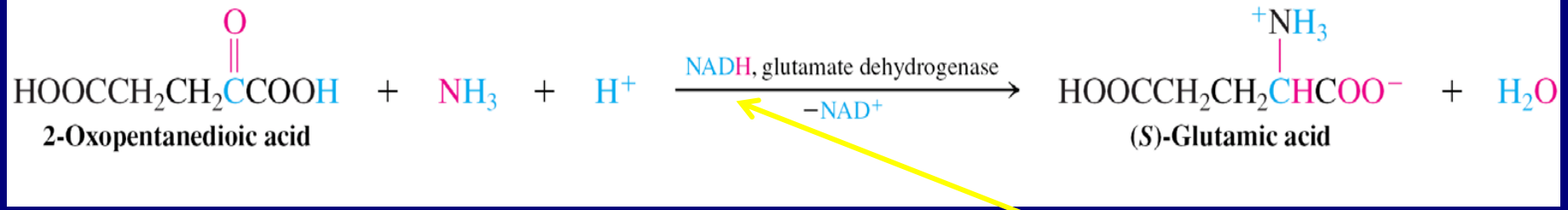
4. Synthesis of Enantiomerically Pure Amino Acids

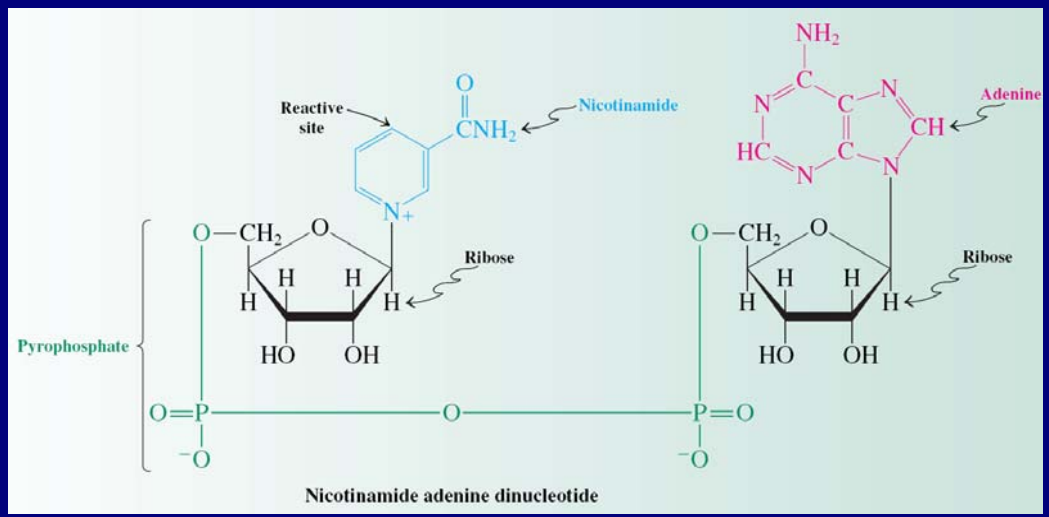
a. Resolution



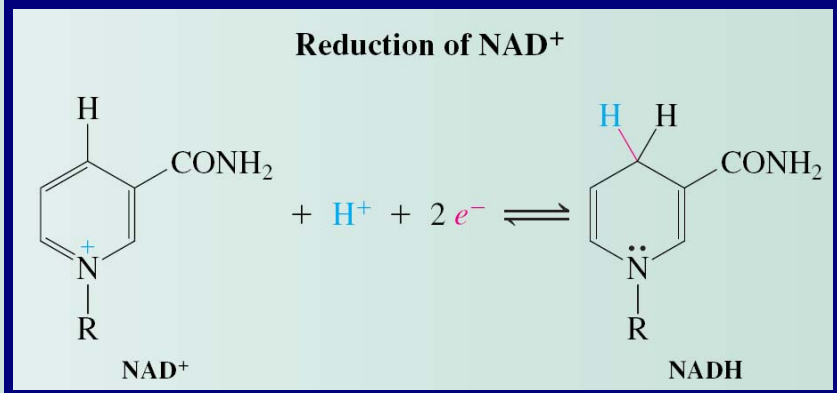
b. Biological methods



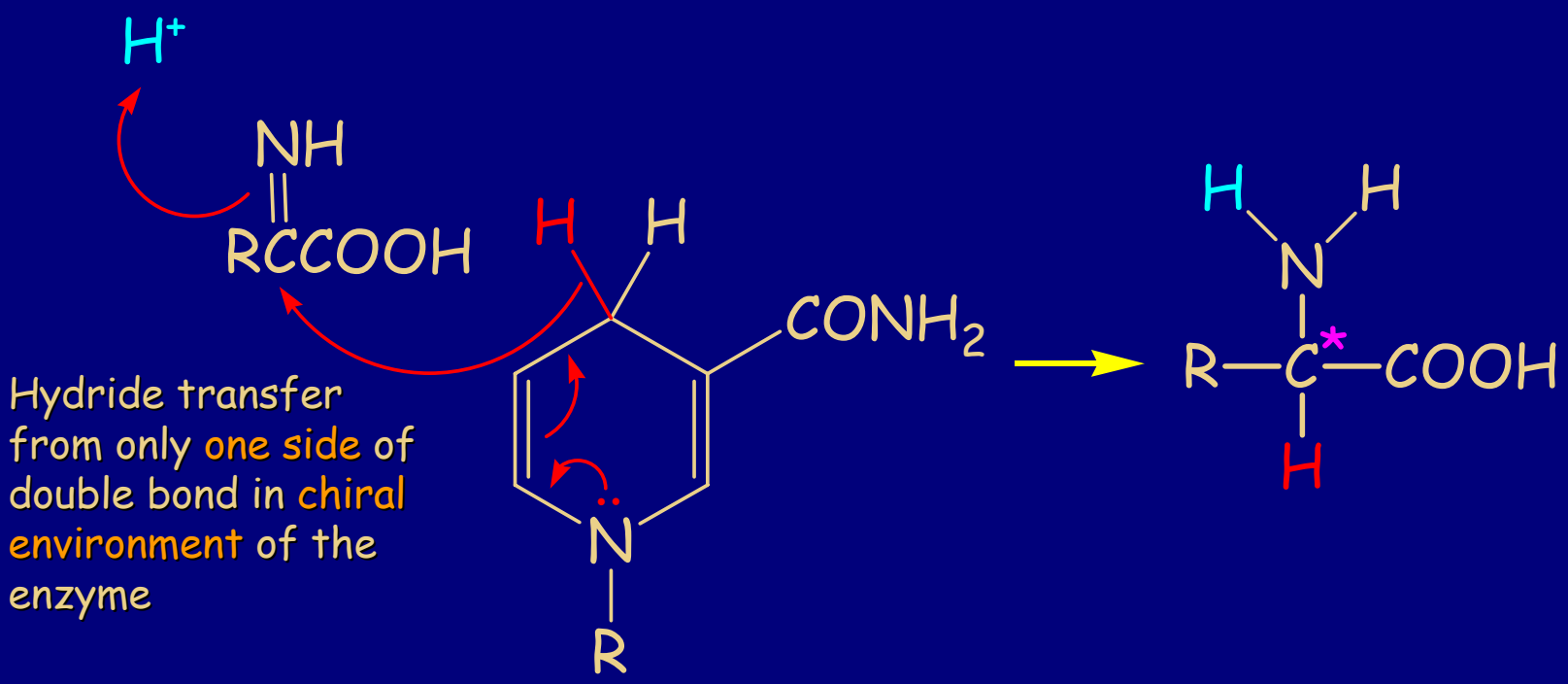




Oxidation

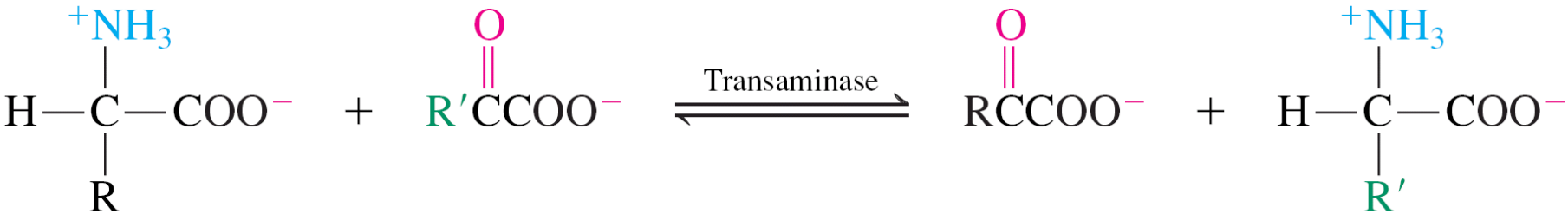


Mechanism:

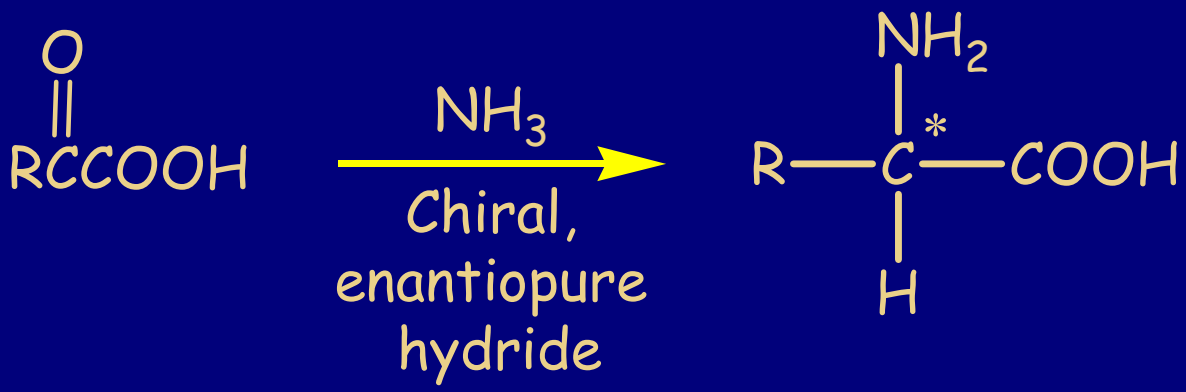


Hydride transfer from only one side of double bond in chiral environment of the enzyme

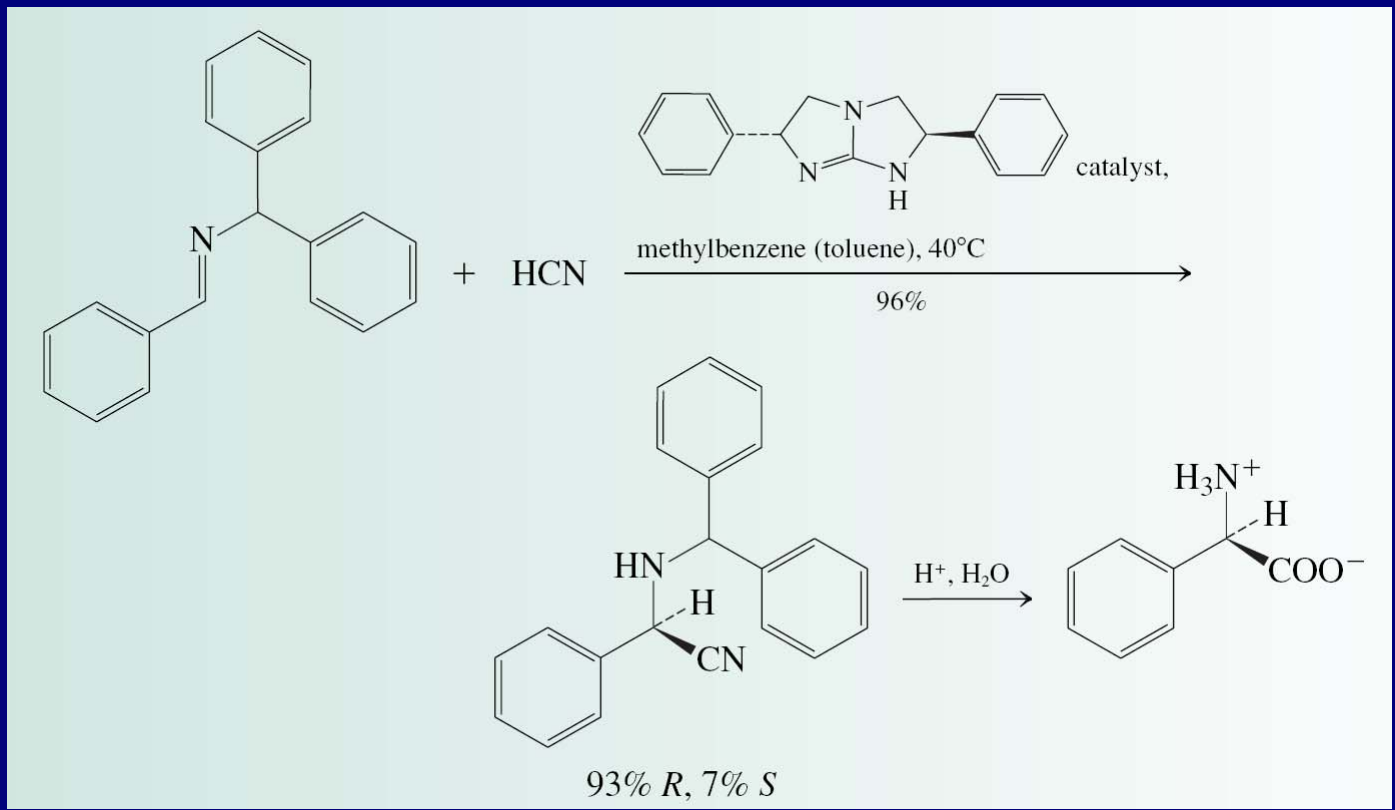
Glutamic acid aminates other α-oxoacids by redox exchange:



c. Using Chiral Auxiliaries

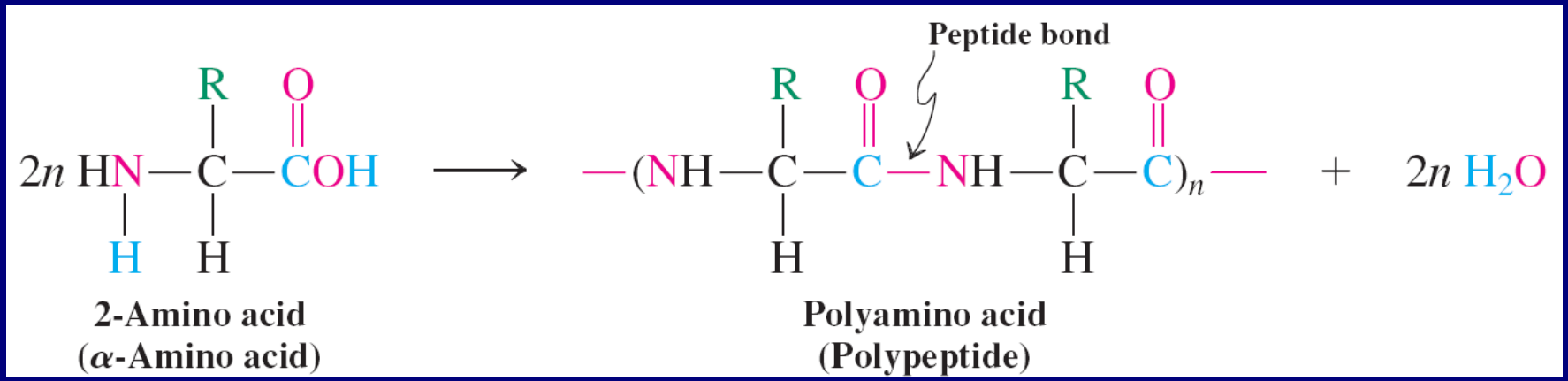


Or enantio-selective Strecker synthesis



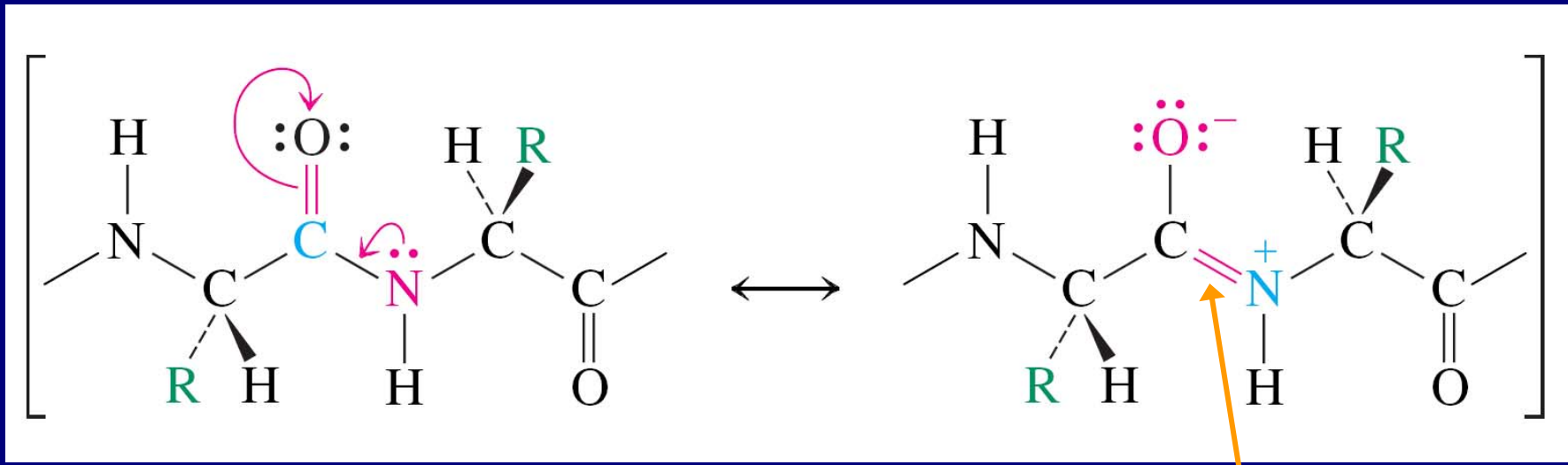
Peptides

Amino acids form peptide bonds



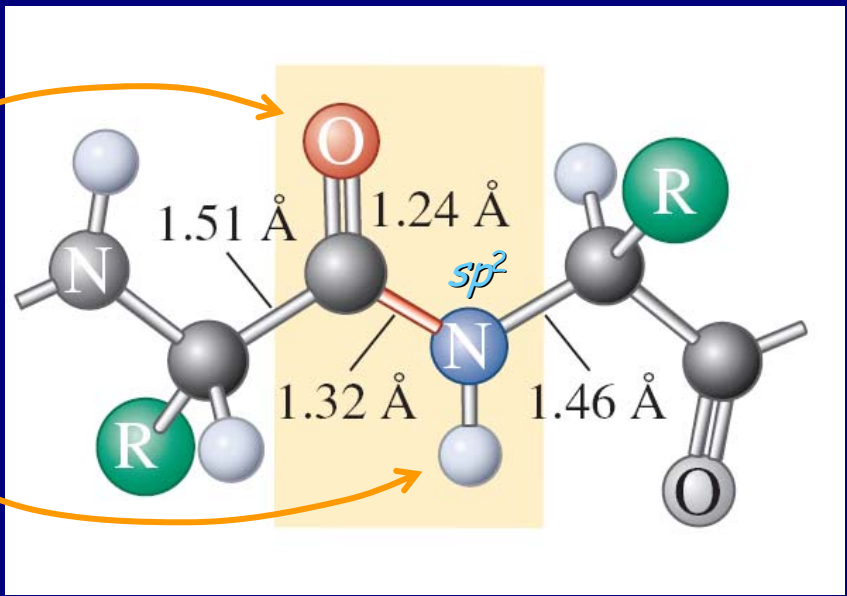
Dimer = dipeptide, trimer = tripeptide, and so on. The chain is arranged in space by H bonding, electrostatic attractions, hydrophobic-hydrophilic interactions (with water), and rigidity of the amide bond.

Rigidity and planarity of the peptide bond

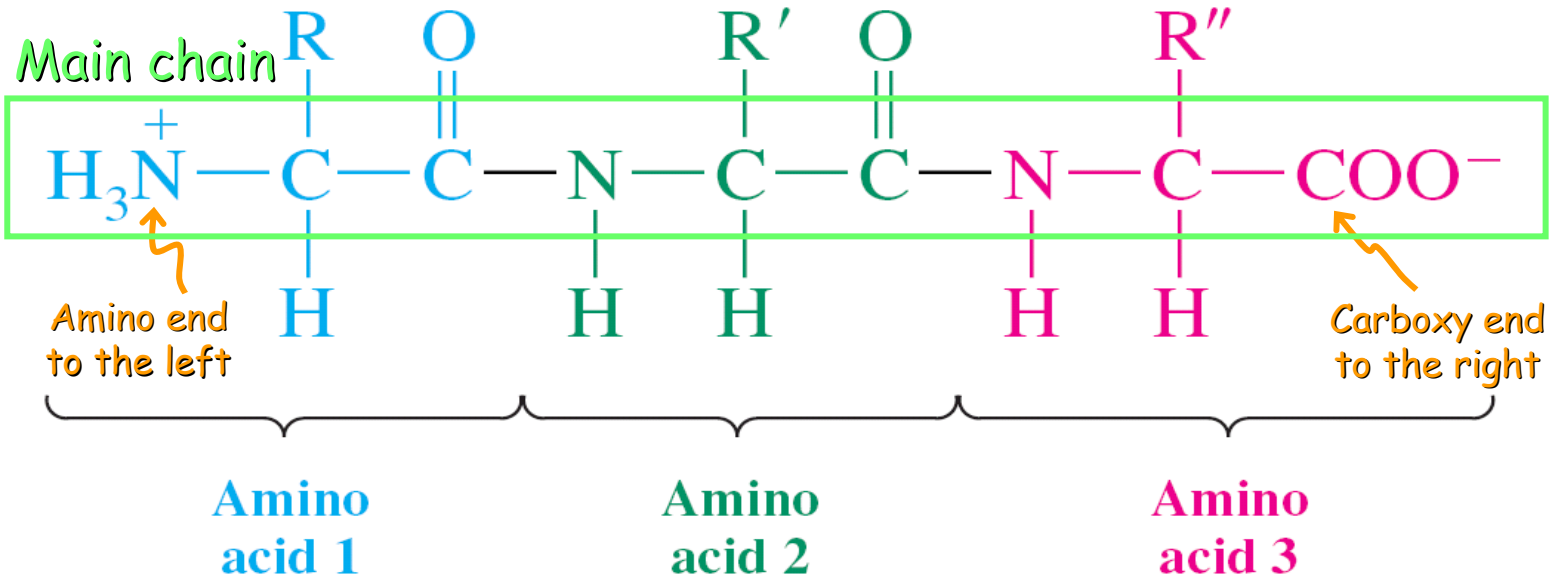


Rigid

Trans



How to Draw the Structure of a Tripeptide

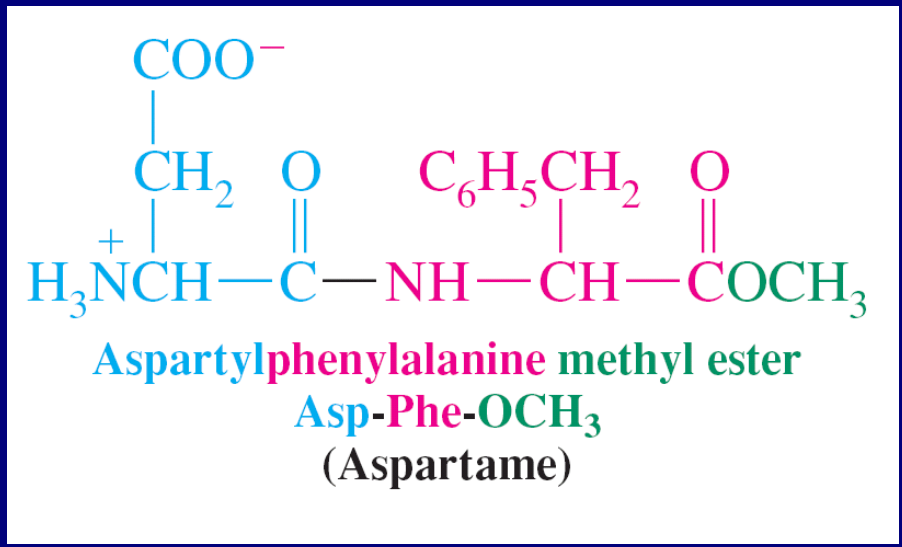


R, R′, R″, etc. are called the side chains.
All stereocenters are assumed to be *S*.

Names: Line up the amino acid names in sequence, from left to right, each ending in "yl", until reaching the terminus.

Three letter code

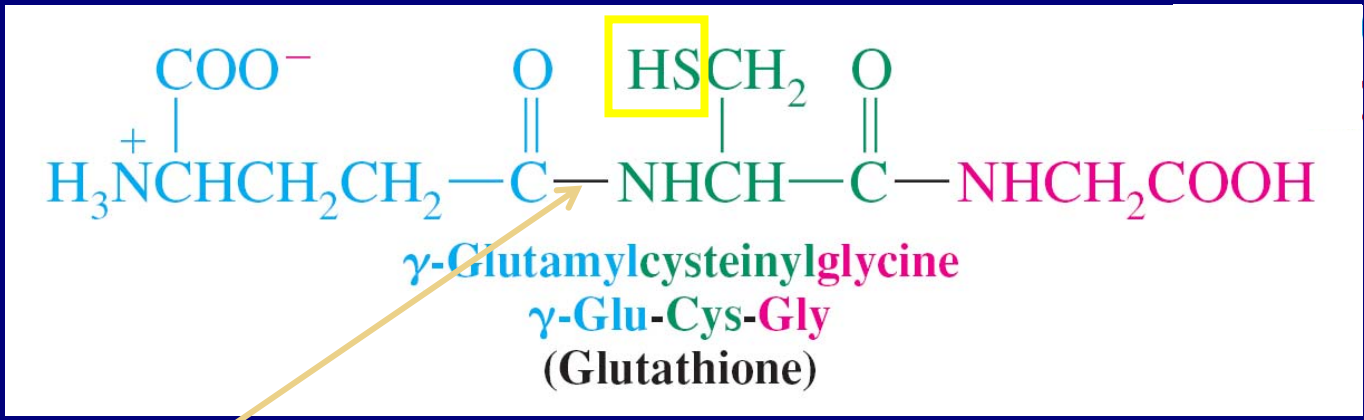
Examples of peptides



Nutrasweet

200 times as sweet as table sugar, only 4 cal/g vs 4 kcal/g. Compare: Fat 9 kcal/g, chocolate 3 kcal/g.

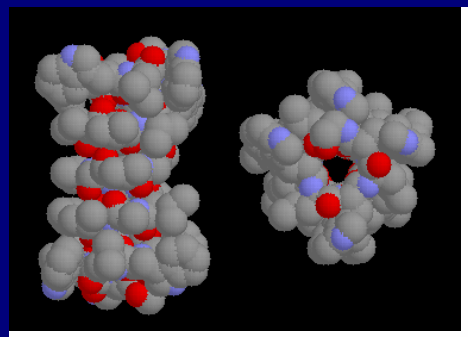
Glutathione



Not a typical peptide bond: Glutamic acid residue makes amide bond with the γ-carboxy group (γ-Glu).

Found in all living cells, particularly in the lens of the eye. Functions as a biological reducing agent by enzymatic oxidation of the cysteine mercapto unit to the disulfide-bridged dimer.

Gramicidine S: Antibiotic (Eye Infection)



Two identical pentapeptides joined head to tail.

Insulin: For diabetes. Disulfide bonds

A chain

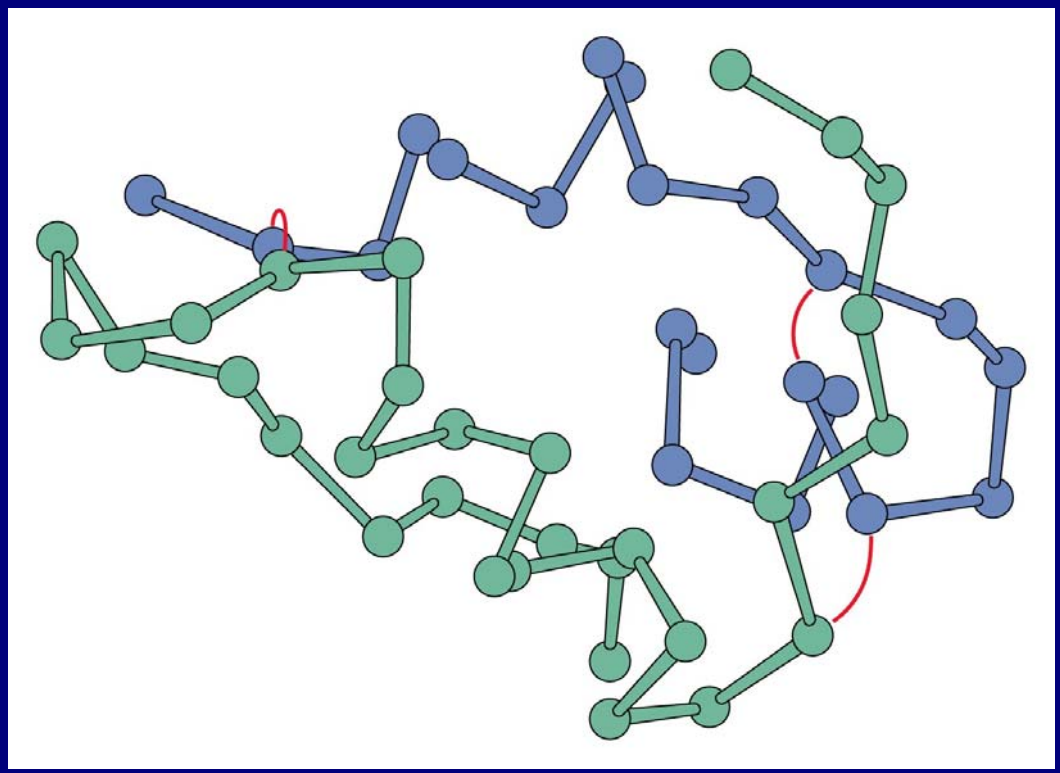
Gly-Ile-Val-Glu-Gln-Cys-Cys-Ala-Ser-Val-Cys-Ser-Leu-Tyr-Gln-Leu-Glu-Asn-Tyr-Cys-Asn

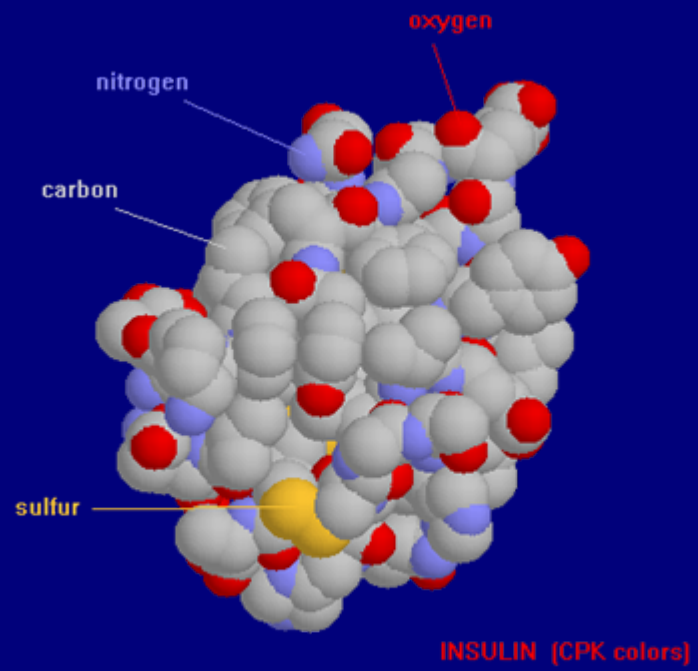
5 10 15 21

B chain

Phe-Val-Asn-Gln-His-Leu-Cys-Gly-Ser-His-Leu-Val-Glu-Ala-Leu-Tyr-Leu-Val-Cys-Gly-Glu-Arg-Gly-Phe-Phe-Tyr-Thr-Pro-Lys-Ala

5 10 15 20 25 30

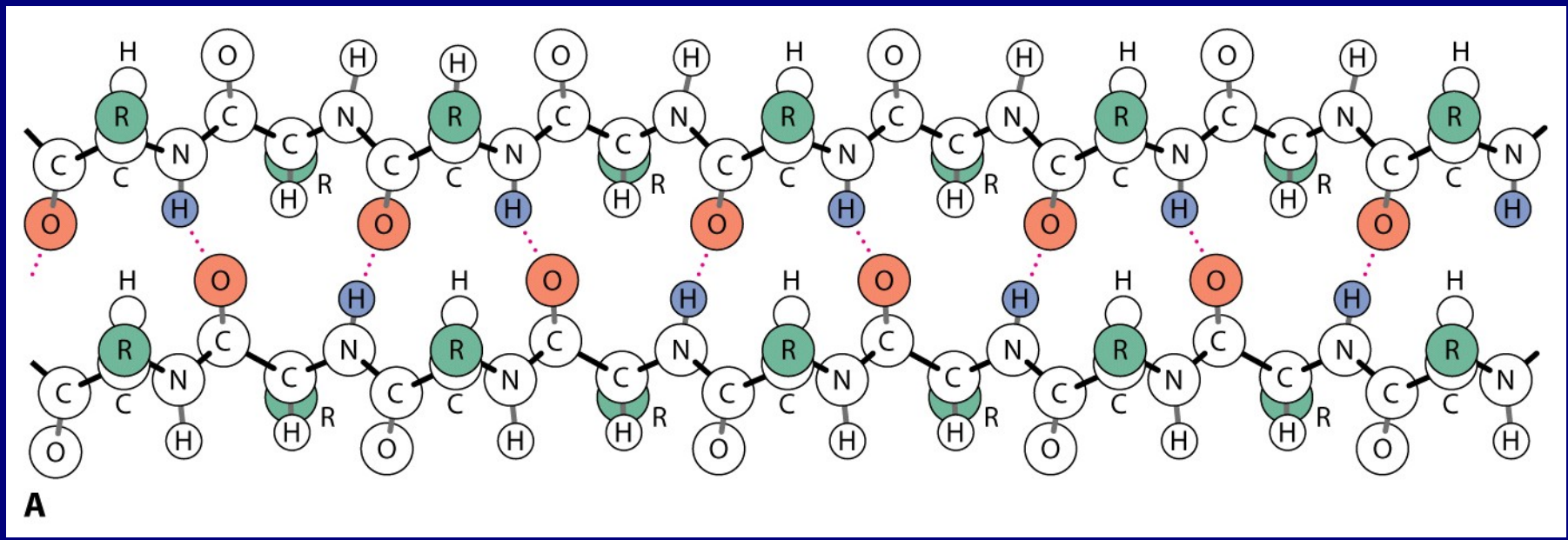




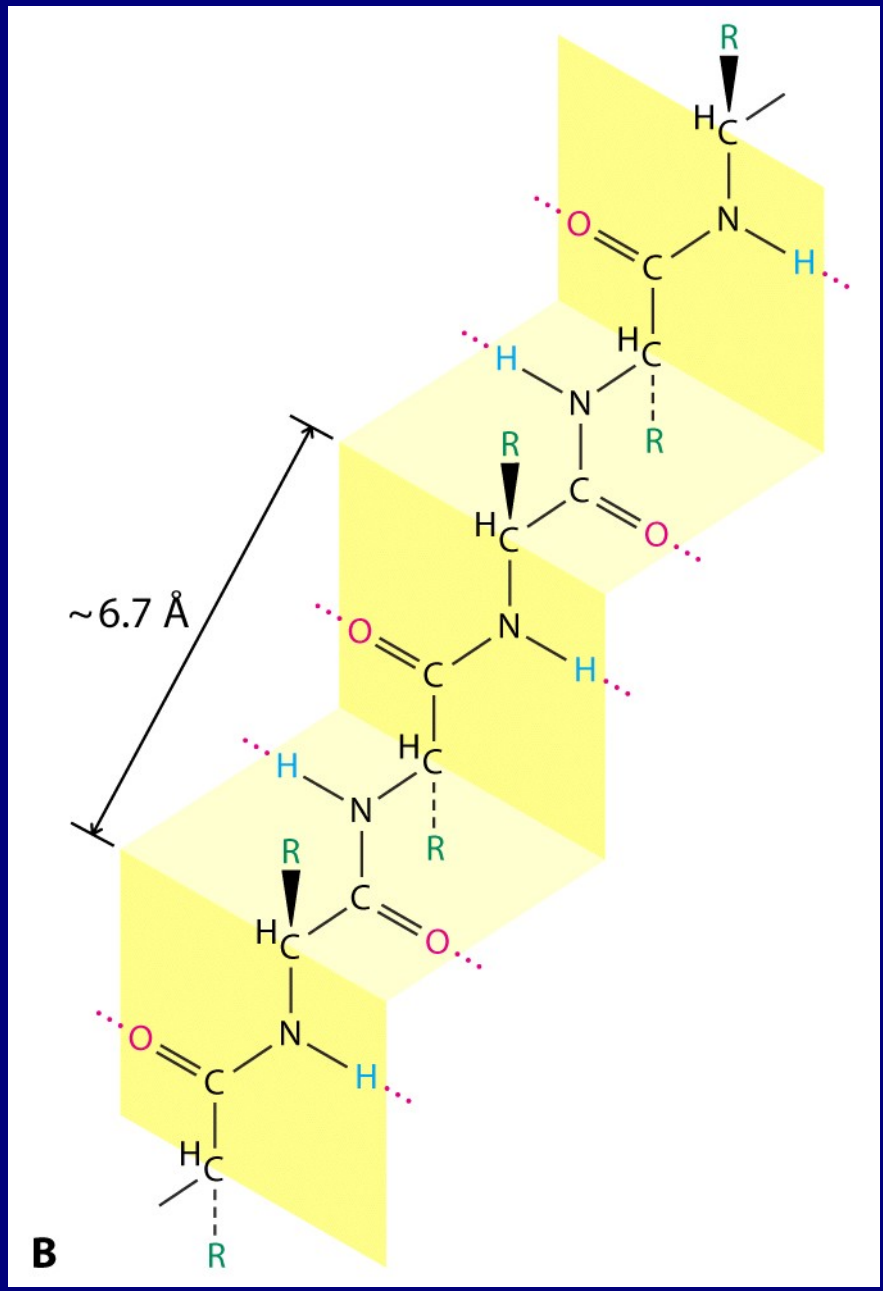
The amino acid sequence is called the primary structure, the three dimensional arrangement is the secondary, tertary, and quaternary structure.

Secondary Structure: H-Bonding

Pleated Sheet Structure

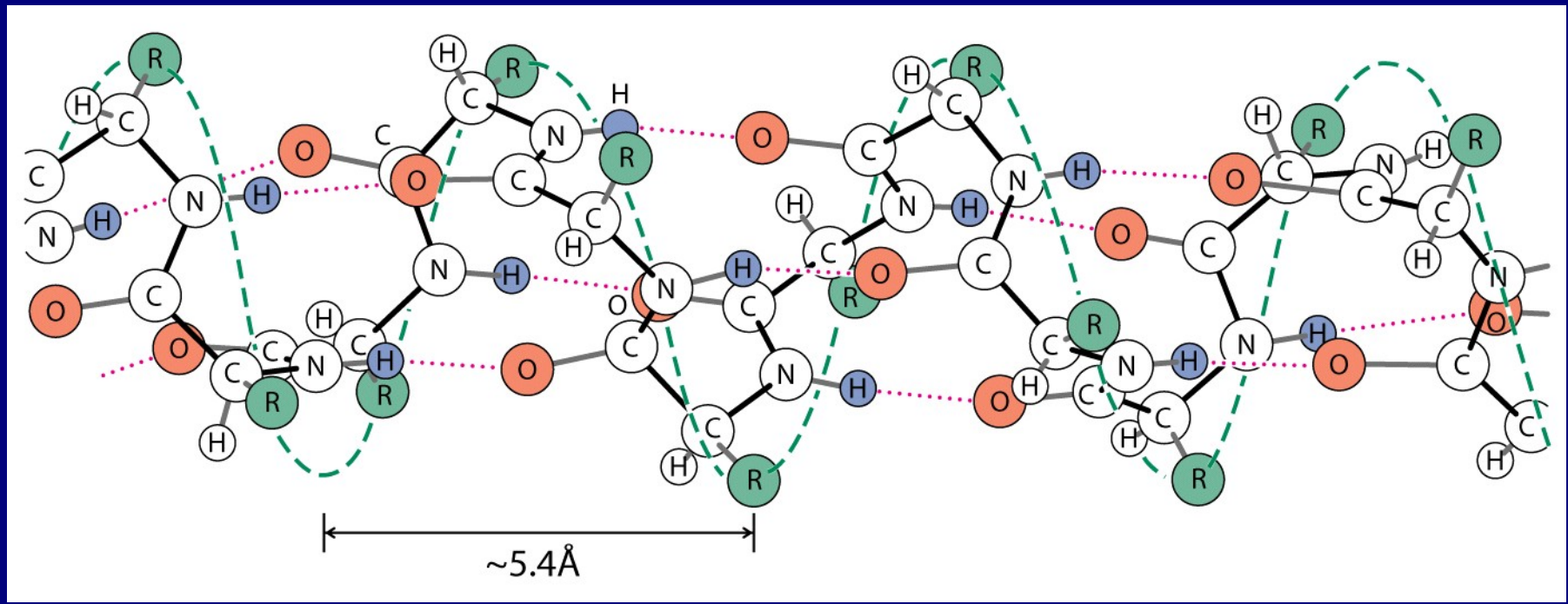


Sheets defined by shaded areas



α - Helix

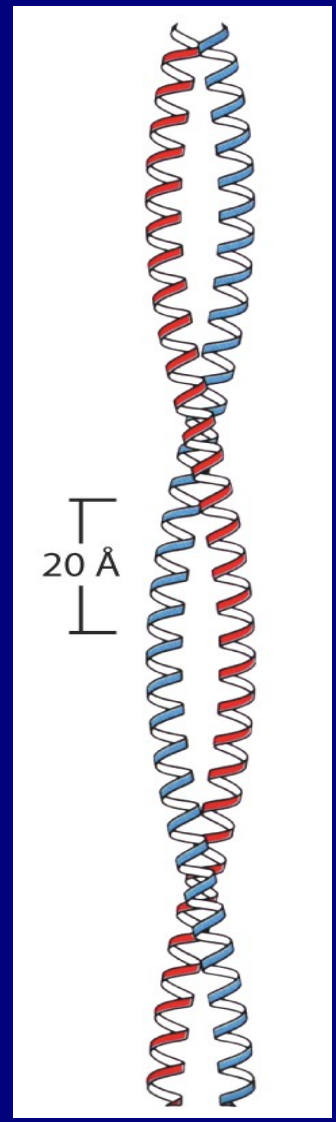
Right-handed spiral held by intramolecular H bonds



3.6 Amino acids per turn; repeat distance 5.4 Å

Tertiary Structure: Further folding, coiling, and aggregation. Denaturation is the breakdown of this structure.

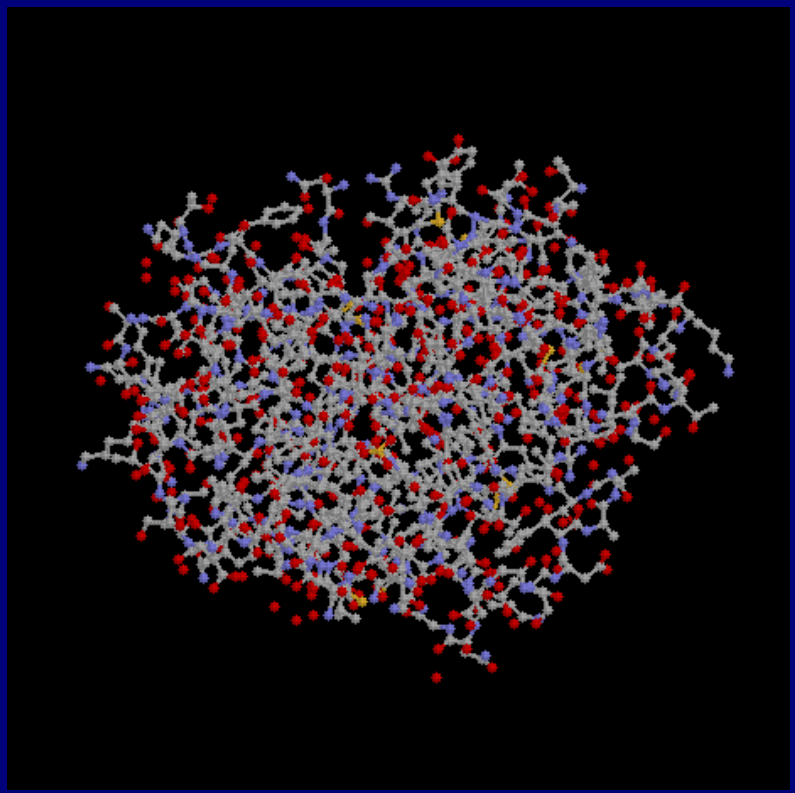
Example:
Superhelix



Typical of fibrous proteins, such as myosin (in muscle), fibrin (in blood clots), and a-keratin (in hair, nails, and wool).

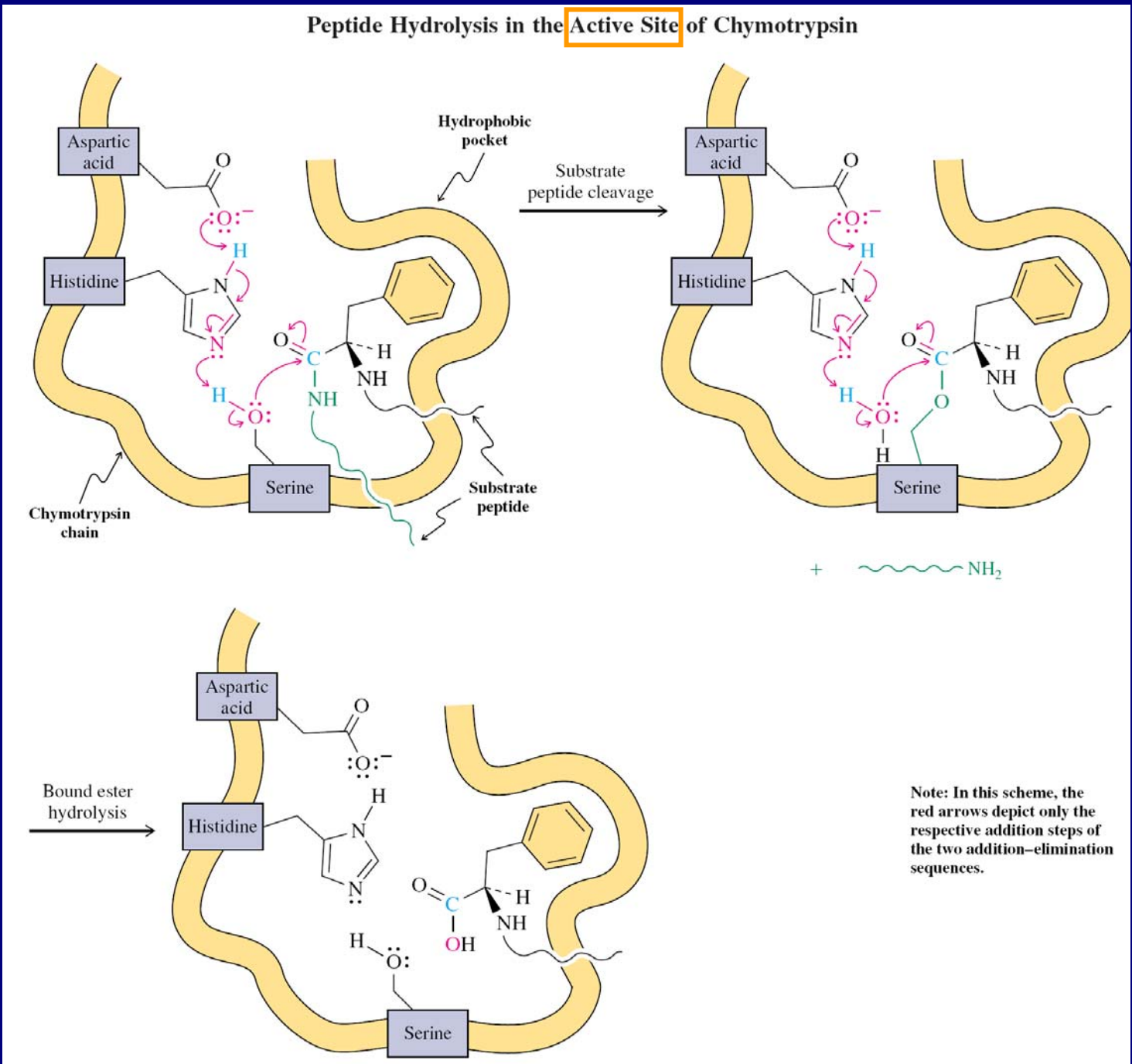
Tertiary structure gives rise to pockets: Active sites or binding sites that provide perfect fit for substrates, e.g., drugs.

Example: Digestive enzyme chymotrypsin



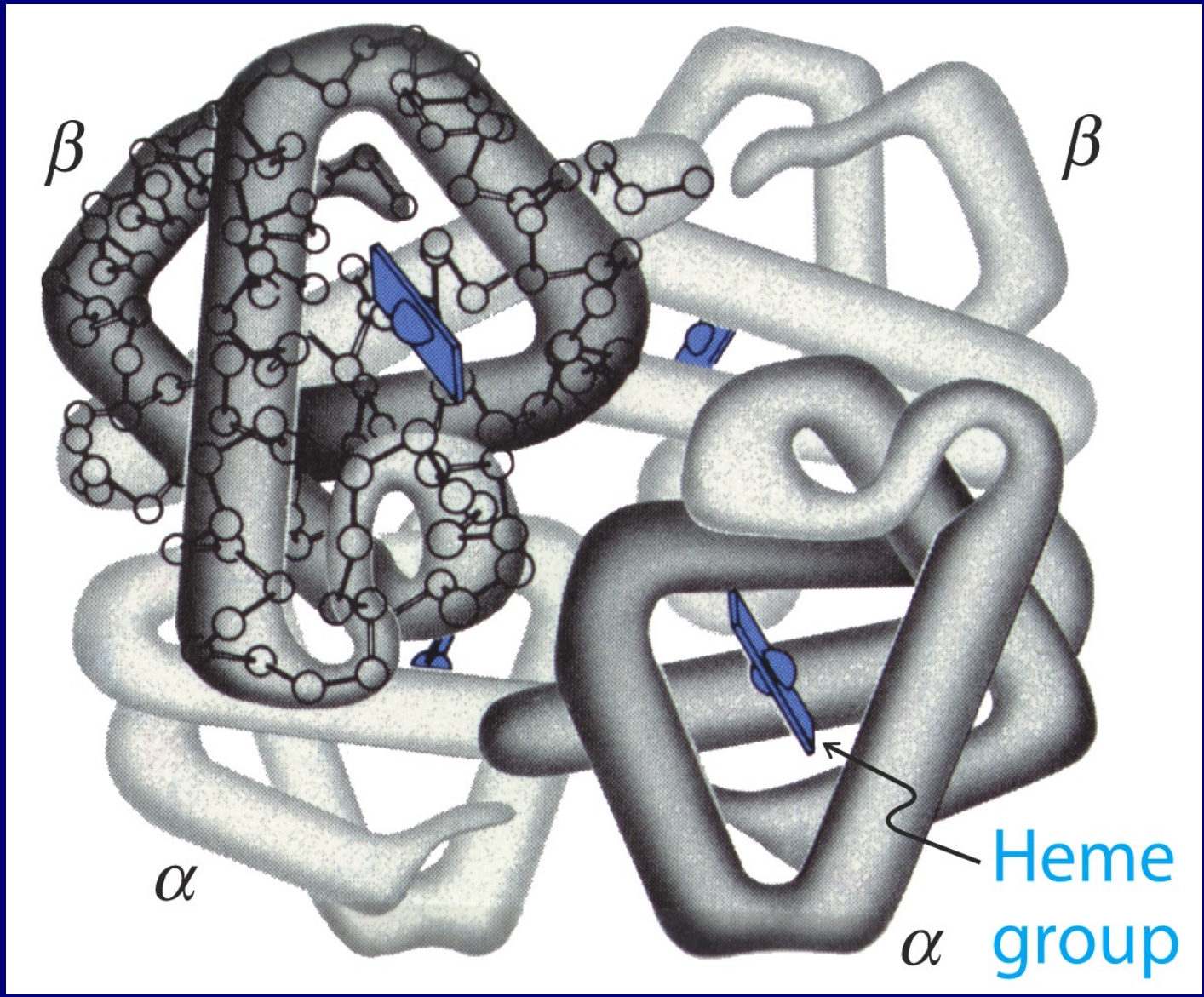
Protein digestion = Amide bond hydrolysis

Four cooperative effects to facilitate proton shuttle



Quaternary Structure: Aggregation of several units

Example:
Hemoglobin



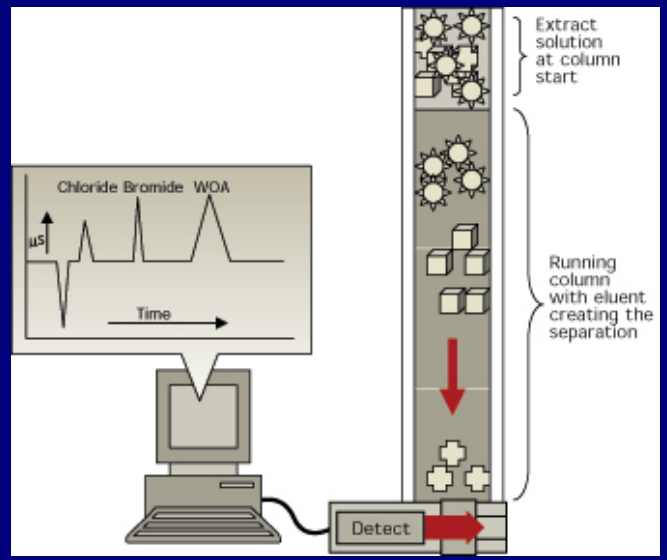
Protein (Polypeptide) Primary Structure Determination

Amino Acid Sequencing

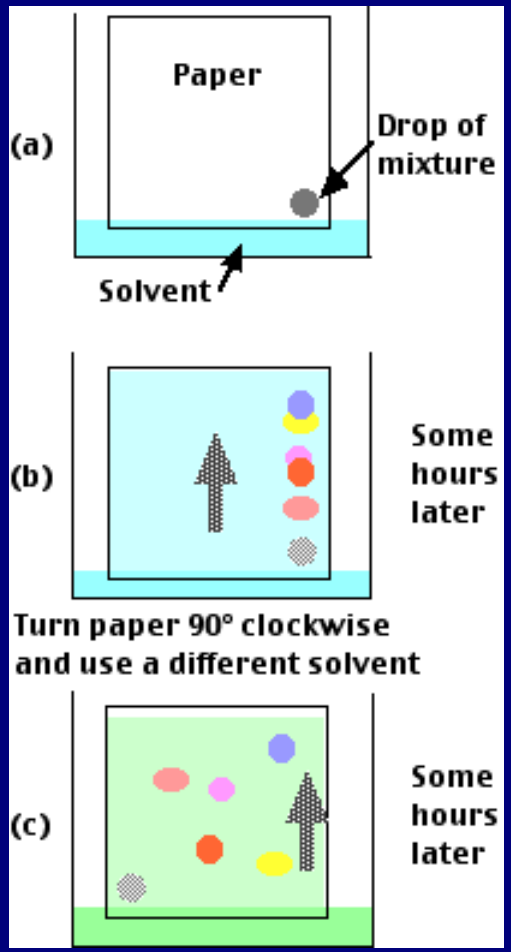
1. Break S-S bridges by oxidation
   → purify pieces

$$R^1\text{-}S\text{-}S\text{-}R^2 \longrightarrow R^1SO_3H + R^2SO_3H$$

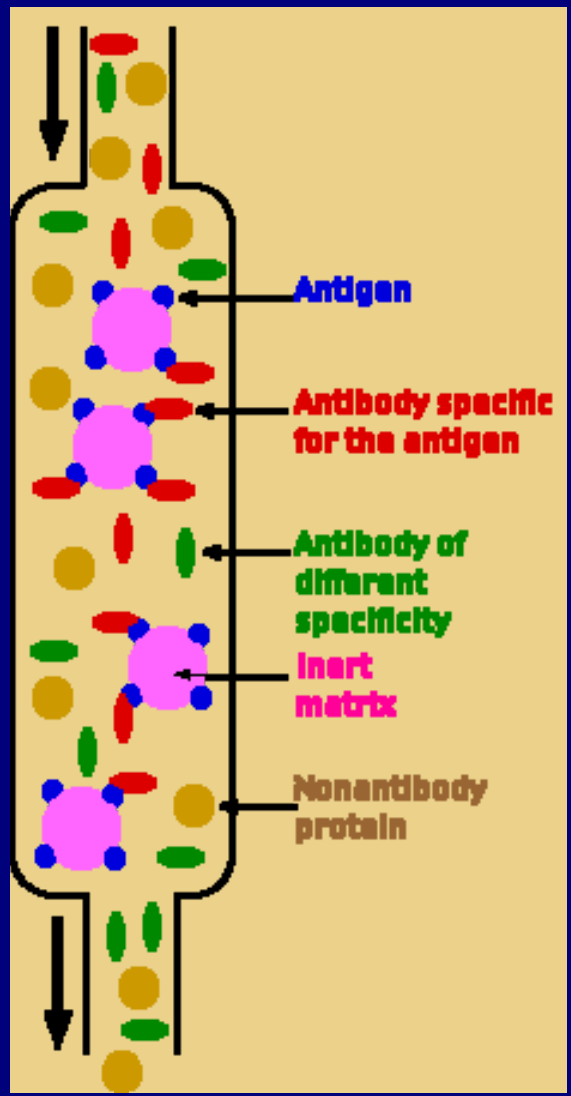
Purification is achieved by various forms of chromatography: Dialysis, gel-filtration, ion-exchange, electrophoresis, affinity chromatography.



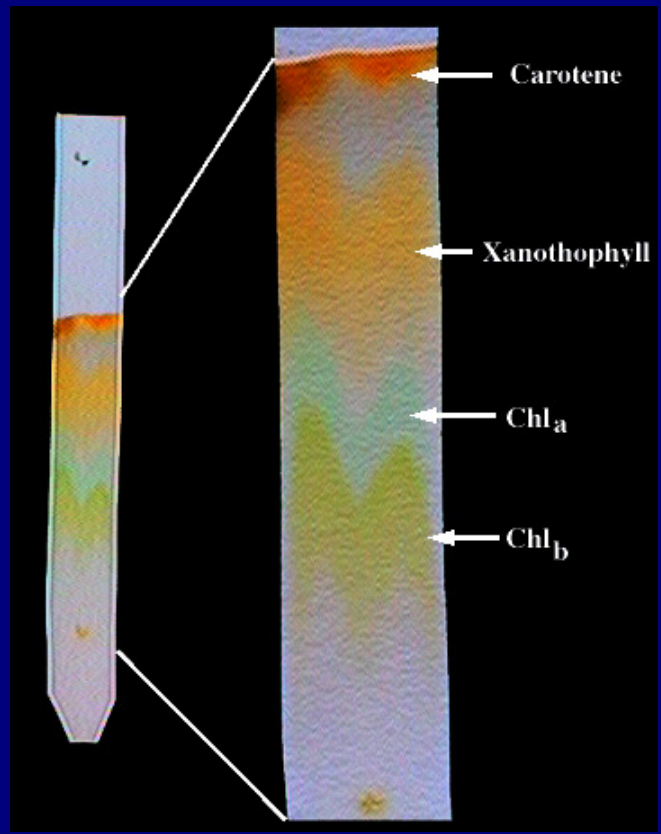
Chromatography



Paper



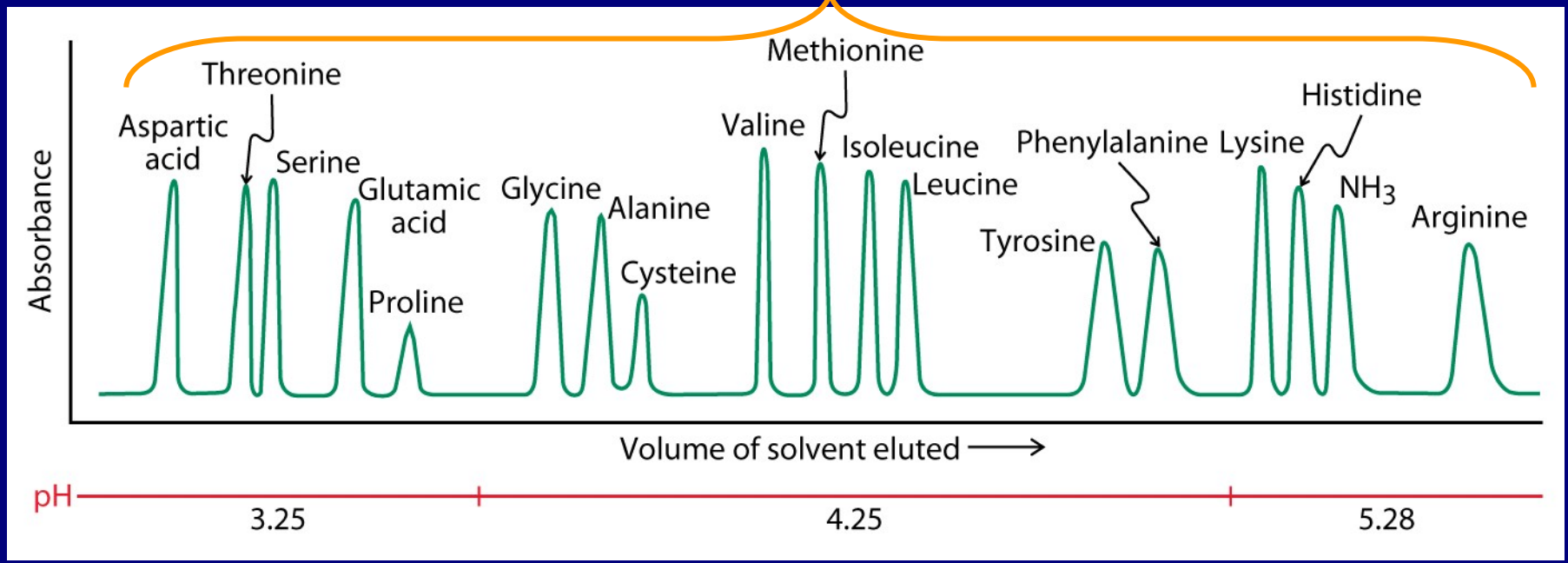
Affinity



Column

2. Amino Acid Analyzer

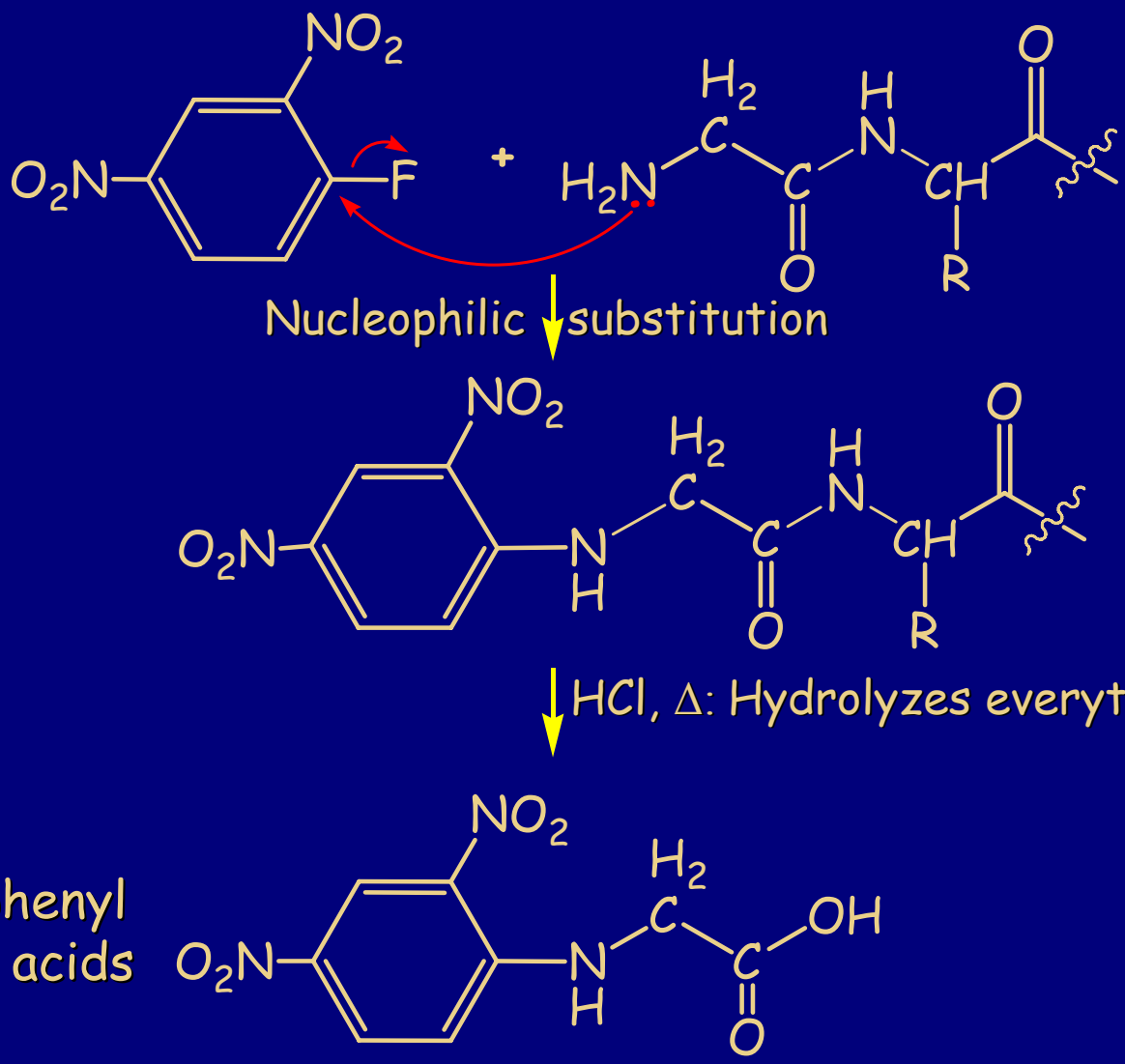
Used after hydrolysis of polypeptides to component amino acids (6 N HCl, 110°C, 24 h). Column chromatography separates and detects them.



Integration gives relative amount of type of amino acid. Now we know composition. What about the sequence?

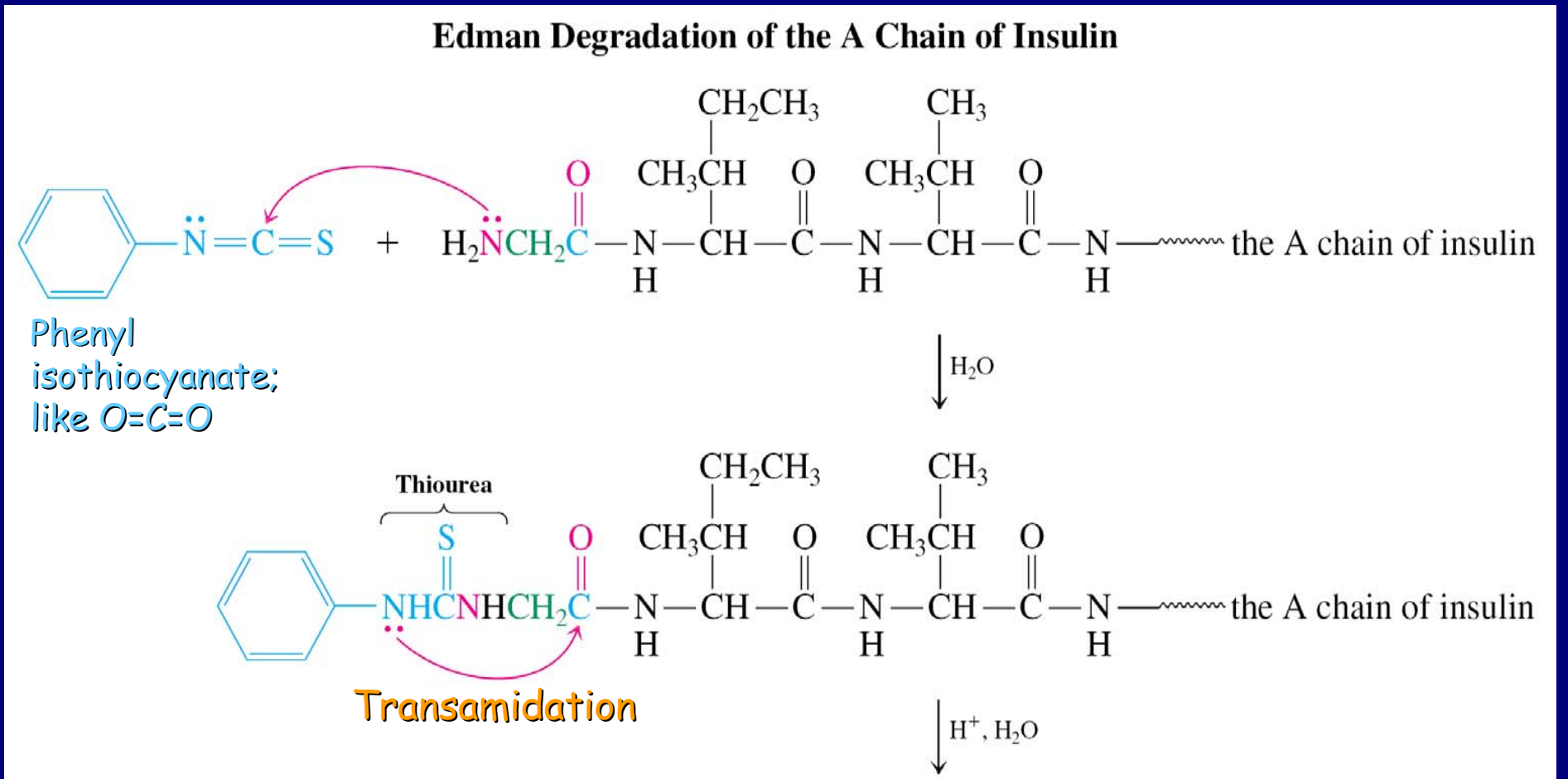
3. Amino Acid Sequencing: One by one

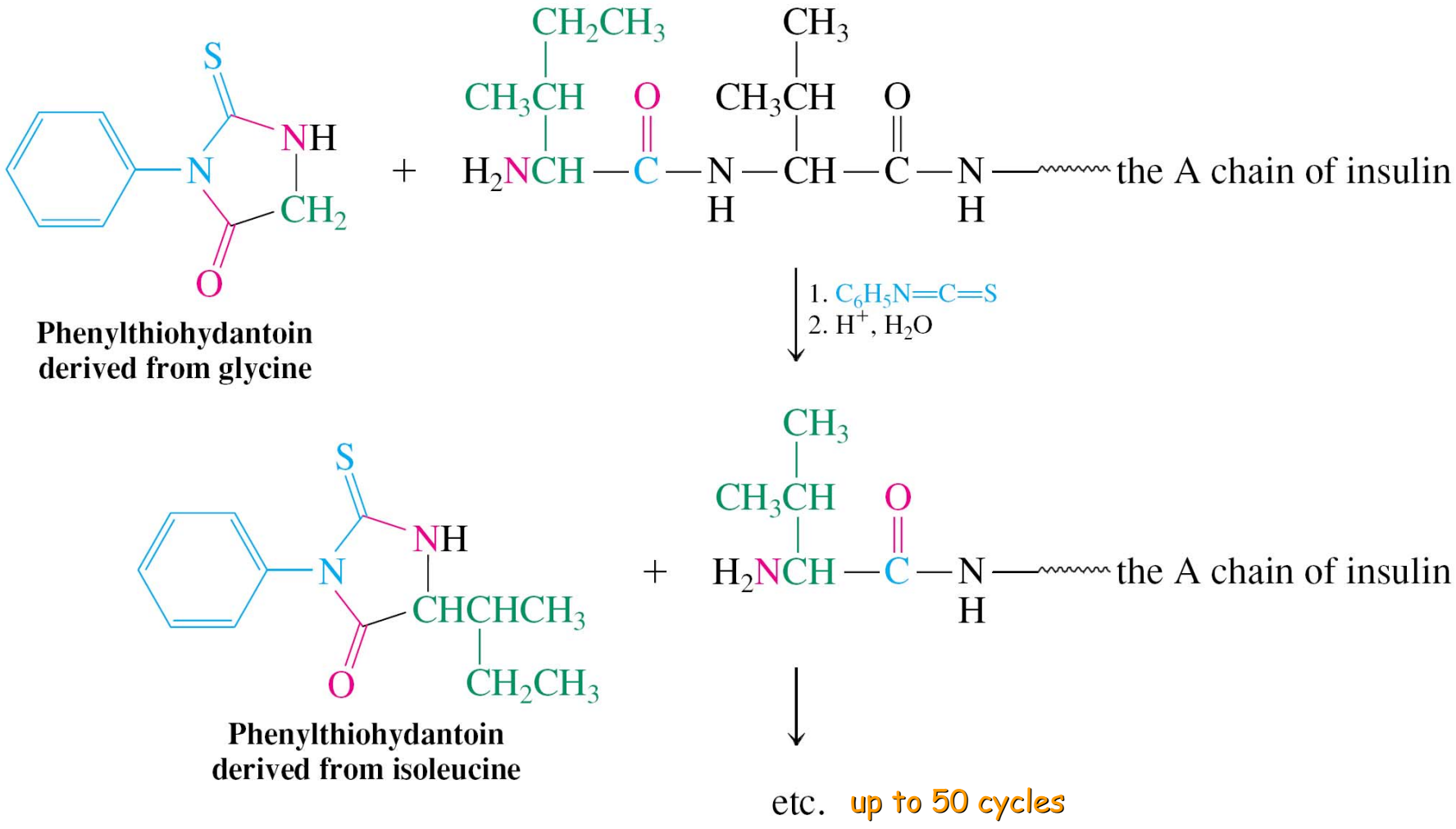
a. Sanger: Degrade from amino terminal end



Analyze: All dinitrophenyl derivatives of amino acids are known

b. Edman degradation– leaves rest of chain intact, allows iterative procedure





All phenylthiohydantoins of amino acids are known. Edman degradation turns unreliable after ~50 amino acids (build up of impurities). Problem solved by chopping up large peptides selectively into smaller (< 50 amino acids) pieces using enzymes.

Enzymatic Cleavage

Followed by finding overlap sequences to establish connectivity in original

Trypsin: Hydrolyzes only at carboxy end of Arg and Lys, e.g.:

Trp-Glu-Arg↯Phe-Phe-Lys↯Ala-Val

Chymotrypsin: Hydrolyzes only at carboxy end of Phe, Trp, Tyr, e.g.:

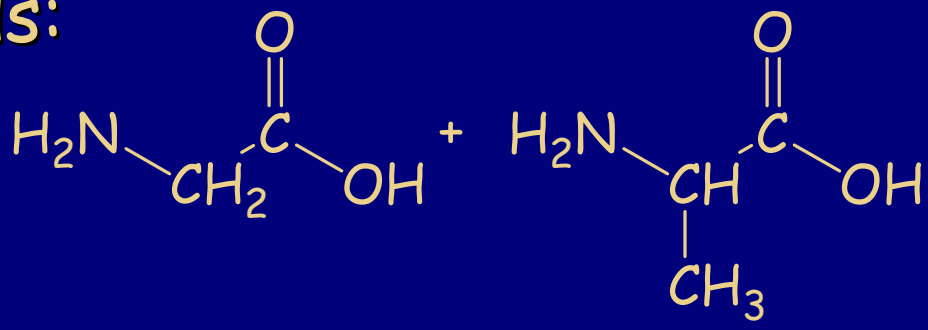
Trp↯Glu-Arg-Phe↯Phe↯Lys-Ala-Val

Thermolysin: Hydrolyzes amino end of Leu, Ile, Val, e.g.:

Trp-Glu-Arg-Phe-Phe-Lys-Ala↯Val

Synthesis of Polypeptides

Protecting groups: Why? We need selectivity in building up the peptide sequence. Consider the synthesis of glycylalanine by dehydration of the component amino acids:



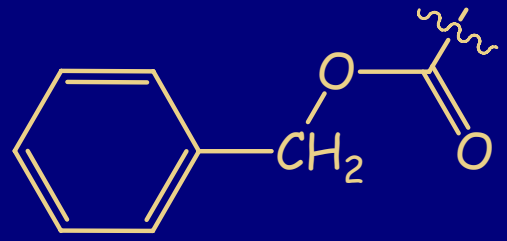
An Attempt at the Synthesis of Glycylalanine by Thermal Dehydration

Gly + Ala $\xrightarrow[-H_2O]{\Delta}$ Gly-Gly + Ala-Gly + Gly-Ala + Ala-Ala + Gly-Gly-Ala + Ala-Gly-Ala etc.

Desired product

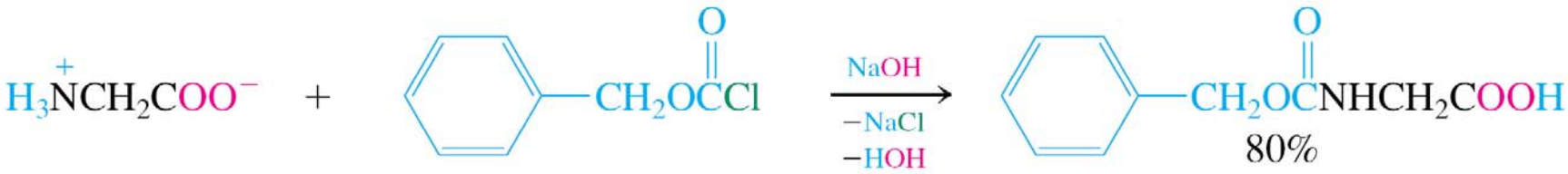
Hence, amino end of Gly and carboxy end of Ala have to be protected.

a. Amino end protection with



Phenylmethoxycarbonyl
(carbobenzoxy, Cbz)

Protection of the Amino Group in Glycine

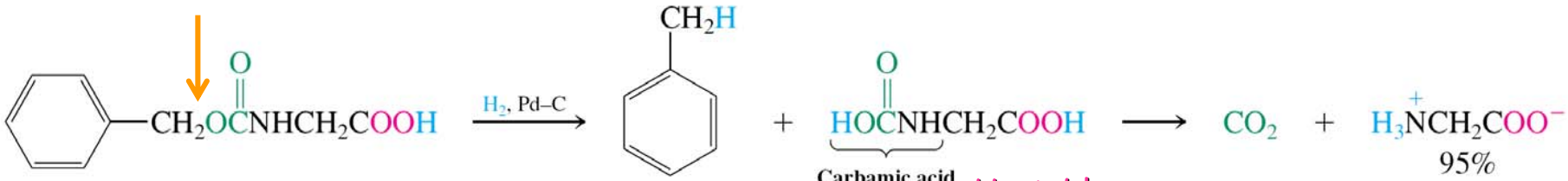


Glycine

Phenylmethyl chloroformate
(Benzyl chloroformate)

Phenylmethoxycarbonylglycine
(Carbobenzoxyglycine, Cbz-Gly)

Deprotection of the Amino Group in Glycine



Carbamic acid function

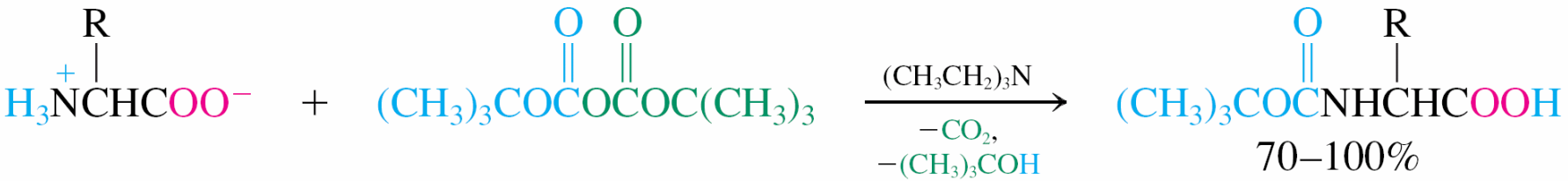
Unstable

Like the hydrogenolysis of benzylic ethers

b. Amino end protection with

Tert-butyloxycarbonyl (Boc)

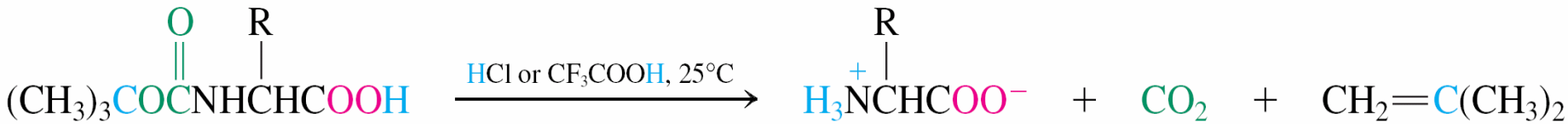
Protection of the Amino Group in Amino Acids as the Boc Derivative



Bis(1,1-dimethylethyl) dicarbonate (Di-*tert*-butyl dicarbonate)

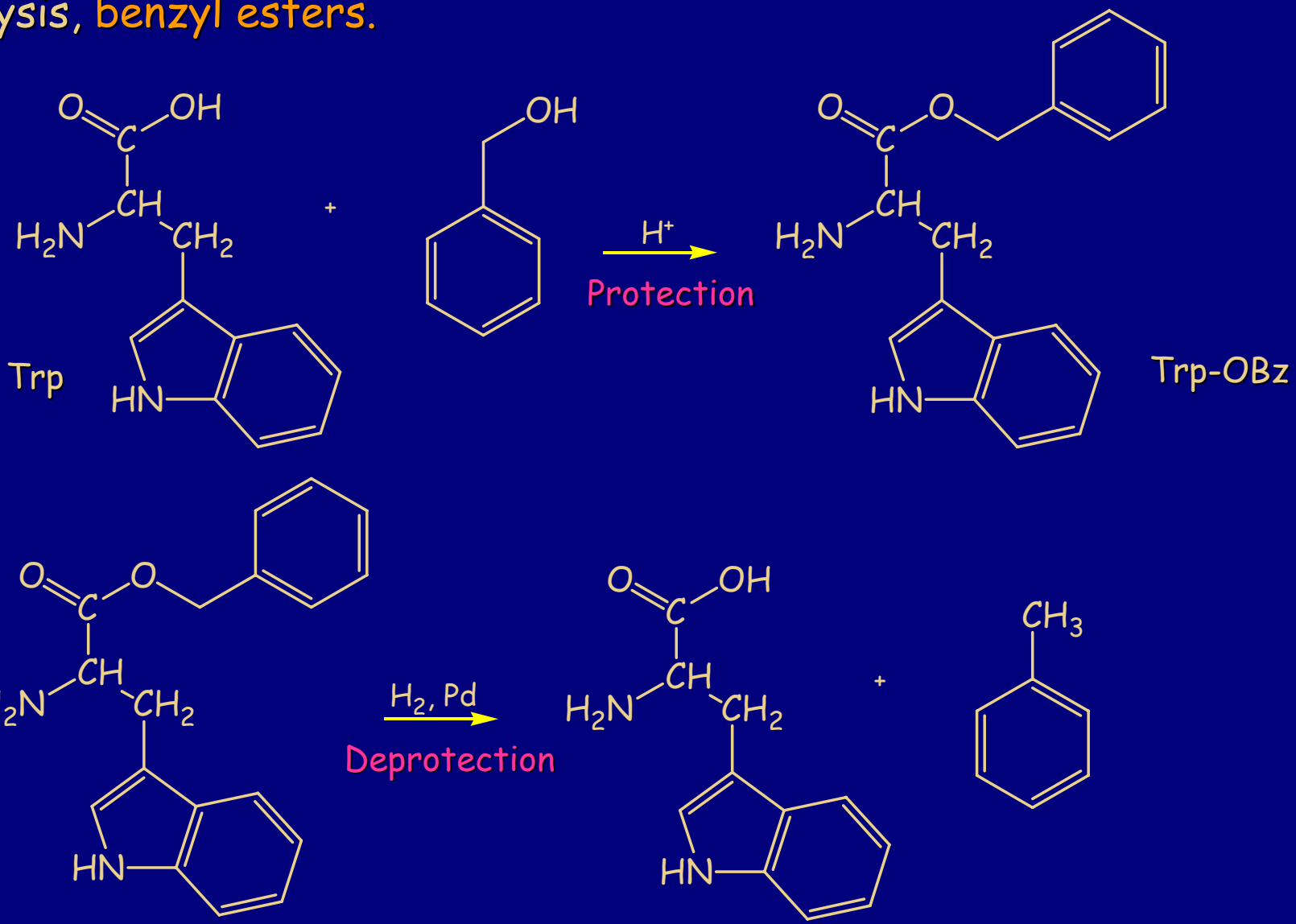
1,1-Dimethylethoxy-carbonylamino acid (*tert*-Butoxycarbonylamino acid, Boc-amino acid)

Deprotection of Boc-Amino Acids (via *tert*-Bu cation)

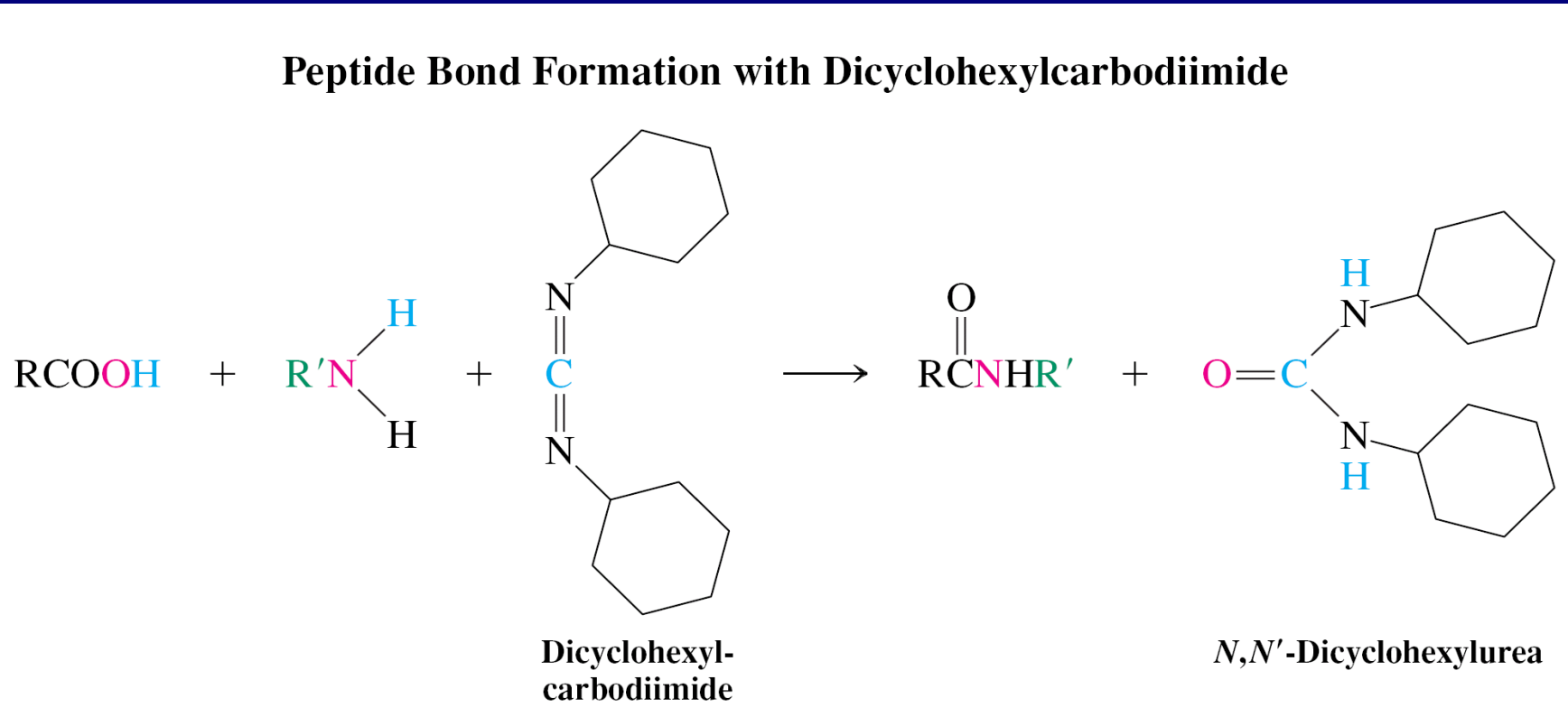


c. Carboxy Protection: Esterification

Simple methyl or ethyl esters. Alternatively, to avoid base (or acid) hydrolysis, benzyl esters.

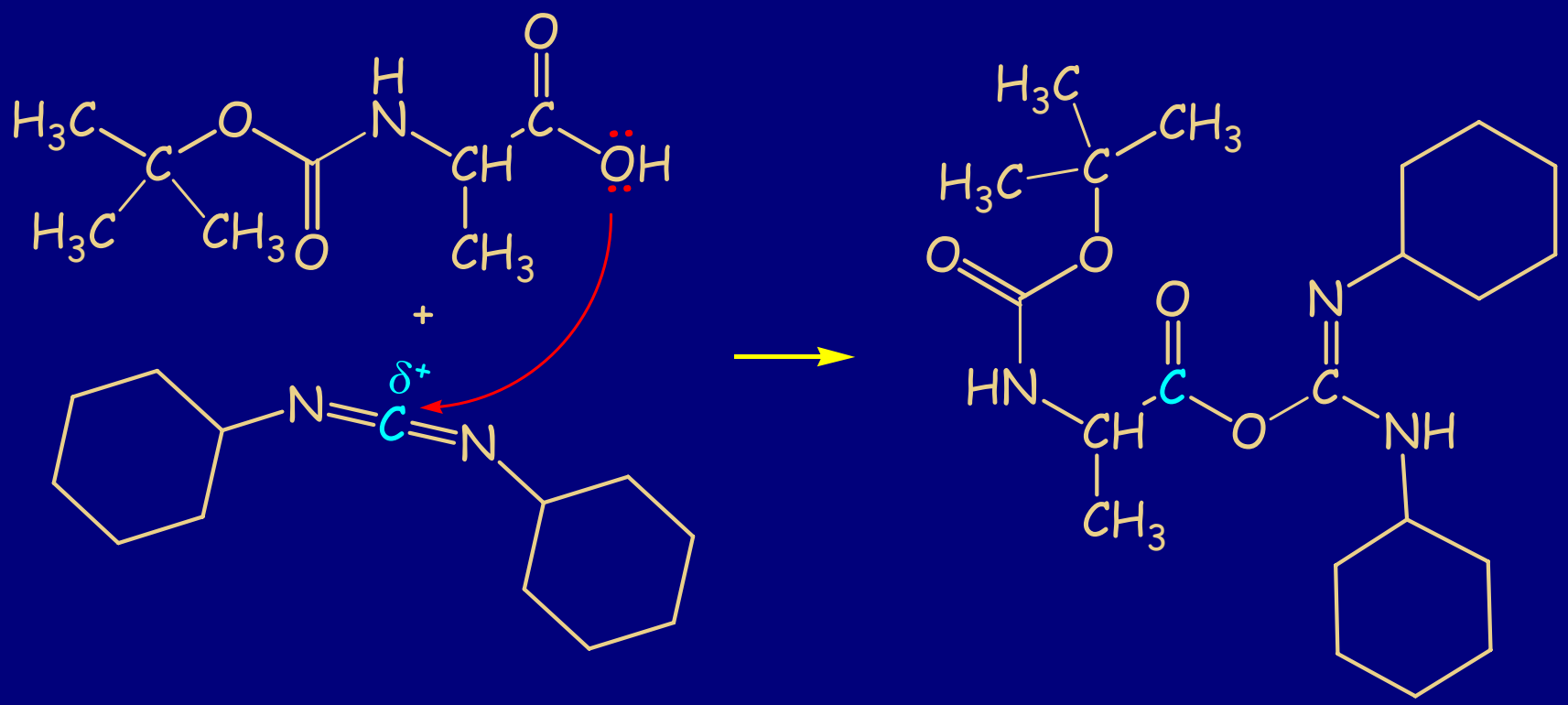


Formation of the amide bond for peptides uses a mild coupling reagent: Dicycloheylcarbodiimide (DCC) as a dehydrating species



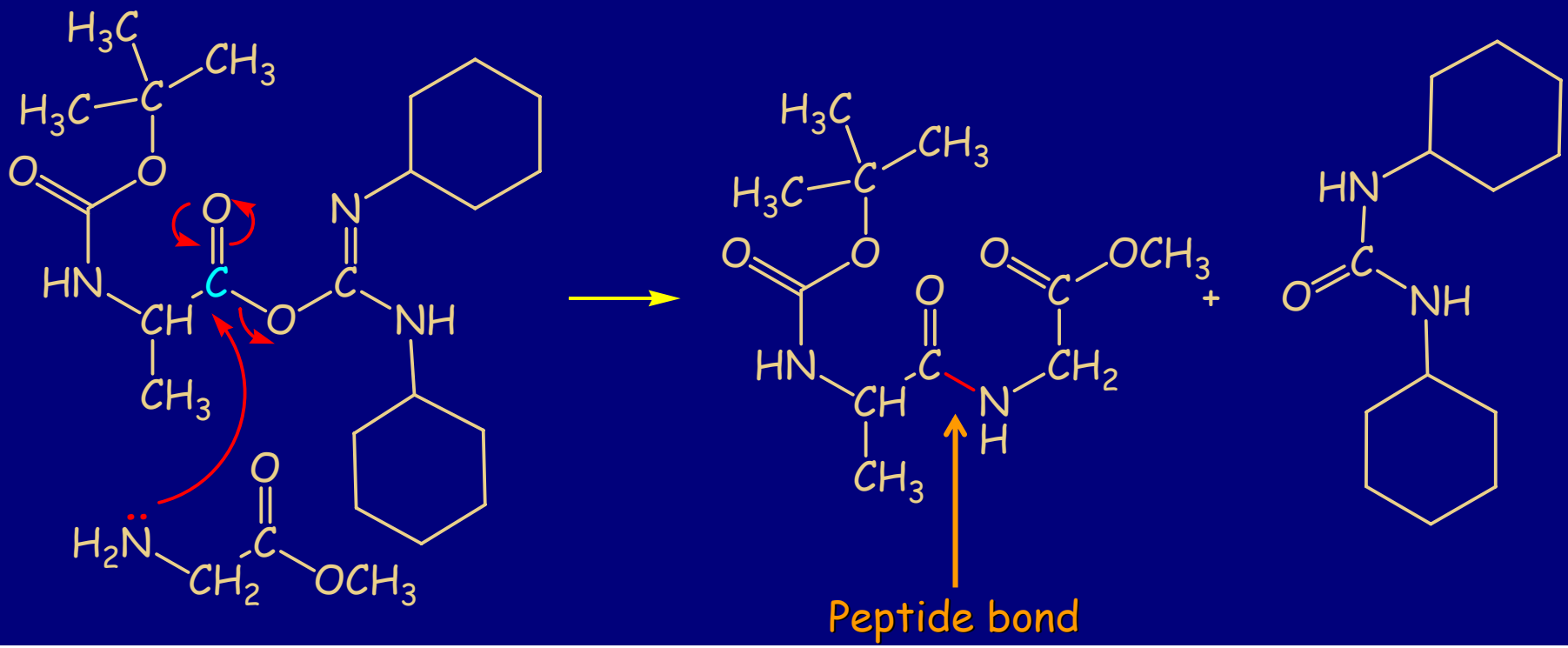
Mechanism:

1. Activation of carboxy group (recall activations to alkanoyl halides or anhydrides) of *N*-protected amino acid



Looks like an anhydride

2. Coupling with amino end of carboxy end protected amino acid to give dipeptide



Preparation of Gly-Ala

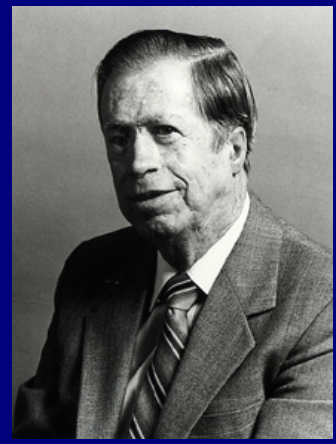
$(CH_3)_3COCNHCH_2COOH$ + $H_2NCHCOCH_2C_6H_5$ (with CH_3) $\xrightarrow{DCC}$ $(CH_3)_3COCNHCH_2CNHCHCOCH_2C_6H_5$ (with CH_3) $\xrightarrow{1.\ H^+,\ H_2O;\ 2.\ H_2,\ Pd\text{–}C}$

Boc-Gly | **Ala-$OCH_2C_6H_5$** | **Boc-Gly-Ala-$OCH_2C_6H_5$**

$H_3\overset{+}{N}CH_2CNHCHCOO^-$ (with CH_3) + $C_6H_5CH_3$ + CO_2 + $CH_2{=}C(CH_3)_2$

Gly-Ala

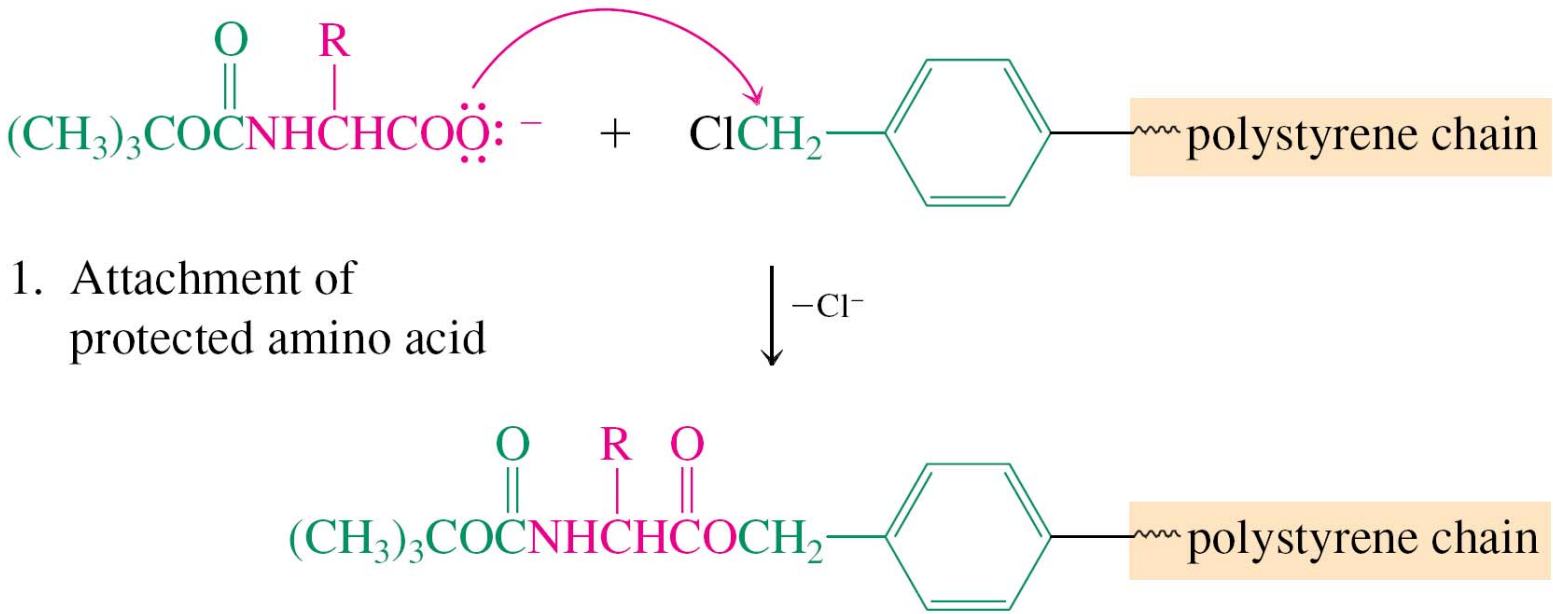
Automation: Merrifield solid-phase peptide synthesis

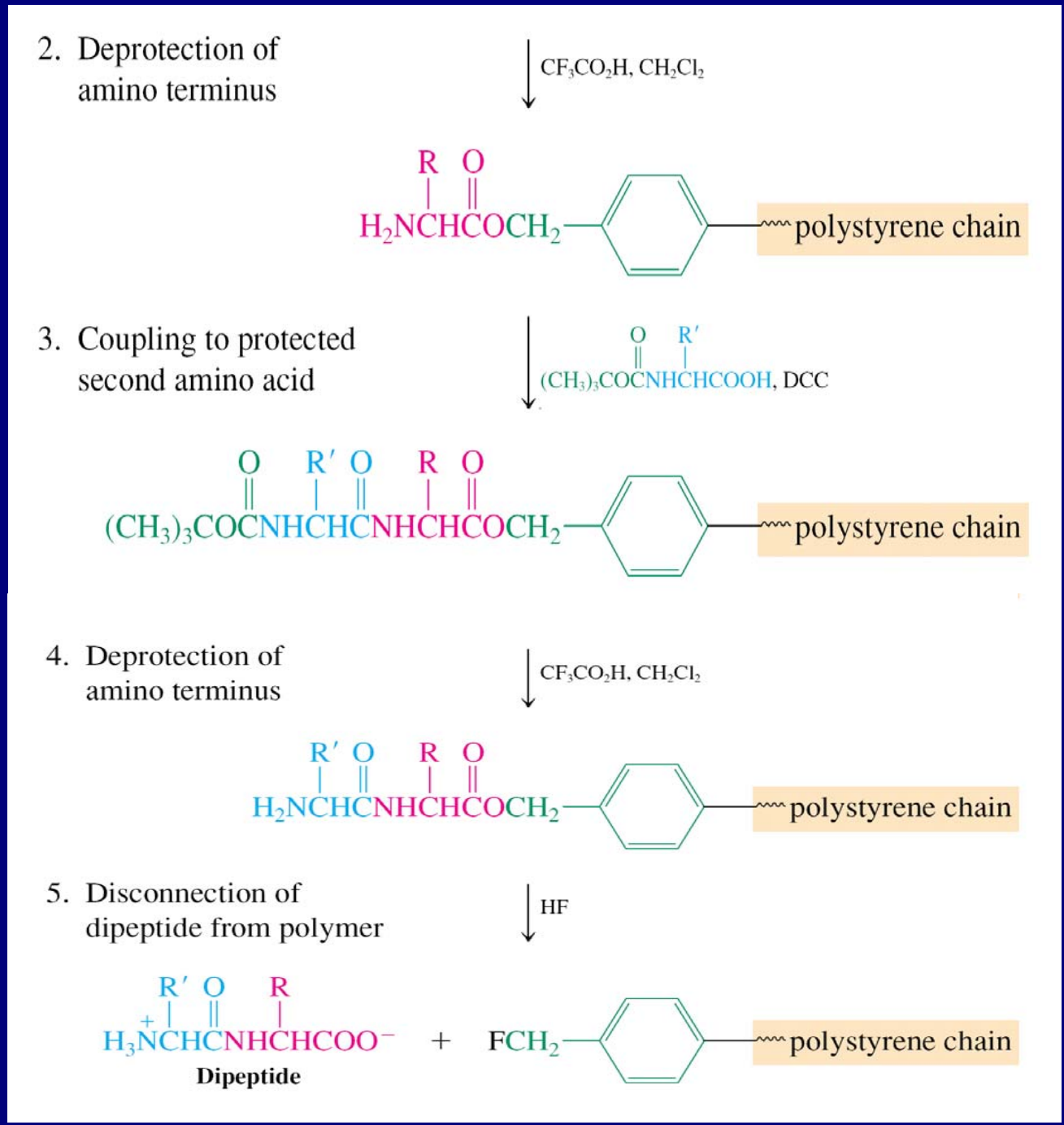


Robert B. Merrifield
b. 1921

Advantages: Robotic "custom made" assembly of any polypeptide; products on the polymer are isolated and purified by simple filtration and washing.

Solid-Phase Synthesis of a Dipeptide



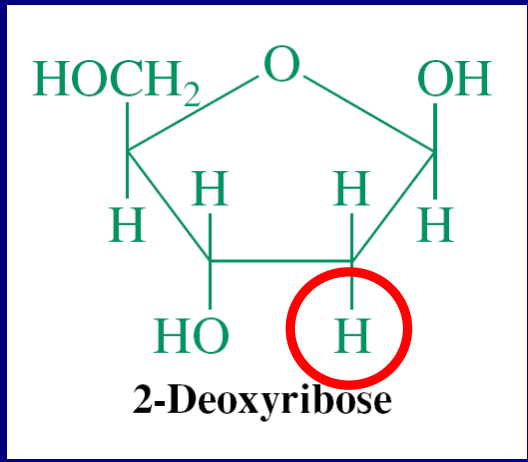


Early mile stone: 1966; Total synthesis of insulin with 51 amino acids; 5000 operations carried out in a few days. Modern extensions: Combinatorial chemistry and total synthesis of complex molecules on solid supports.

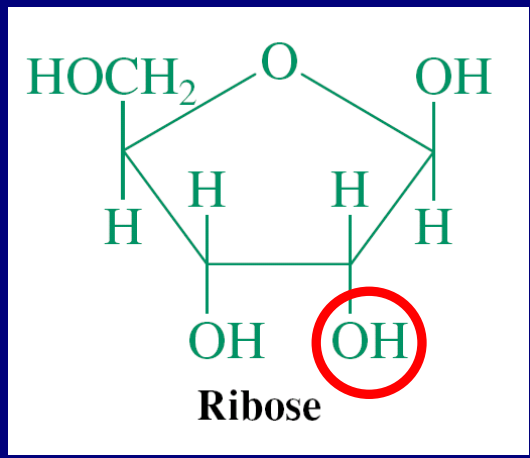
DNA and RNA: Natural Polymers Containing the Blueprint of Life

Life (in this context) is the synthesis of proteins, which run our (any) body. The information is stored in DNA: Deoxyribose nucleic acid

The information is "read" (expressed) with the help of RNA: Ribonucleic acid



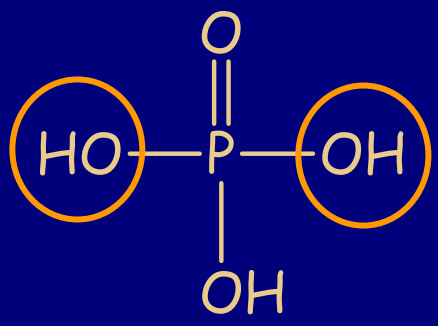
2-Deoxyribose



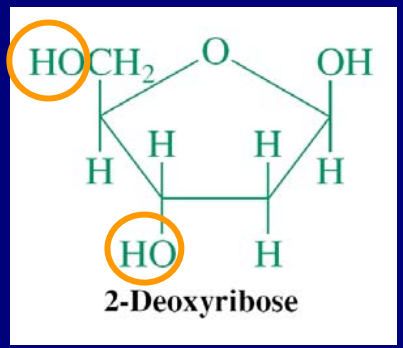
Ribose

Structure of Nucleic Acids

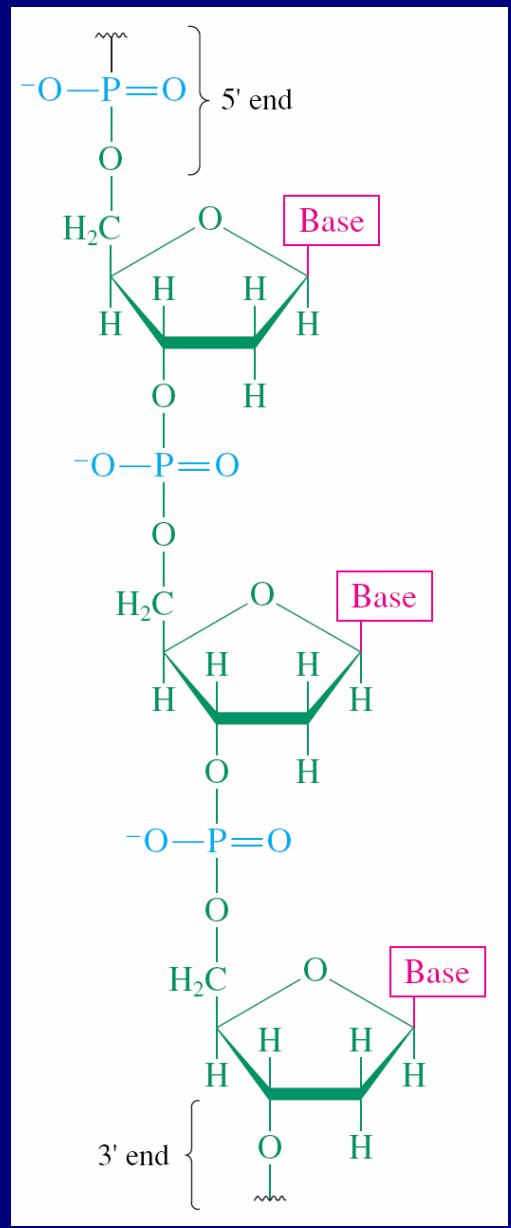
Example: DNA chain. Backbone is a polymer of the sugar linked by phosphate groups as a diester.



Phosphoric acid



2-Deoxyribose

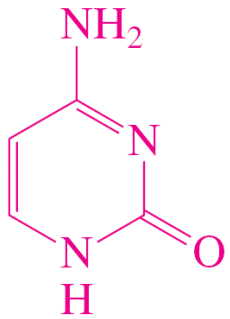


"Monomer" is called a nucleotide

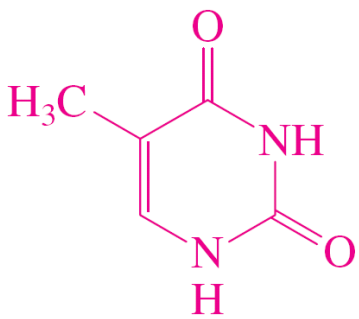
The information lies in the sequence of the bases attached to the anomeric carbon: There are four bases.

Bases: All aromatic heterocycles

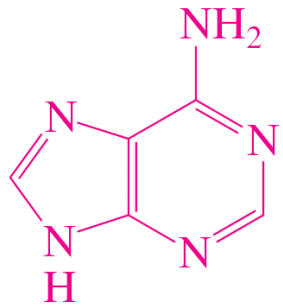
For DNA:



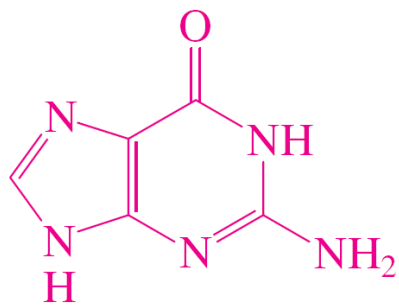
Cytosine (C)



Thymine (T)

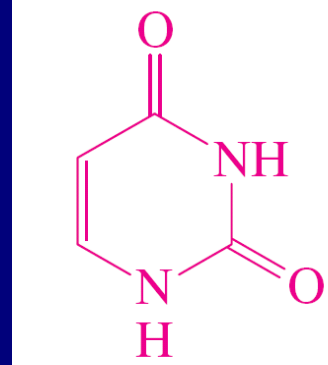


Adenine (A)



Guanine (G)

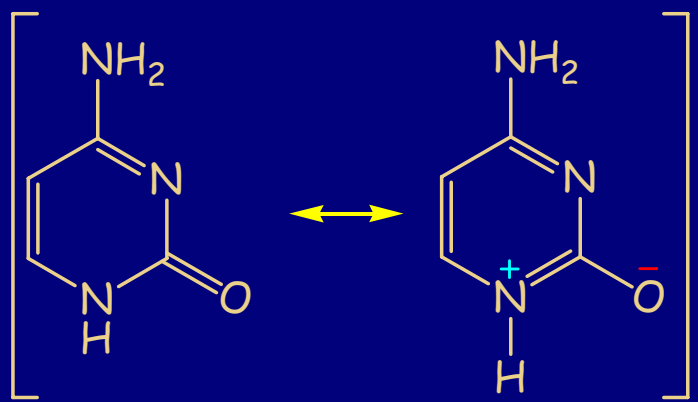
For RNA: C, A, G and



Uracil (U)

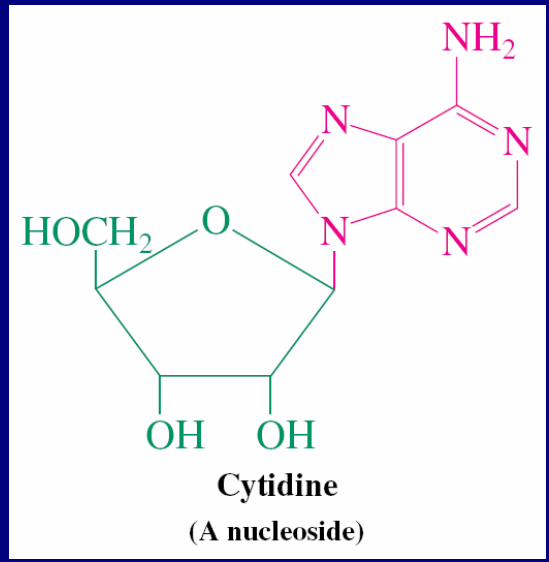
Instead of thymine

To visualize the aromaticity in the cyclic amides, formulate the dipolar resonance form, e.g. cytosine:



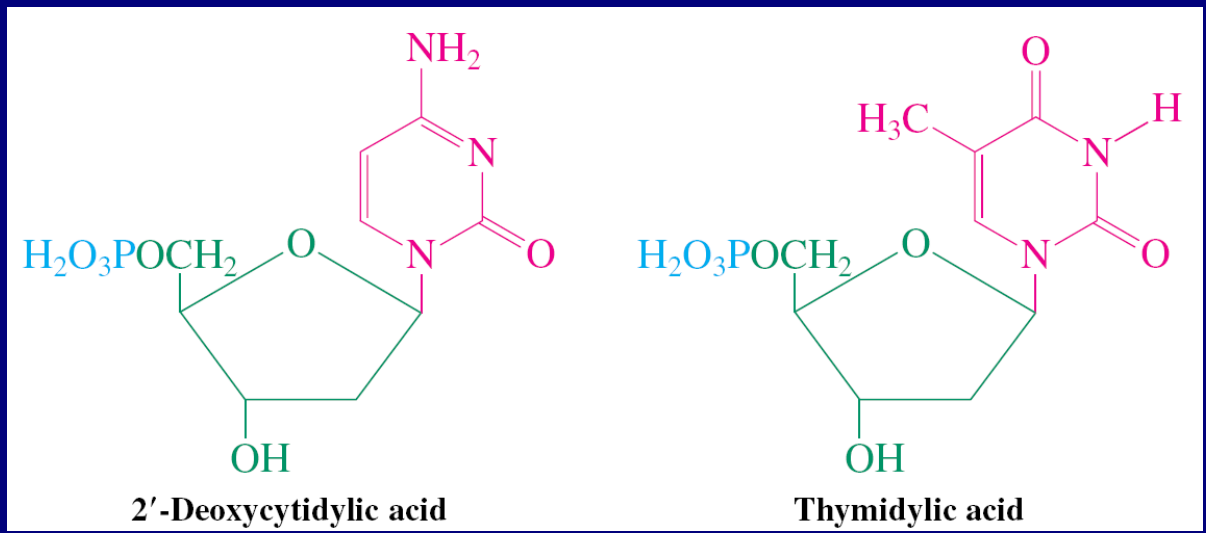
Naming the pieces:

Sugar-base compound:
A nucleoside



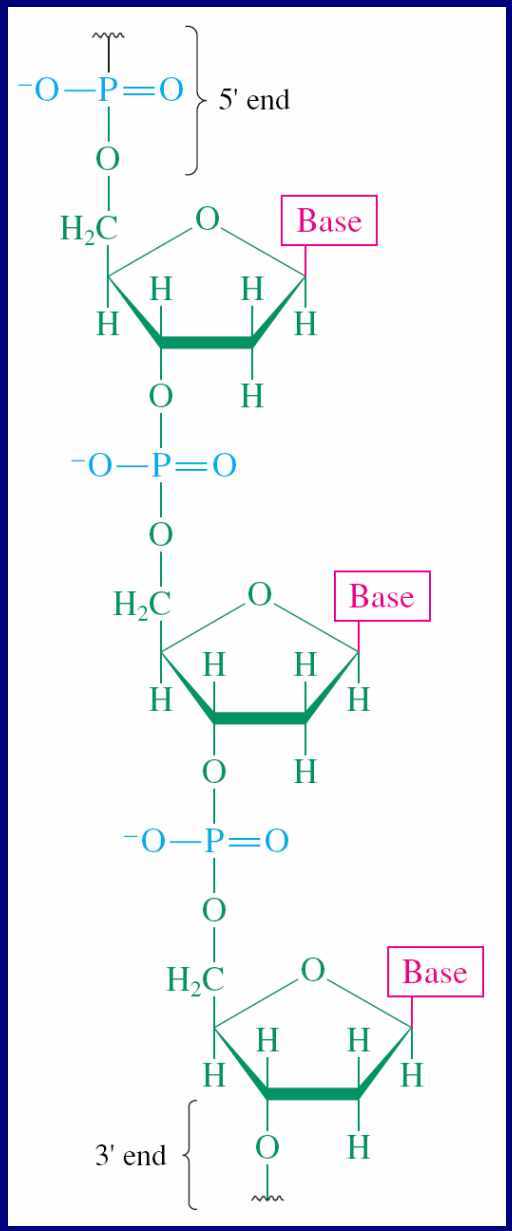
Cytidine
(A nucleoside)

Sugar-base-phosphate monomer piece of polymer:
A nucleotide



2′-Deoxycytidylic acid

Thymidylic acid

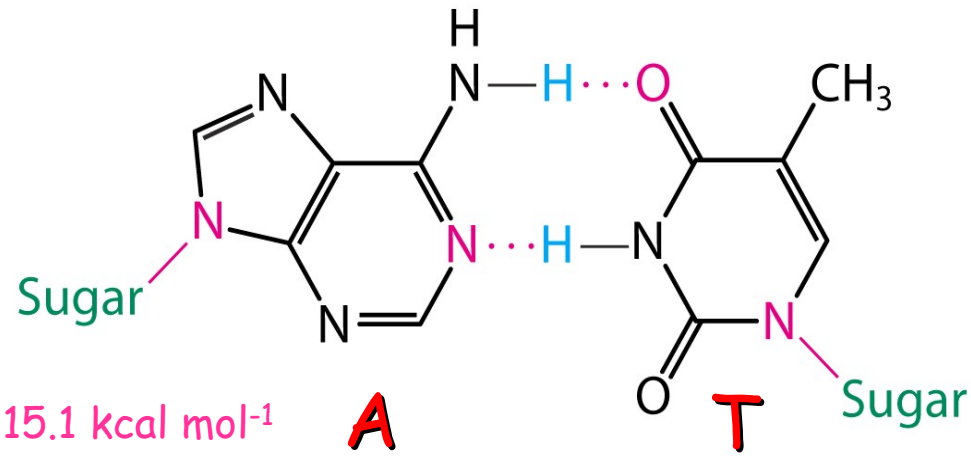


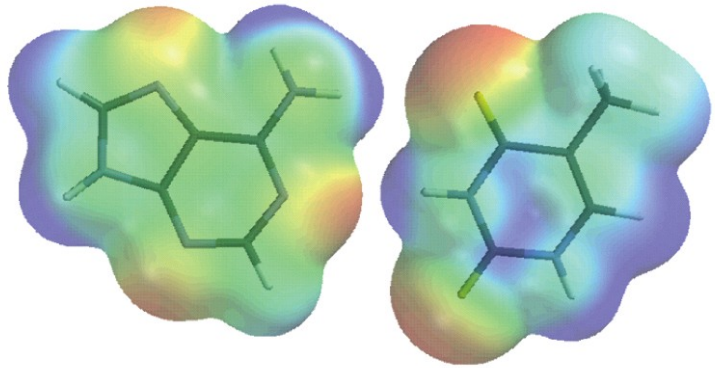
Primary structure

DNA forms extraordinarily long chains (up to several centimeters) with molecular weights of as high as 150 billion. Like proteins, they adopt secondary and tertiary structures: Double helix!

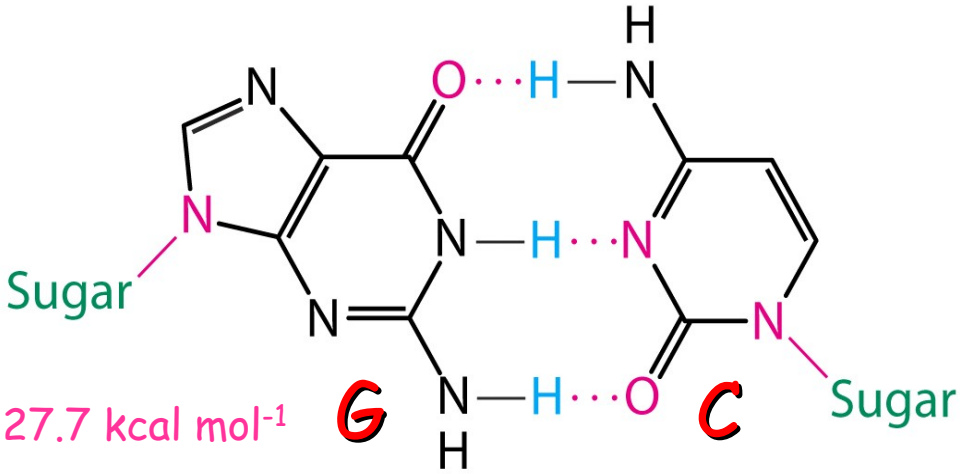
Human chromosomes

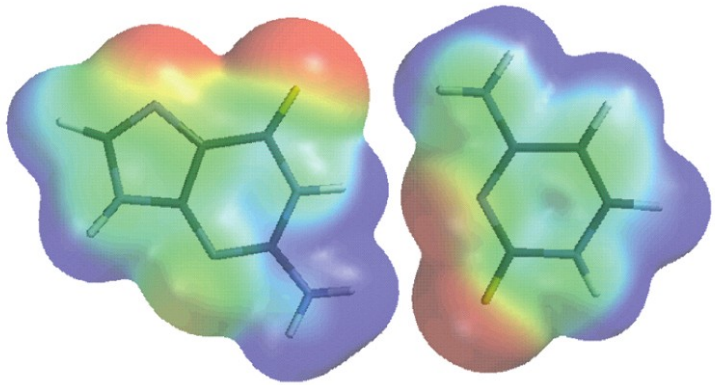
Watson-Crick: Hydrogen bonding through complementary pairs: In DNA A-T, G-C (1:1 ratio)





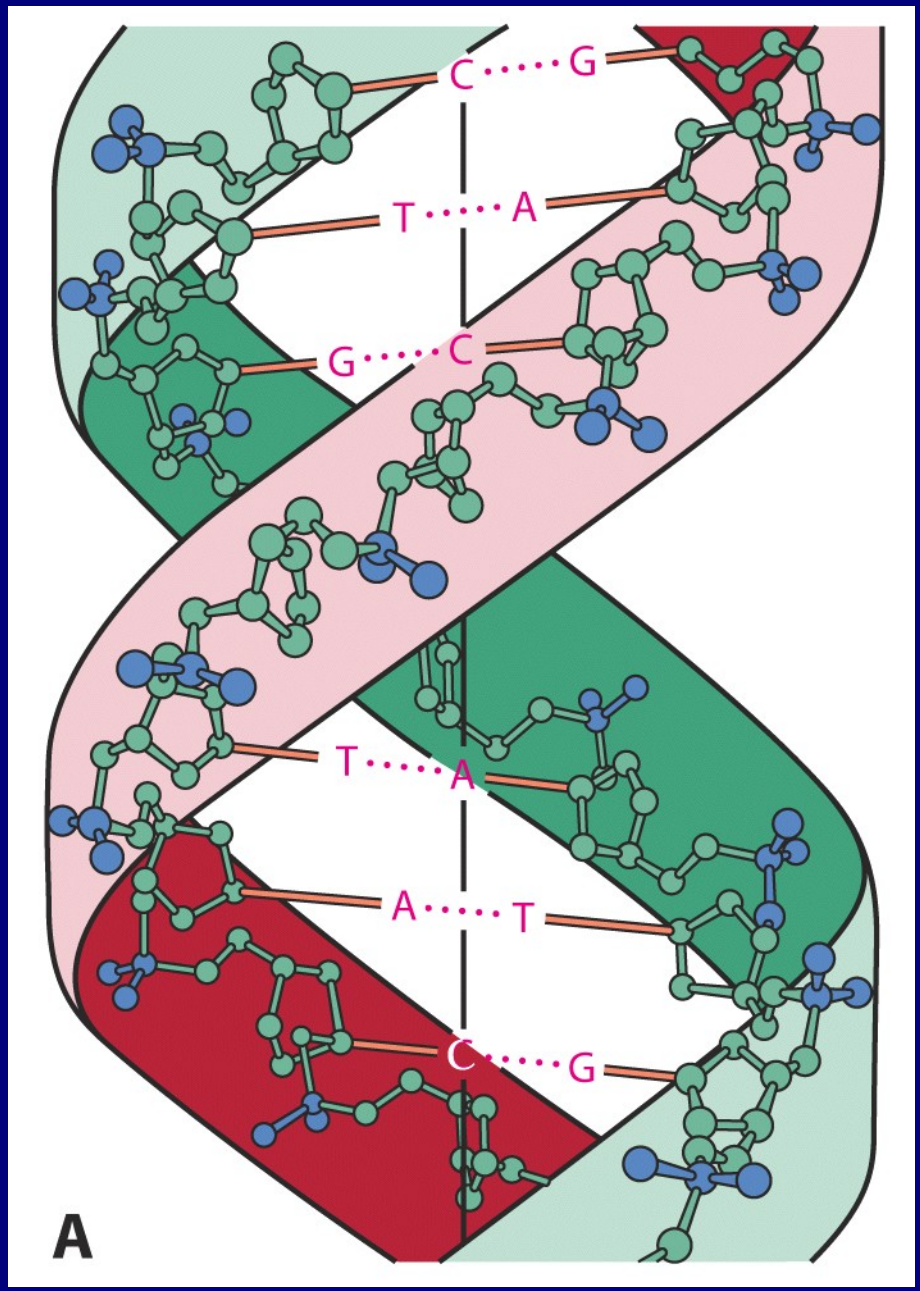
Adenine–thymine

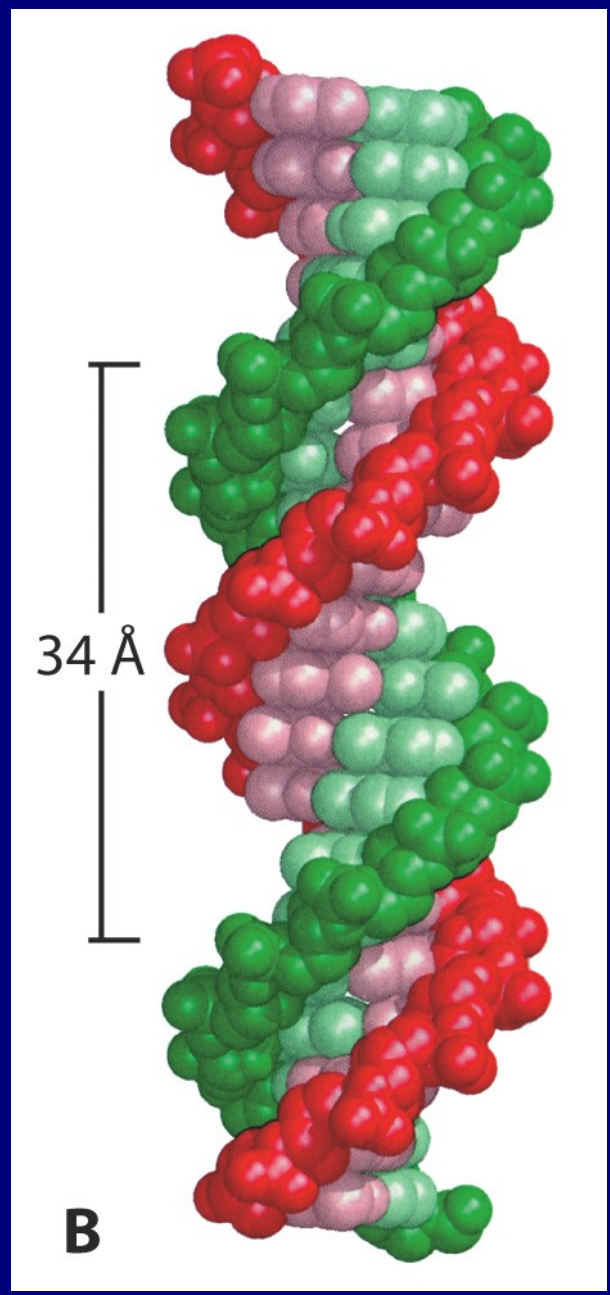




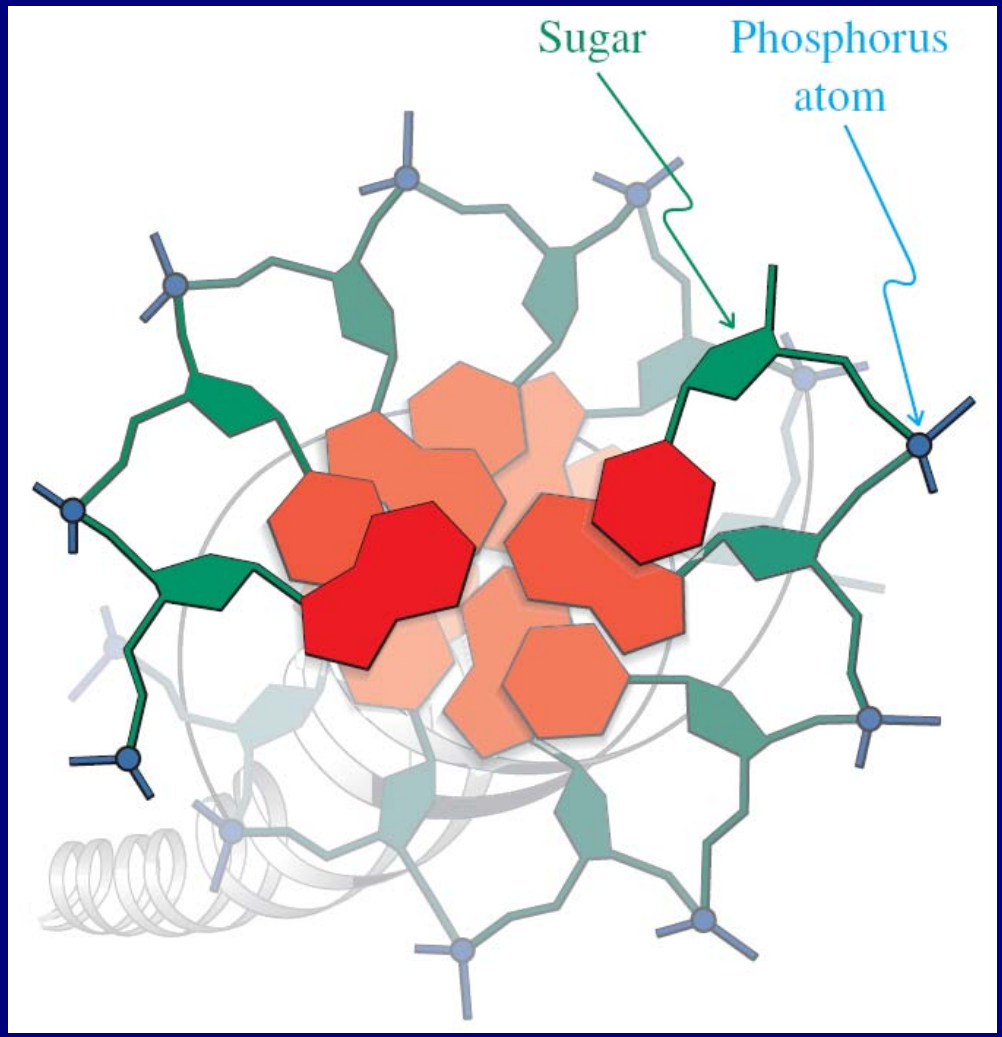
Guanine–cytosine

Gives rise to double helix.....


C ····· G
T ····· A
G ····· C
T ····· A
A ····· T
C ····· G
A


34 Å
B

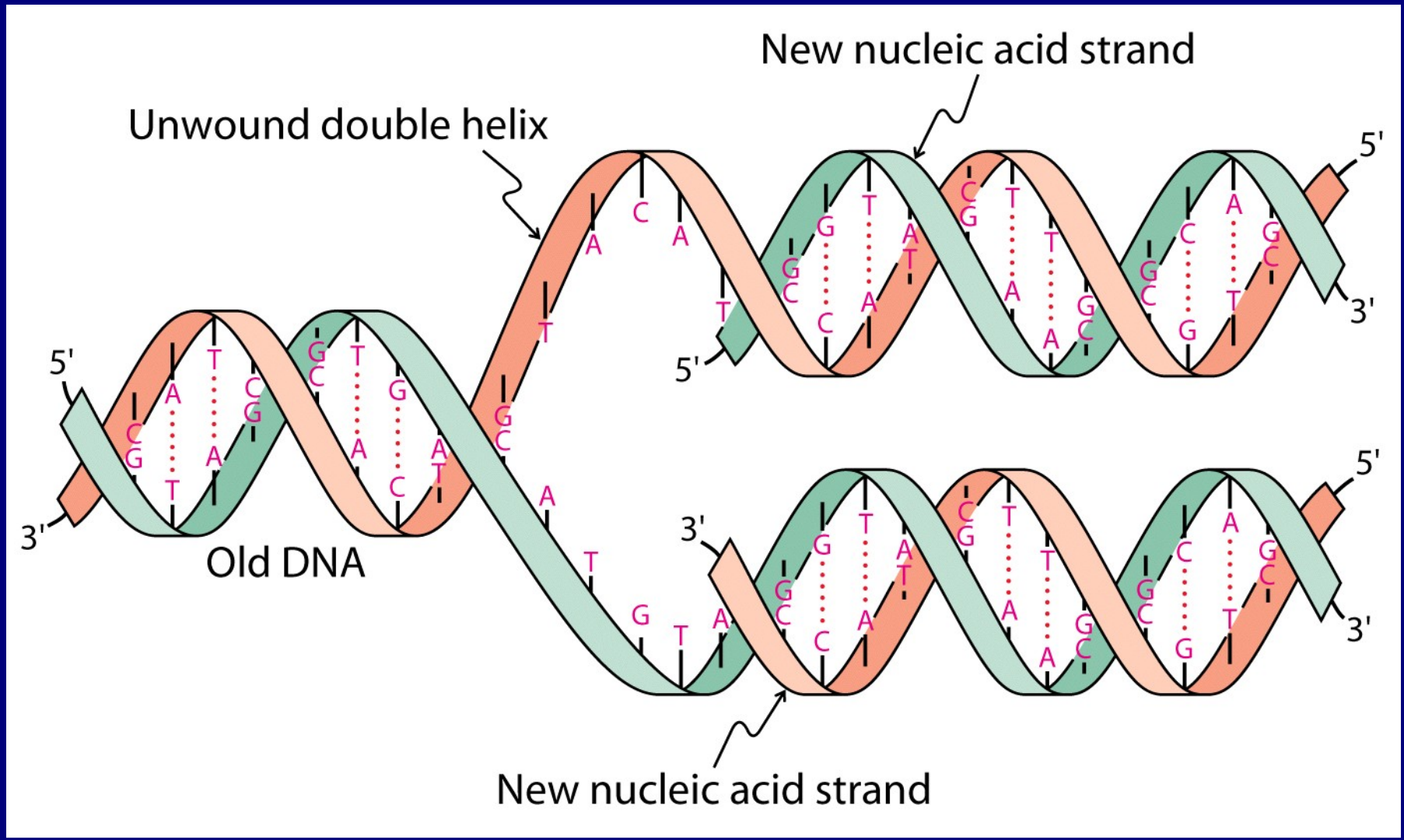
View down the helical axis



Complementarity:

A—G—C—T—A—C—G—A—T—C
T—C—G—A—T—G—C—T—A—G

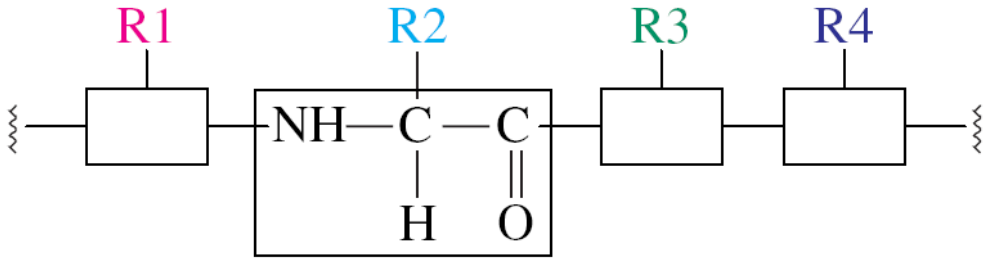
Replication of DNA



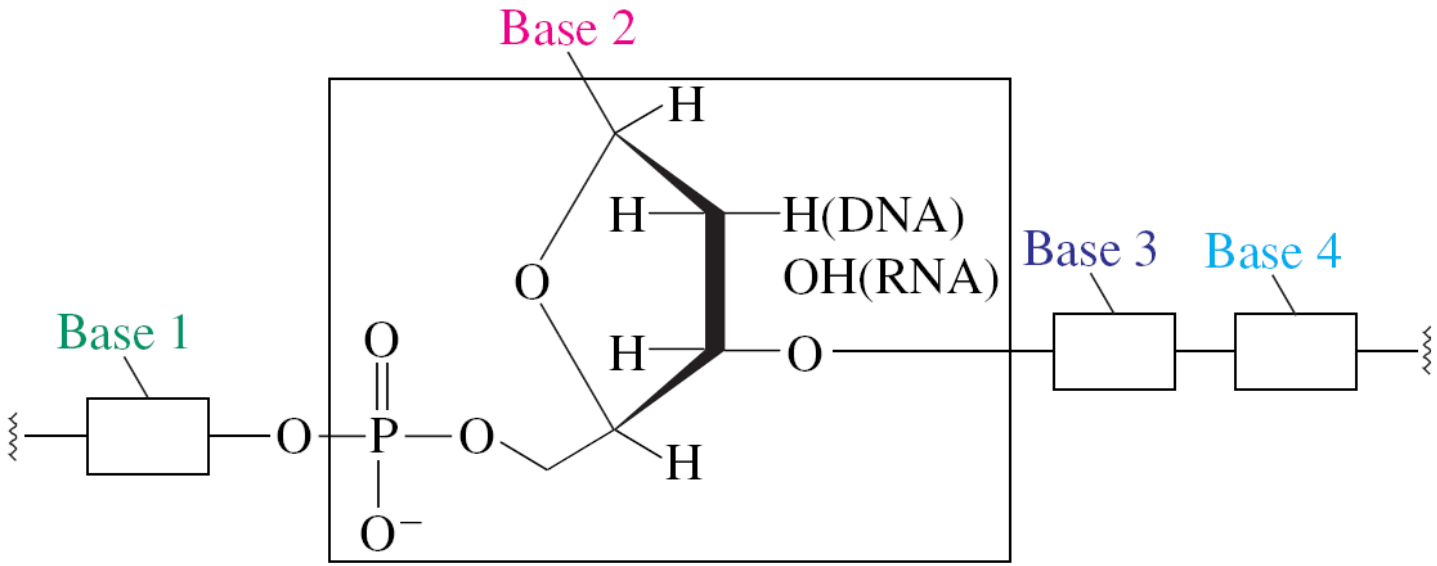
In humans 2.9 billion base pairs; error rate 1 in 10 billion!

Information storage: Peptides versus nucleic acids

A polypeptide R = 20 letters: The 20 natural amino acids



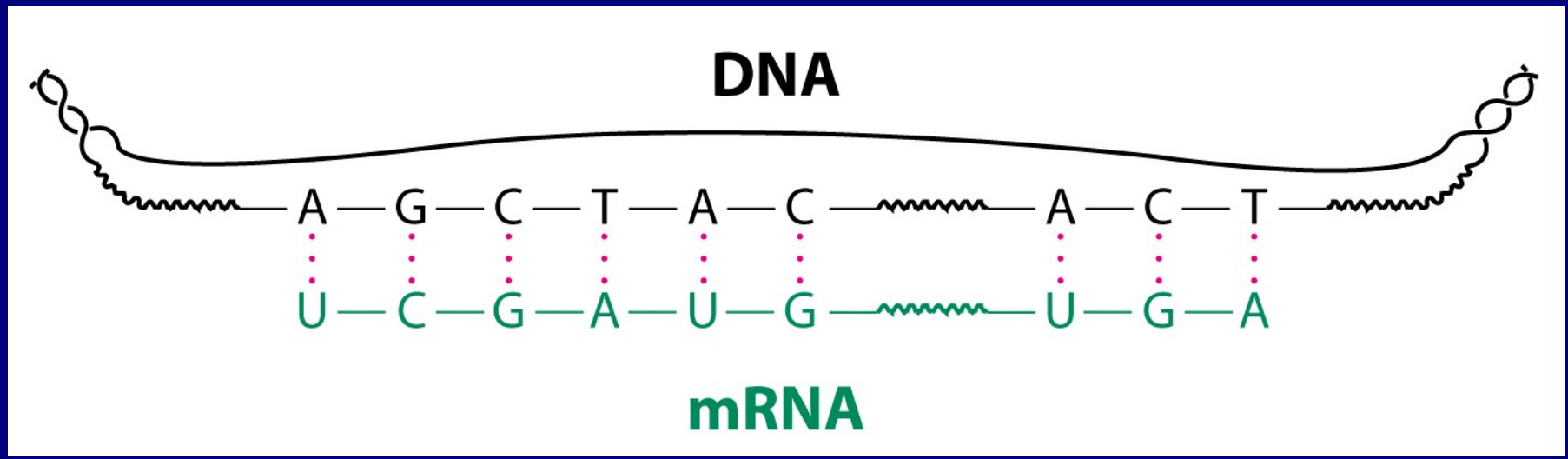
A nucleic acid Four letters: The four bases



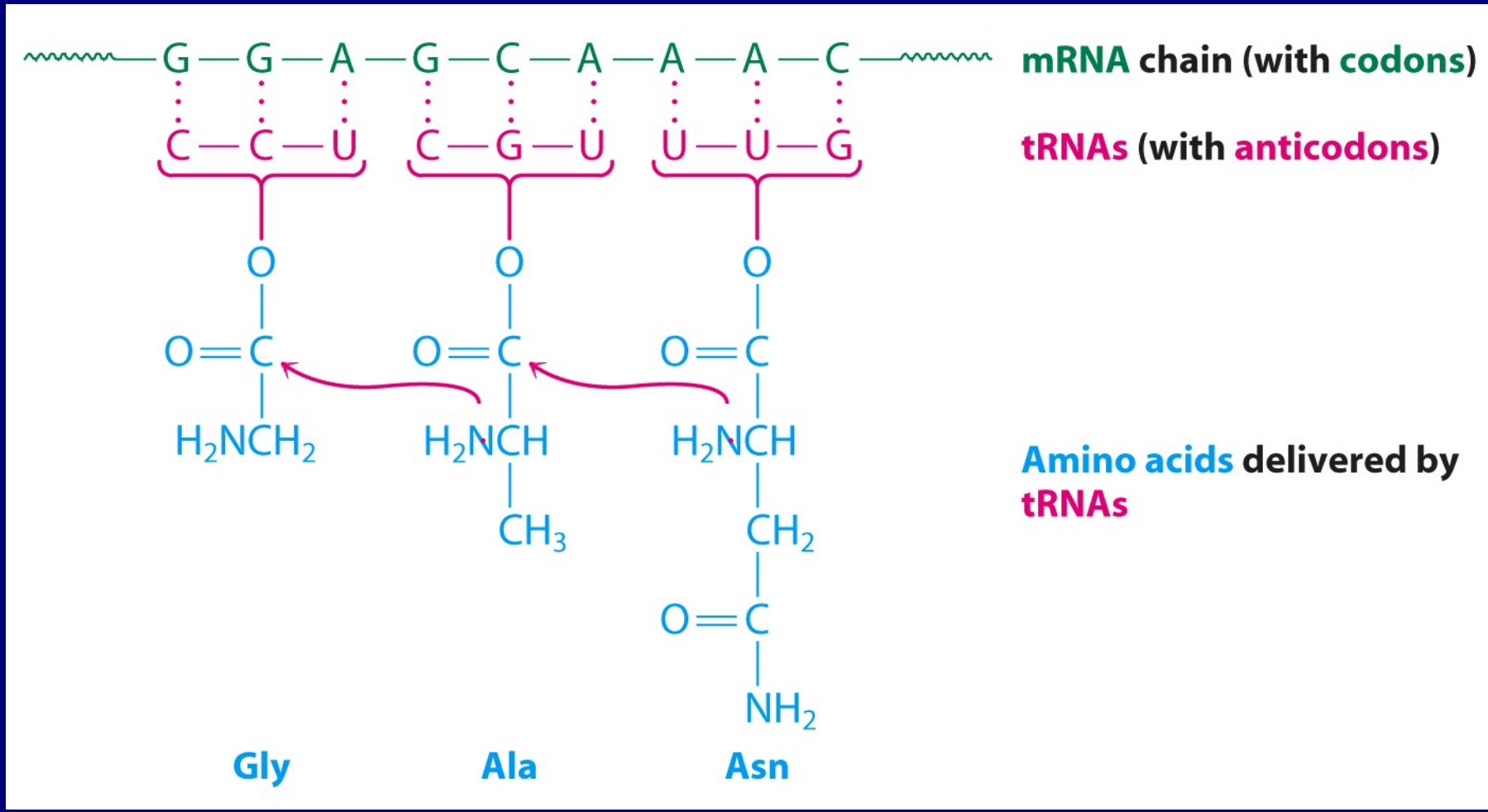
Base sequence contains information for protein synthesis

DNA → mRNA → polypeptide

Messenger RNA



mRNA copies a piece of DNA to be used to construct a particular peptide; transfer RNA (tRNA) delivers the amino acids, and a catalyst (ribosome) puts the peptide together, following the blueprint: Three base sequences (codons) that translate into a specific amino acid.



Three base code (codon): # of combinations

$4^3 = 64$ → more than enough for 20 amino acids

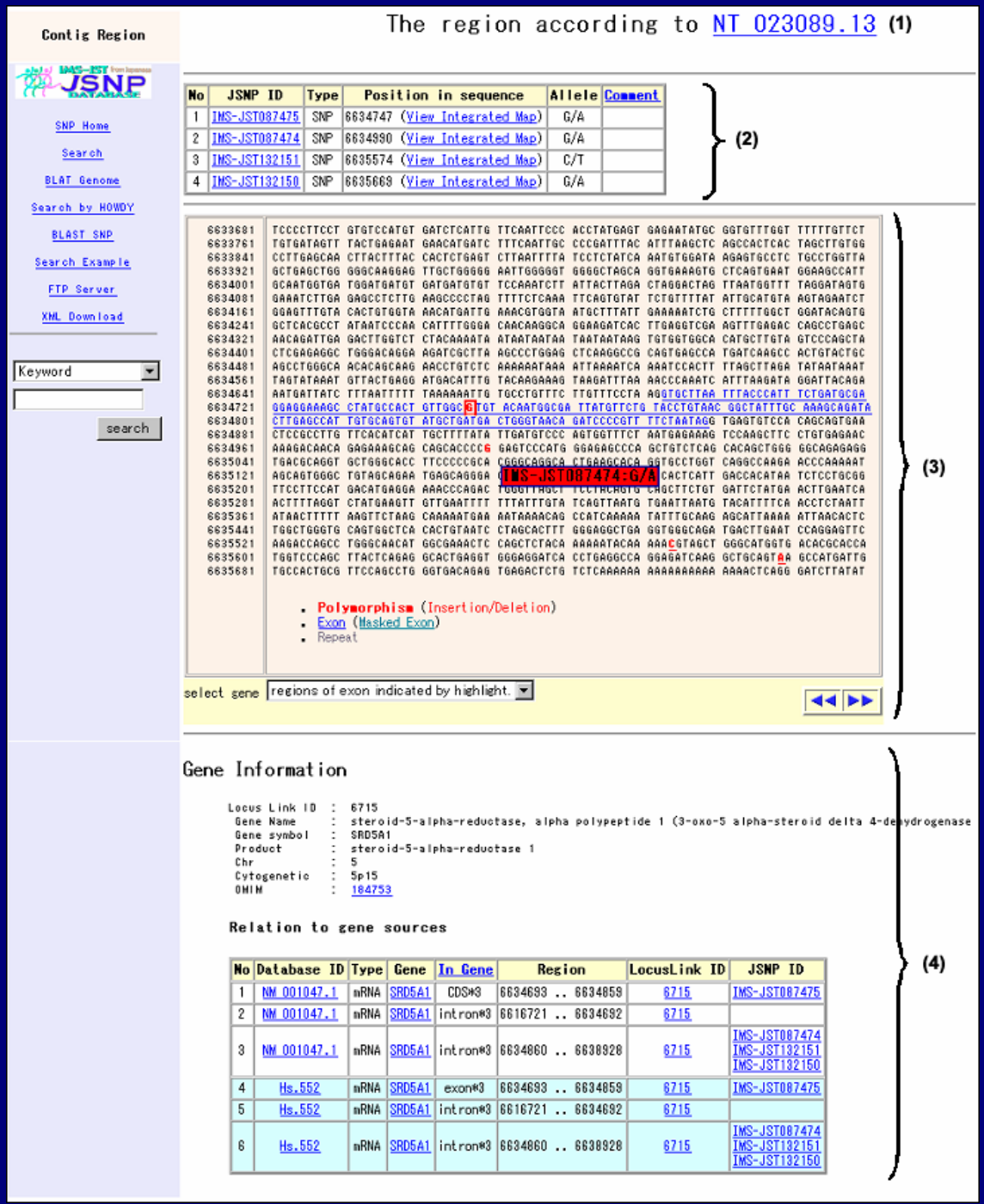
TABLE 26-3 Three-Base Code for the Common Amino Acids Used in Protein Synthesis

Amino acid	Base sequence	Amino acid	Base sequence	Amino acid	Base sequence
Ala (A)	GCA	His (H)	CAC	Ser (S)	AGC
	GCC		CAU		AGU
	GCG				UCA
	GCU	Ile (I)	AUA		UCG
			AUC		UCC
Arg (R)	AGA		AUU		UCU
	AGG				
	CGA	Leu (L)	CUA	Thr (T)	ACA
	CGC		CUC		ACC
	CGG		CUG		ACG
	CGU		CUU		ACU
			UUA		
Asn (N)	AAC		UUG	Trp (W)	UGG
	AAU				
		Lys (K)	AAA	Tyr (Y)	UAC
Asp (D)	GAC		AAG		UAU
	GAU				
		Met (M)	AUG	Val (V)	GUA
Cys (C)	UGC				GUG
	UGU	Phe (F)	UUU		GUC
			UUC		GUU
Gln (Q)	CAA				
	CAG	Pro (P)	CCA	Chain initiation	AUG
			CCC		
Glu (E)	GAA		CCG	Chain termination	UGA
	GAG		CCU		UAA
					UAG
Gly (G)	GGA				
	GGC				
	GGG				
	GGU				

Genetechnology: A Revolution

The Human Genome

Deciphering the sequence of 2.9 billion base pairs



The Future ?

http://funnyjunk.com/pages/world.htm/