

Lec 3: More Nucleic Acids, Proteins.

Homework

Reading, ECB, Chpt 1 and Chpt 2
(a lot, but...)

Hw #2, due Jan. 30. Posted on web this afternoon.

Change to schedule: DNA twisting and bending flexibility, originally scheduled for March 24th (Lec. 18), will be either next time or time-after. You learn about some single molecule experiments and DNA flexibility.

Quiz today on introductory reading

A lot to cover!

Quiz #1 (covering Stryer)

1. a) List the 3 kingdoms of life.
b) You are a member of which kingdom?

a) bacteria, archea, eukarya/eukaryotes

b) eukaryotes

2. The folding of proteins is driven primarily by changes in entropy.

What molecules in the system exhibit the greatest change in entropy?

Water

3. What are the two chemical differences possible between DNA and RNA?

Additional OH group off of sugar, 2' OH

Uracil instead of Thymine base

4. A disease caused by mutation in the CFTR gene locus is called _____. The mutation leads to a decrease in the salt secretion by a transport protein.

Cystic fibrosis

Your questions from Lec. 2

1. Graduate students who need 4 credits → easy.

Sign up via web for Physics 597 (independent study) with me. **My # is 25016.** No extra effort on your part. Grade in 597 = grade in Physics 498Bio.

Scientific

2. Difference between major & minor groove

-- see today's viewgraphs

3. Why is one end of DNA called 5'? The other 3'?

-- the nomenclature of sugar #'s. (today)

4. More about DNA folding – why a meter long can compact into a few microns

-- have a special section on DNA bending and twisting with magnetic traps next time or time-after.

5. What if mis-match: how fix it?

-- Recognize by change in radius.

-- Mechanism: polymerases, helicases; later.

6. Ribosome– how does it know which 3 RNA letters to equate with which proteins.

-- uses special RNA: tRNA (some today)

**DNA is a linear polymer of nucleotides.
Backbone is held together by covalent
sugar-phosphate bonds.**

DNA Backbone is negatively charge.
(One charge per nucleotide)

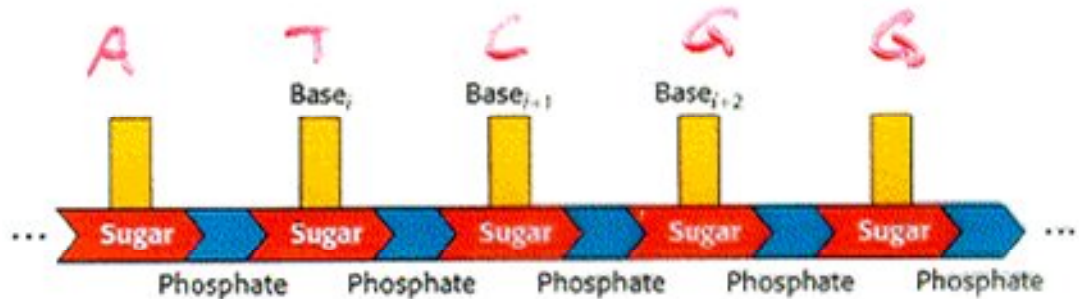
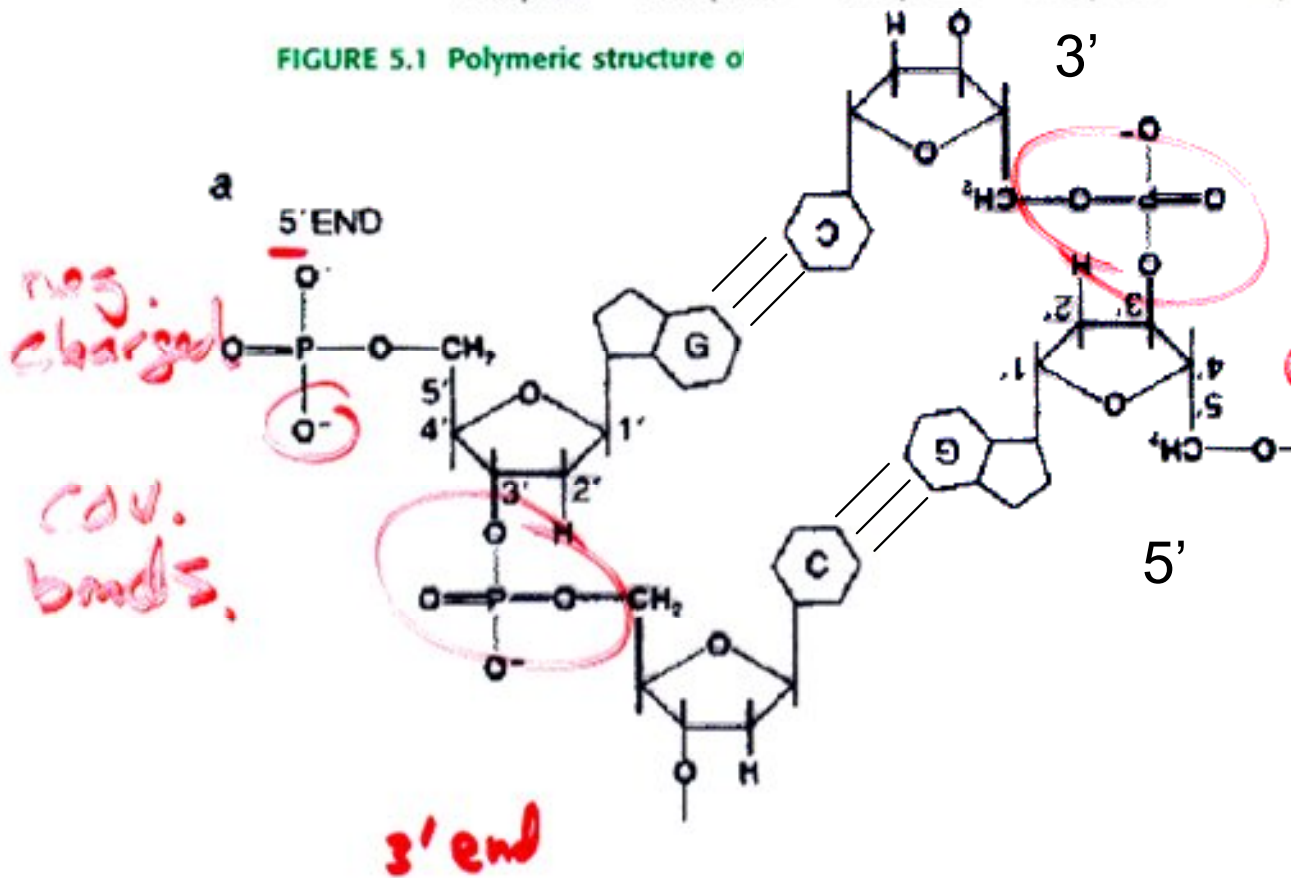


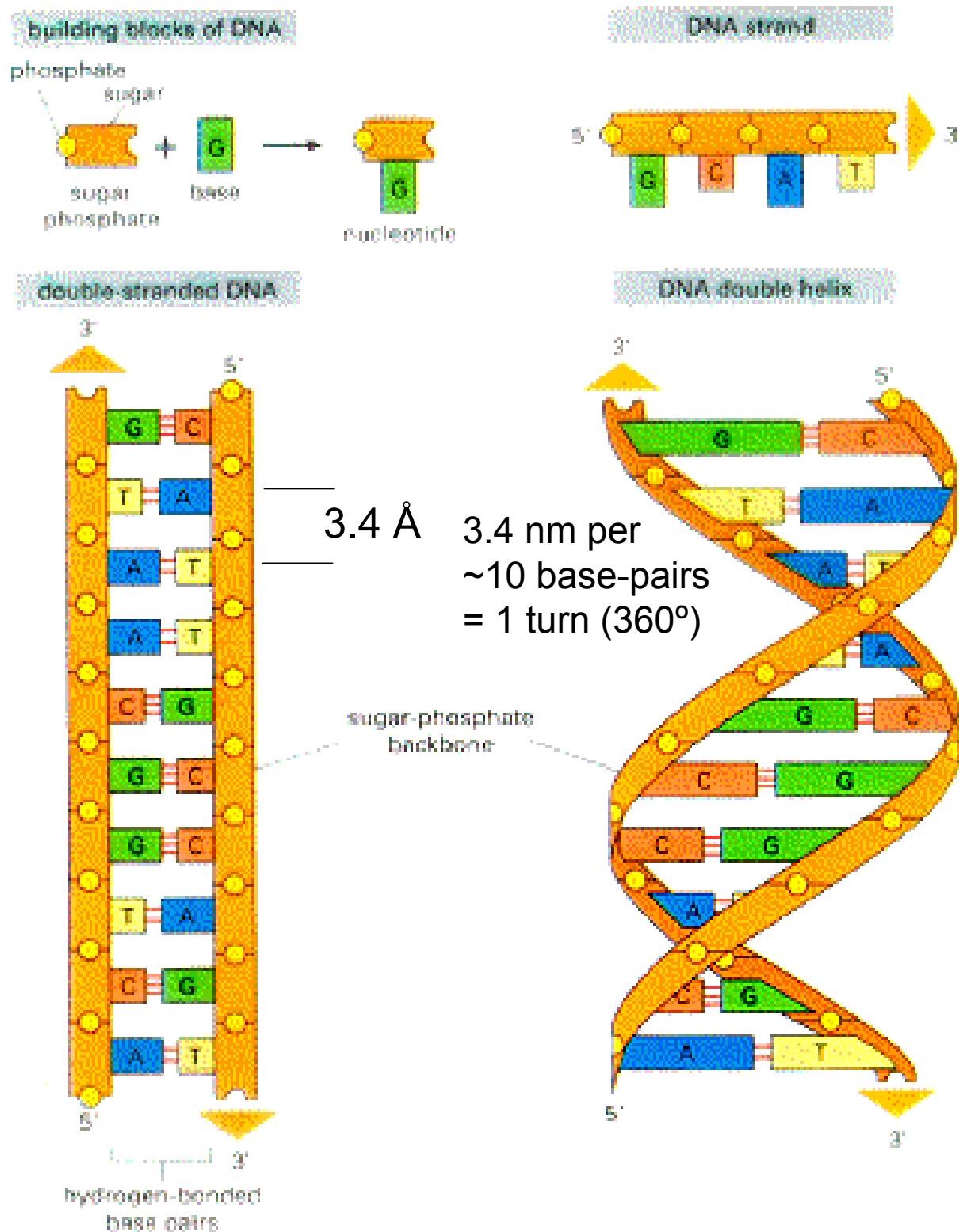
FIGURE 5.1 Polymeric structure o



**The storage of information
is in the linear arrangement of nucleotides.**

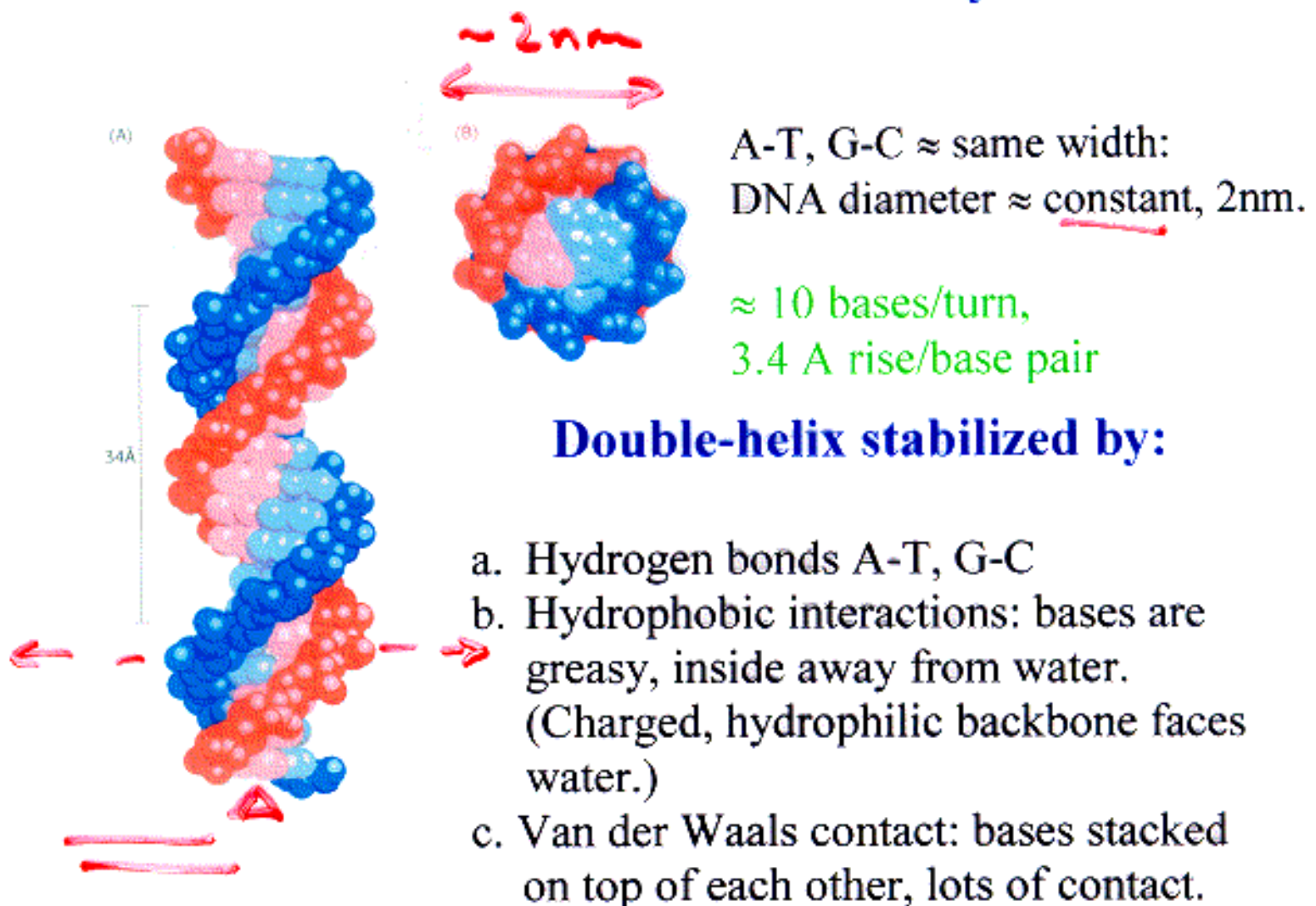
If add salt to solution, what is effect on melting Temp?
Melting temp = Temp. at which DNA strands come apart.

DNA is a double helix of anti-parallel strands



Must come apart for bases to be read.

DNA Dimensions & Stability



Destabilized by:

charge-repulsion of negative charged backbone.

Bottom line: A-T \approx 2kT, G-C \approx 4kT

Each one fairly weak, but many leads to stable structure. However, can be unzipped with input of energy (ATP-scale of Energy).

Major Groove— more exposed than Minor Groove

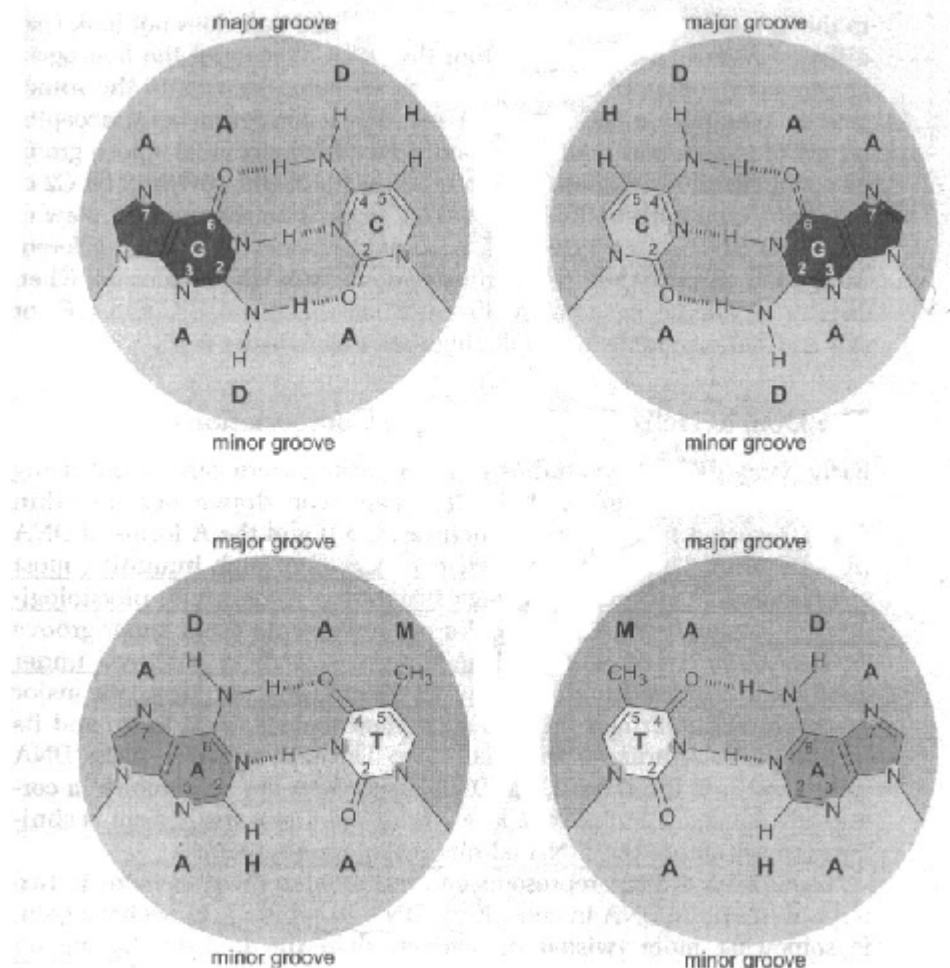
120° vs. 240° makes more accessible; base pairs identifiable

A: Hydrogen Bond Acceptor;

D: H-bond donor

M: Methyl Group;

H: non-polar hydrogen



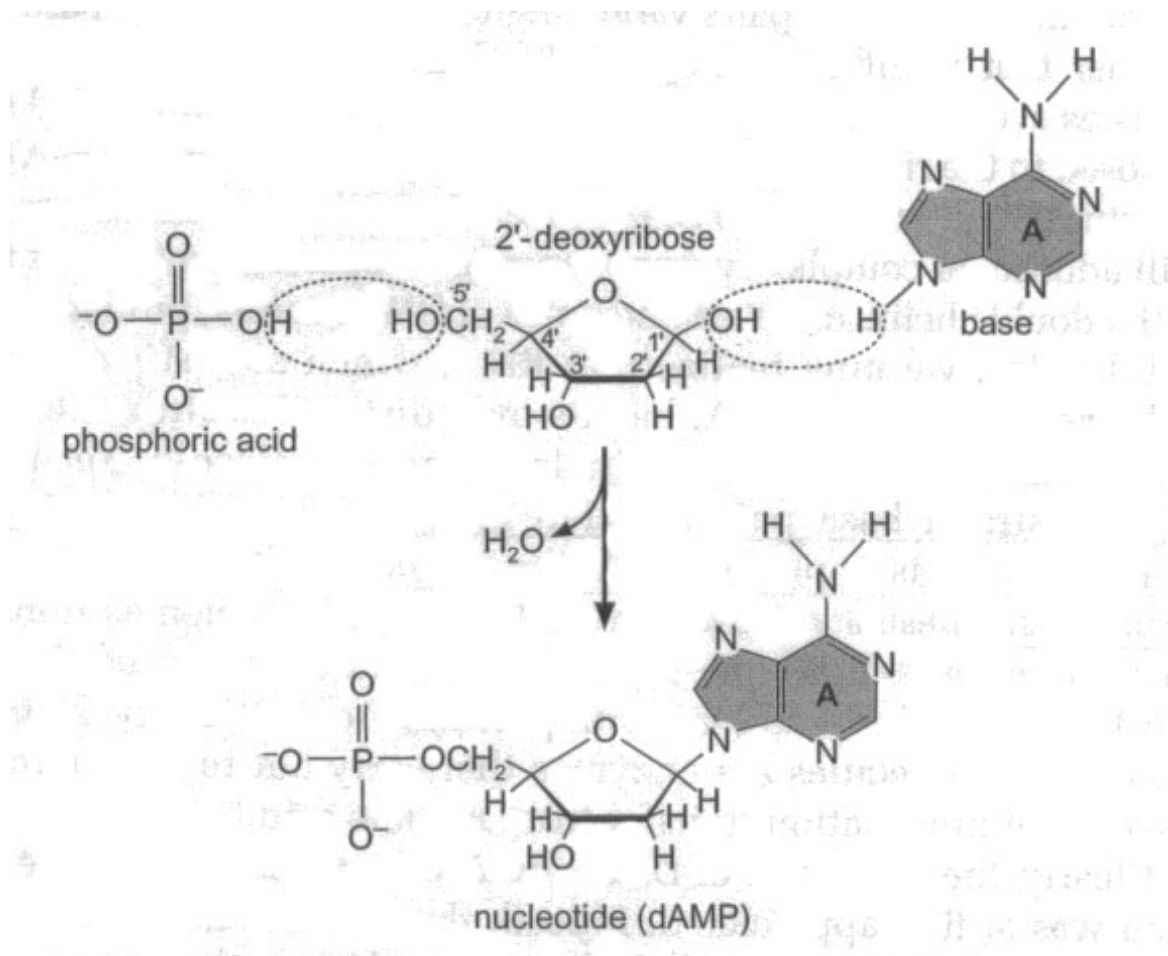
Minor groove: small size of less able to accommodate amino acids of protein interrogating DNA.

Major groove: each base pair forms unique pattern, can be recognized [G-C (AADH); C-G (HDAA)]

A-T, T-A (AHA=AHA) and G-C, C-G all look similar (ADA=ADA)

How to make nucleotide

(Example of **condensation reaction**,
like amino acids & peptide bonds)



Free H_2O : lots of entropy gained

Reaction wants to go.

What is Entropy?

$$\Delta S$$

Qualitative: Degree of disorder

Gas will expand from a bottle at finite temperature because there are more accessible states available to it, than staying put.

(Temperature gives molecules a little bit of energy to access states)

Quantitative:

$$\Delta S = k_B \ln W$$

where $W = \#$ accessible states.

Reaction that produces H_2O , which, in general, is free to diffuse everywhere, is highly favorable.

DNA → RNA → Proteins

Central Dogma of Biology

DNA: series of 4 nucleotides (bases): A,T,G,C

↓ Transcription [DNA & RNA similar]

RNA: series of 4 nucleotides (bases): A,U,G,C

↓ Translation [RNA & Proteins different]

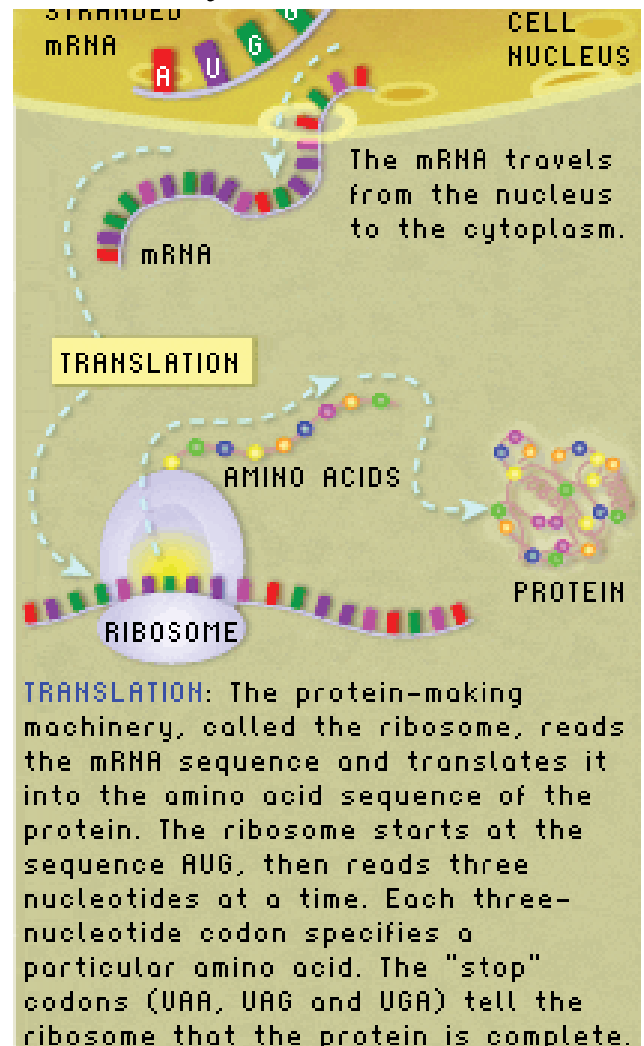
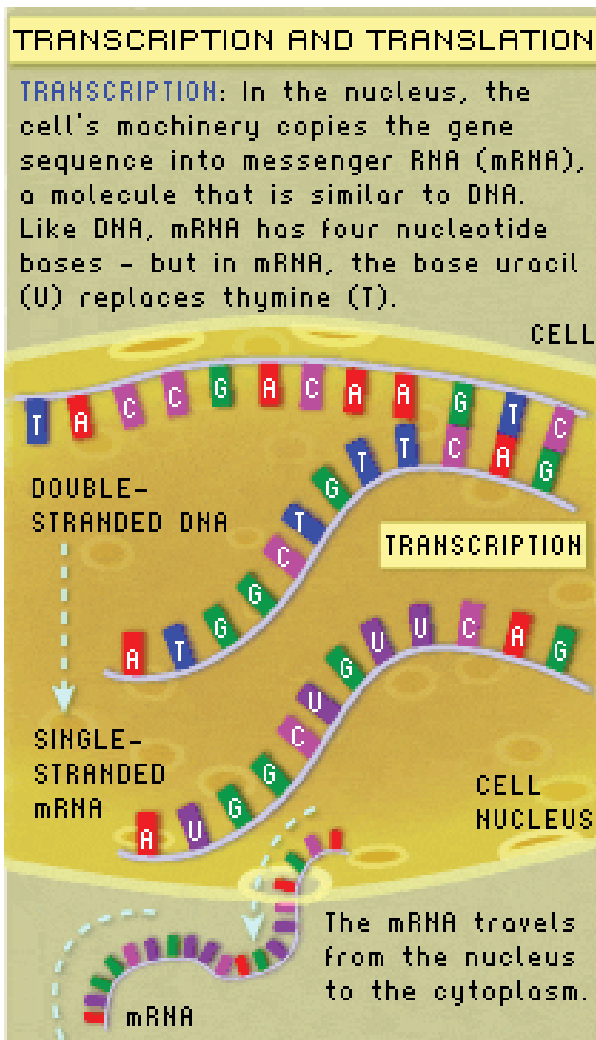
Proteins: series of 20 amino acids: Met-Ala-Val-...

each coded by 3 bases → amino acid

AUG → Methionine; GCU → Alanine; GUU → Valine

Proteins are 3-D strings of linear amino acids

Do everything: structure, enzymes...



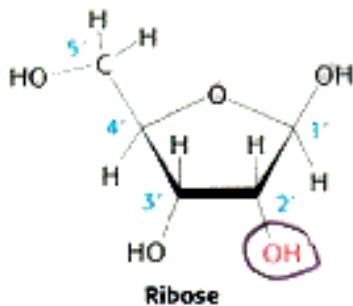
Difference between RNA and DNA is the Sugar + 1 Base

RNA is a string of nucleotides, just like DNA

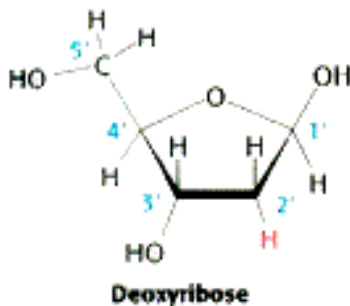
Sugar: Pentose (5 membered ring)

= Ribose: ribonucleic acid (RNA)

= Deoxyribose (2' H) = deoxyribonucleic acid (DNA).

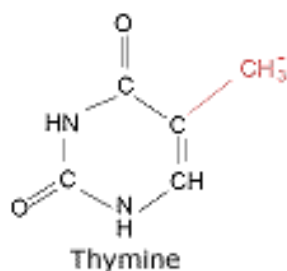
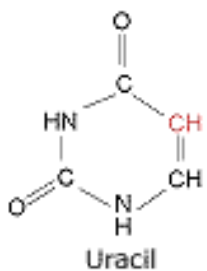


→ RNA → more diverse chemically + structurally
→ storage of info (short-term)
→ Chemical rxns → ENZYME



→ DNA is more chemically stable than RNA. Better for long-term, high reliability storage of information

RNA substitutes Uracil for Thymidine



Uracil will base pair with many groups.

Methyl group restricts uracil (thymine) to pairing only with adenine.

This greatly improves the efficiency of DNA replication, by reducing the rate of mismatches, and thus mutations. Also, methylation protects DNA from viruses.

RNA codon table

This table shows the 64 codons and the amino acid each codon codes for.
The direction is **5' to 3'**.

X = A,U,G,C

1st base		2nd base			Inverse table
		U	C	A	G
	U	UUU (Phe/F) Phenylalanine UUC (Phe/F)Phenylalanine UUA (Leu/L) Leucine UUG (Leu/L)Leucine	UCX (Ser/S) Serine	UAU (Tyr/Y) Tyrosine UAC (Tyr/Y)Tyrosine UAA Ochre (Stop) UAG Amber (Stop)	UGU (Cys/C) Cysteine UGC (Cys/C)Cysteine UGA Opal (Stop) UGG (Trp/W) Tryptophan
	C	CUX (Leu/L)Leucine	CCX (Pro/P) Proline	CAU (His/H) Histidine CAC (His/H)Histidine CAA (Gln/Q) Glutamine CAG (Gln/Q)Glutamine	CGX (Arg/R) Arginine
	A	AUU (Ile/I) Isoleucine AUC (Ile/I)Isoleucine AUA (Ile/I)Isoleucine AUG (Met/M) Methionine , <i>Start</i>	ACX (Thr/T) Threonine	AAU (Asn/N) Asparagine AAC (Asn/N)Asparagine AAA (Lys/K) Lysine AAG (Lys/K)Lysine	AGU (Ser/S)Serine AGC (Ser/S)Serine AGA (Arg/R)Arginine AGG (Arg/R)Arginine
	G	GUX (Val/V) Valine	GCX (Ala/A) Alanine	GAU (Asp/D) Aspartic acid GAC (Asp/D)Aspartic acid GAA (Glu/E) Glutamic acid GAG (Glu/E)Glutamic acid	GGX (Gly/G) Glycine

Notice/Recall that 3 bases cause more than 1 AA.

Inverse table

Ala	GCU, GCC, GCA, GCG	Leu	UUA, UUG, CUU, CUC, CUA, CUG
Arg	CGU, CGC, CGA, CGG, AGA, AGG	Lys	AAA, AAG
Asn	AAU, AAC	Met	AUG
Asp	GAU, GAC	Phe	UUU, UUC
Cys	UGU, UGC	Pro	CCU, CCC, CCA, CCG
Gln	CAA, CAG	Ser	UCU, UCC, UCA, UCG, AGU, AGC
Glu	GAA, GAG	Thr	ACU, ACC, ACA, ACG
Gly	GGU, GGC, GGA, GGG	Trp	UGG
His	CAU, CAC	Tyr	UAU, UAC
Ile	AUU, AUC, AUA	Val	GUU, GUC, GUA, GUG
START	AUG, GUG	STOP	UAG, UGA, UAA

Notice/Recall that more than 1 tri-base = AA.

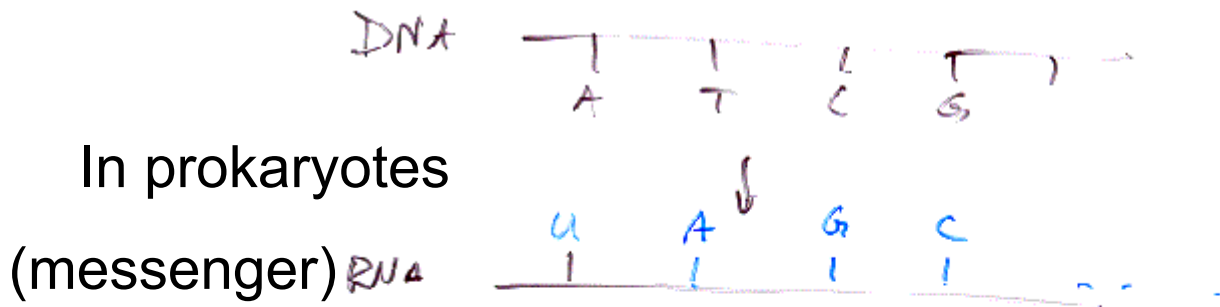
$4^3 = 64$ possible aa \rightarrow actually only 20

Also, notice start & stop codons.

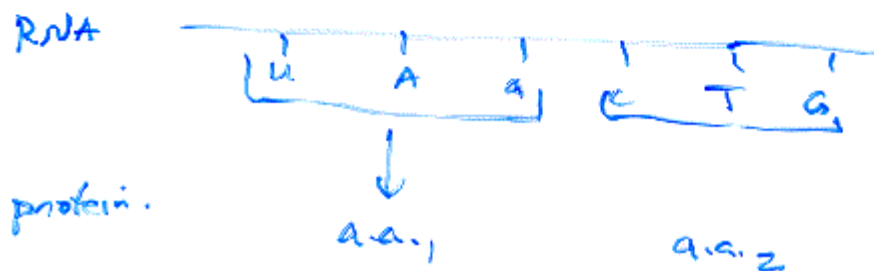
Can tell on DNA where protein starts/stops.

RNA is made from DNA

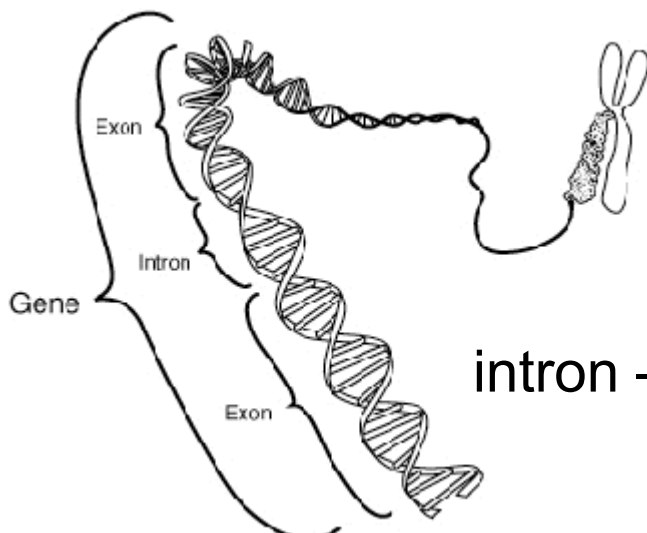
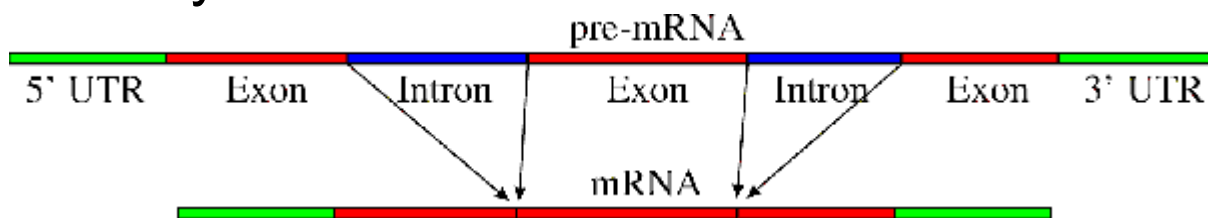
Introns vs. exons in Eukaryotes



prokaryotes introns
are only found in
tRNA and rRNA



In eukaryotes



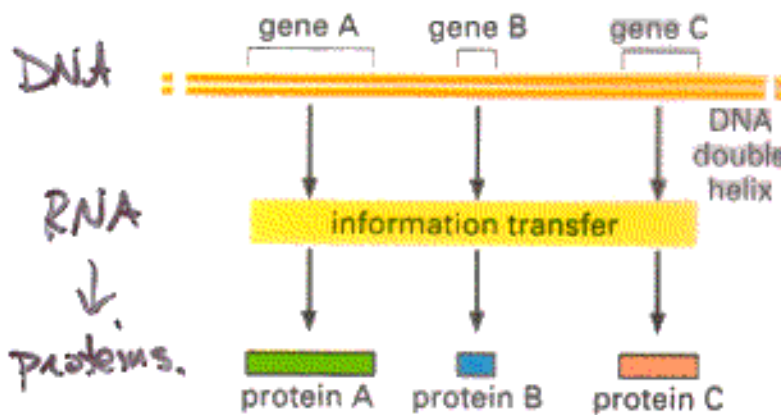
intron – “non-coding region” deleted

1993 Nobel Prize in Medicine to Phillip Allen and Richard J. Roberts

<http://en.wikipedia.org/wiki/Intron>

The structure of DNA provides a means for ...

a. Information storage & access.



Gene = length of DNA coding for polypeptide.

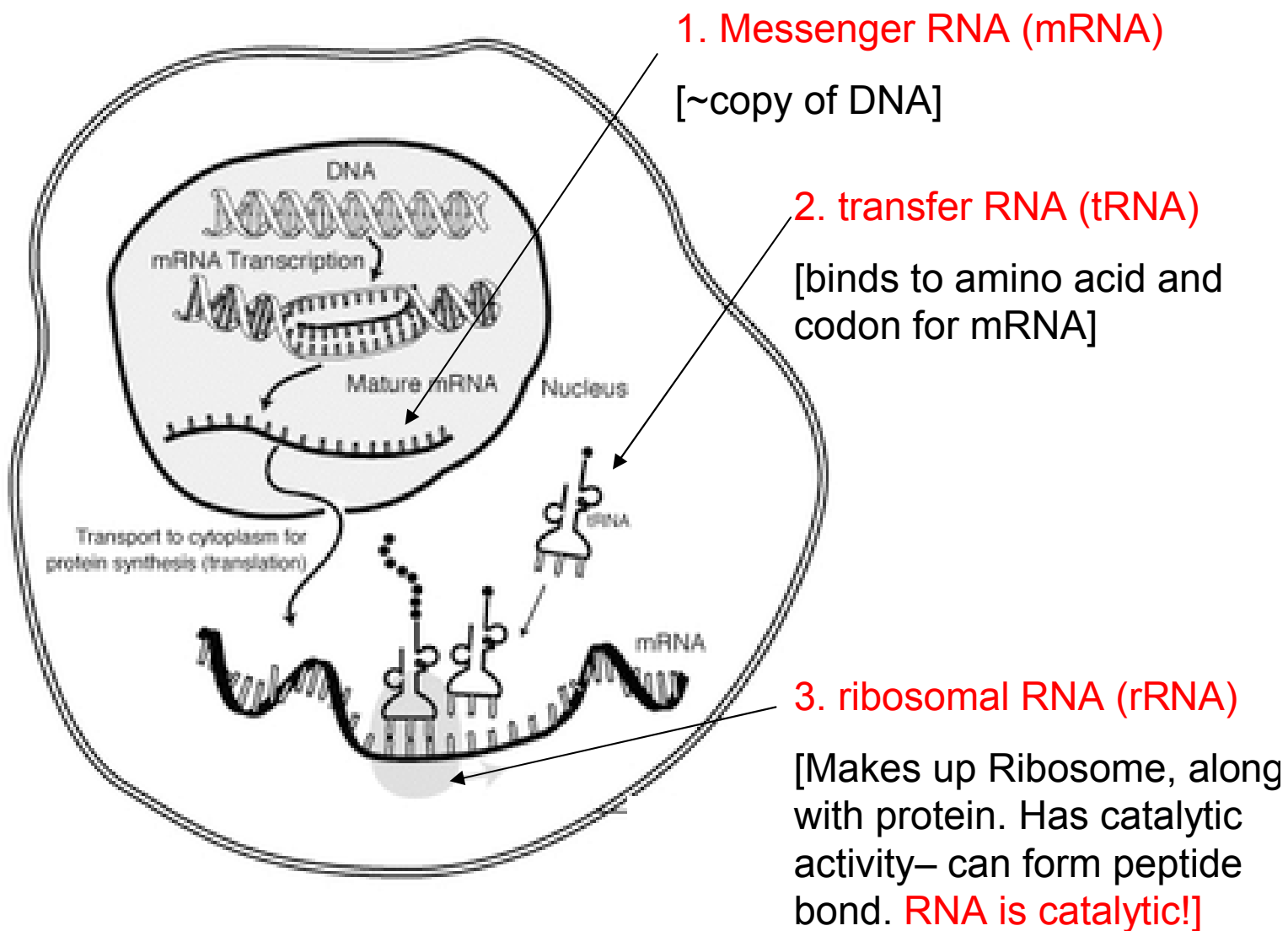
Chunk of letters - tells story
Story - polypeptide / protein

β -globin gene: One of two types of polypeptides making up hemoglobin.

In eukaryotes (humans) ~ 5% of base pairs in DNA code for a protein.
Other 95% useless? remnant from our evolutionary past? Structural role?

CCCTCTCGAGCCACACCCCTAGGGTTGGCCA
ATCTCTCTCCAGGAGCAGGGAGGGCAGGG
CCAGGGCTGGGCATAAAGTCAGGGCAGG
CCATCTATTTGCTTACATTTGCTTCTGAGAC
AACTTTGTTTCACTAGCAACTCAAGCAAGCA
CCAGGGTGGACCTGACTCCTGAGGAGNACT
CTGGGGTTACTGCCCTGTGGGGCAAGGTGA
ACCTGGATGAAGTTGGTGGTGAAGGCCCTGG
GCAGTTGGTATCAAGGTTAGCAAGACNGGT
TTAAGGAGACCAATAGAGACTGGGCATGTG
JAGACAGGAGGACTCTTGGGTTCTGATA
GGCACTGACTCTCTCTGCTTATTTGGTCTAT
TTTCCGACGCTTACGCTGCTGGTCTCTAC
CCTTGGACCCAGAGGTTCTTTGAGTCTCTT
GGGATCTGTGCACTCCTGTATGCTGTATG
GGCAGCCCTAAGGTTGAAGGCTCATGGCAG
AAGTGGCTGGTCTCTTTAGTGATGGCTG
GCTCAGCTGGACAGGCTCAAGGGCAGCTTT
GGCAGACTGAGTGGCTGACCTGTGACAG
CTGACGCTGGATCCGAGGACTTCAGGGTG
AGTCTATGGGACCTTGTATGTTTCTTTCC
CCTCTCTTTCTATGGTTAAGTTCATGTCAT
AGCAAGGGGAGGAGTACAGGCTACAGTTT
ACAATGGGAAACAGAGCAATGATTGCATCH
GTGTCGAAGTCTCAGGATGCTTTAGTTTC
TTTTATTTGCTGTTCAATCAATTTGTCTC
TTTTGTTTAACTCTGCTTCTCTTTTCTT
CTTCTGCGCAATTTTACTATTTACTCTAA
TGCCCTTAACATTTGTGTATAACAAAGGAAA
TATCTCTGAGATACATTAACTTAACAA
AAAGCTTTACACAGTCTGCTAGTACAT
ACTATTTCGAATATATGCTGCTTATTTGC
ATATTCATAATCTCCTACTTATTTCTCT
TTATTTTAACTGATACATTAATCAATAC
ATCTTATGGGTTAAGGTTAATGTTTCAA
TATGTTACACATTTTGAAGCAATGAGGT
AATTTTGCATTTGTAATTTTAAAGAAATGCT
TCTCTCTTTTAAATATCTTTTGTGTTATC
TTATTTCTAATCTTTGCTTATCTCTTCT
TTTCAGGGCAATTAATGATACAAATGATGAT
GGCTCTTTGCAGCATTCTAAAGAAATACRG
TGATTAATTTCTGGGTTAAGGCAATAGCAAT
ATTTCTGCACATAAATAATTTCTGCATATA
ATTTGAACTGATTAAGGCTTTCATATG
CTAATAGGAGCTACAAATCCAGCTACCATTC
TGCTTTCTATTTATGGTTGGGATAAGGCTG
GATTATTTCTAGTCCAGCTAGGGCTTTT
GCTAATCATGTTCAATCTCTTATCTTCT
CCACAGTTCTTGGGCAAGCTCTGGGCTG
TGTGCTGGCTCATCTTTGGCAAGAAAT
CACTCCACAGTGCAGGCTGCTATCAGAA
ACTGGTGGCTGGTGTGCTTAATGGCTGGC
CCACAGTACCTAAGCTGGCTTTCTTGC
TGTCCAATTTCTATTAAAGGTTCTTTGTT
CCCTAAGTCCAACTACTAACTGGGGATA
CTATGAAGGGCTTGGCATCTGGATCTG
CCCTAATAAAAACATTTATTTCTATTGCA
GGAAGCATTTAAATATTCTGAATATTT
ACTAAAGGGGAAATGGGGAGGCTAGTGA
TTTAAAGCATAAAGAAATGATGAGCTGTT
AAACCTTTGGAAATACACTATATCTTAA
CTCATGAAAGAGGTTGAGGCTGCAACAG
CTATGCAATTTGGCAAGGCTGCTGATGC
CTATGCTTATTTATGCTTCAAGAAAGGAT
CTTTGAGAGGCTTGAATTTGAGGTTAAG
TTTGTCTATGCTGATTTTACATTAAT
TGTCTAGCTGCTGCTGATGAGTCTTTCT

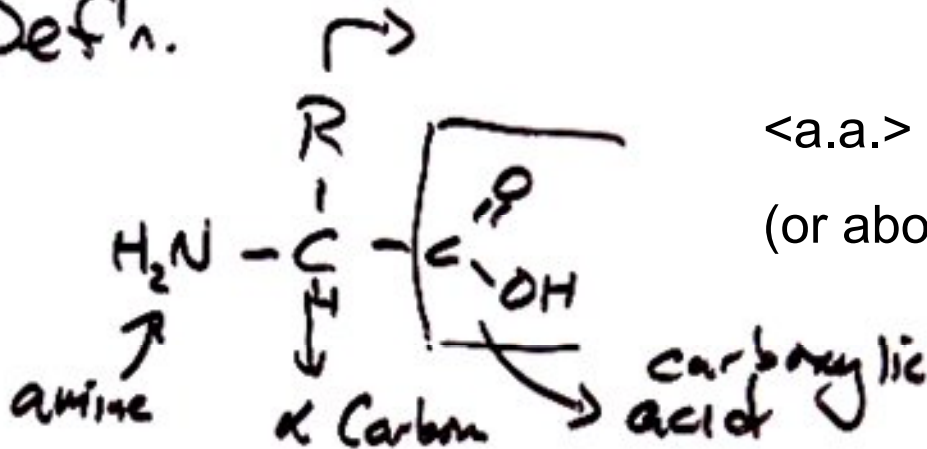
RNA has 3 different uses, called 3 different names, (mRNA, tRNA, rRNA)



http://en.wikipedia.org/wiki/Messenger_RNA

Amino acids

Def'n.

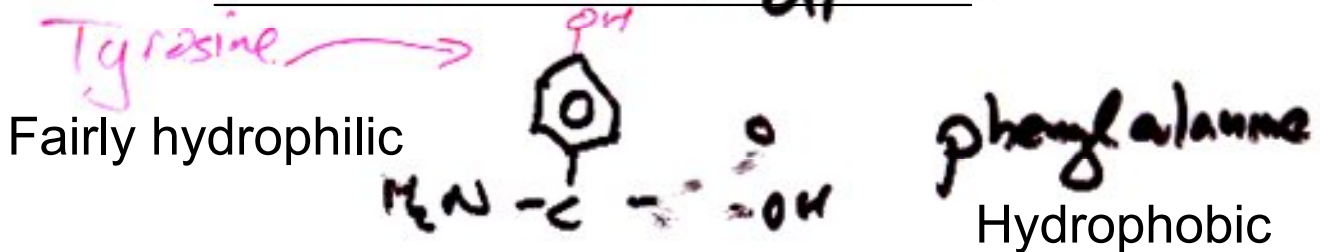
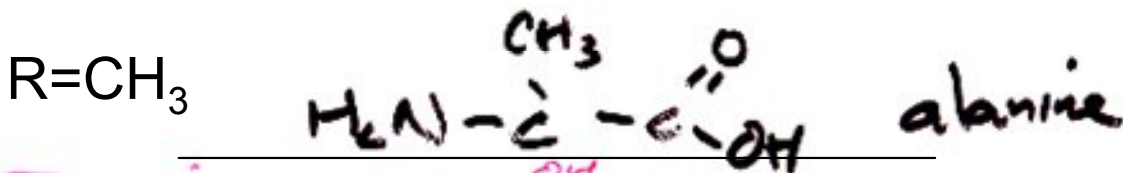
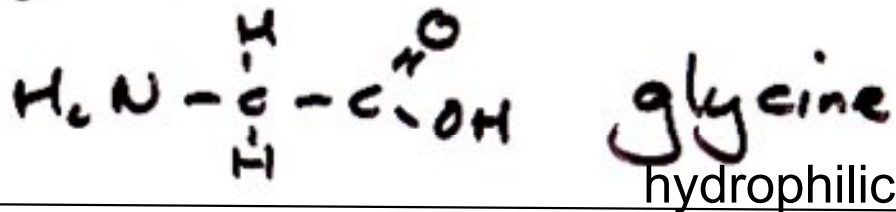


<a.a.> ~110 grams/mole
(or about 100g/mole)

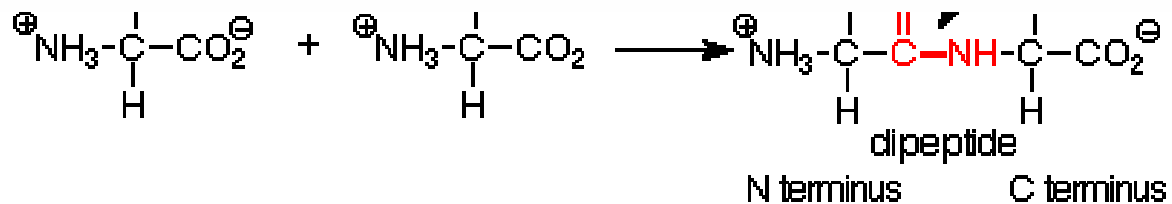
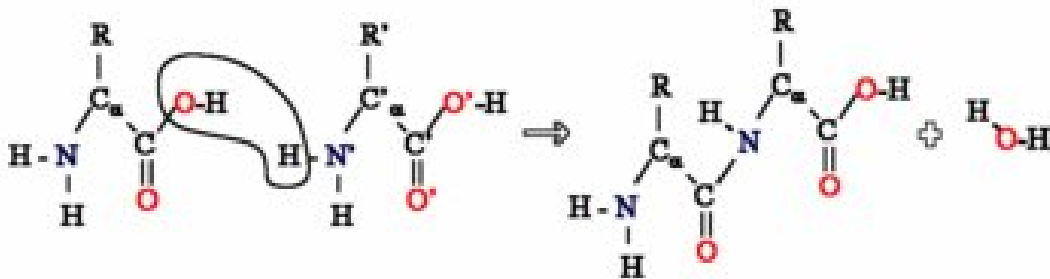
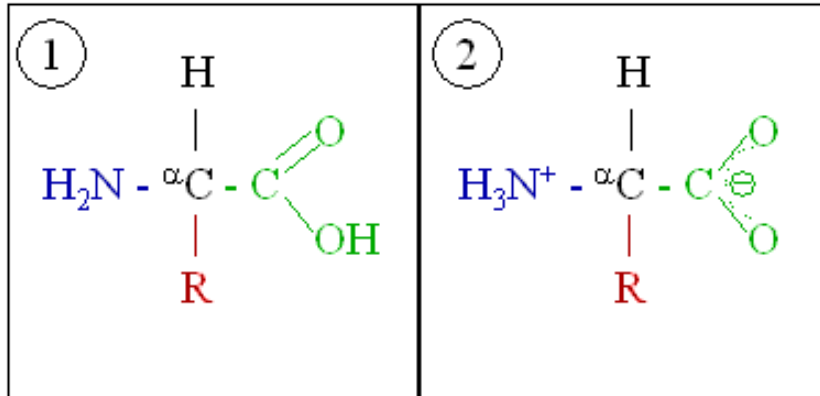
R - general designation for
some organic structure
= side group.

20 diff. types of a.a.
" " R groups.

Simplest $\text{R} = \text{H}$



Amino Acids undergo condensation reaction to form peptides



Free H₂O : lots of entropy gained



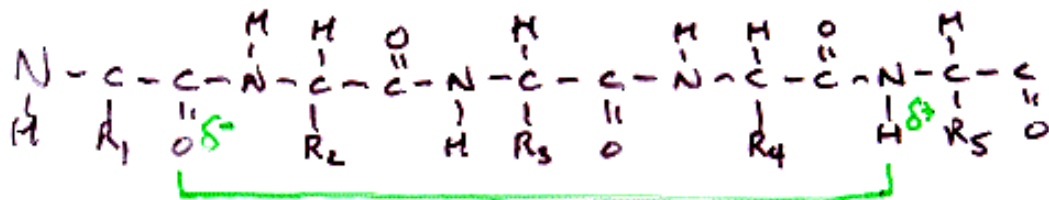
Linear sequence of ~ 20 amino acids

Can get enormous diversity and
function with Proteins

Important Example(s) of
arrangements of a.a.'s
(called secondary structure)

α -helix ; β -sheet

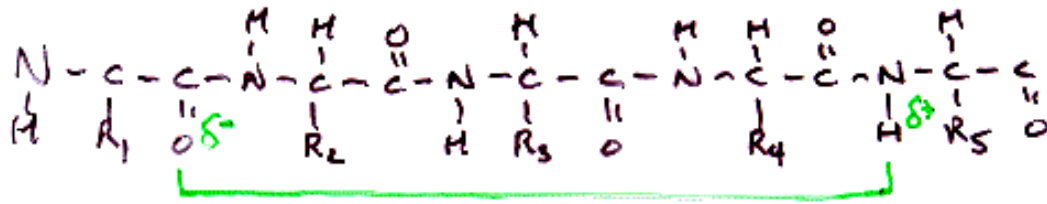
α -helix



$\text{C}=\text{O}$ of one a.a. can H-bond to NH
group of a.a. 4 a.a. away.

Secondary Structure

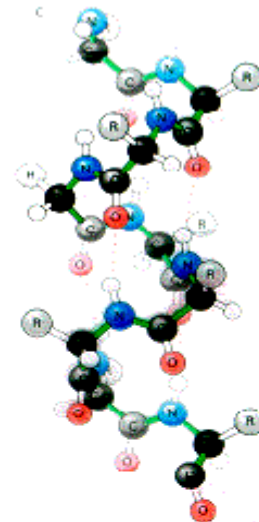
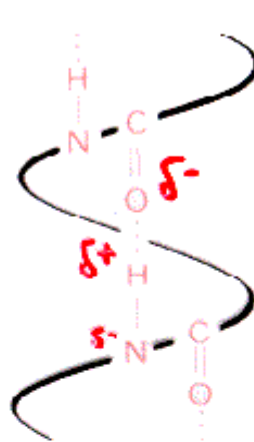
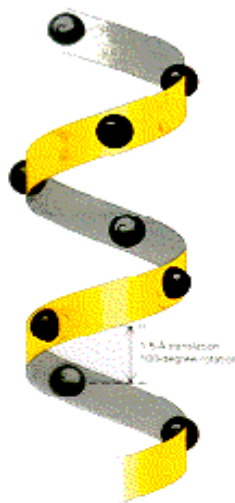
α -helix, β -sheets



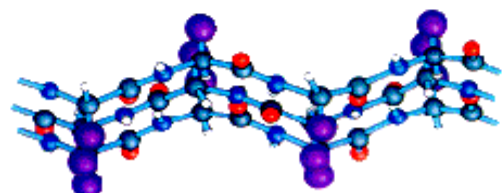
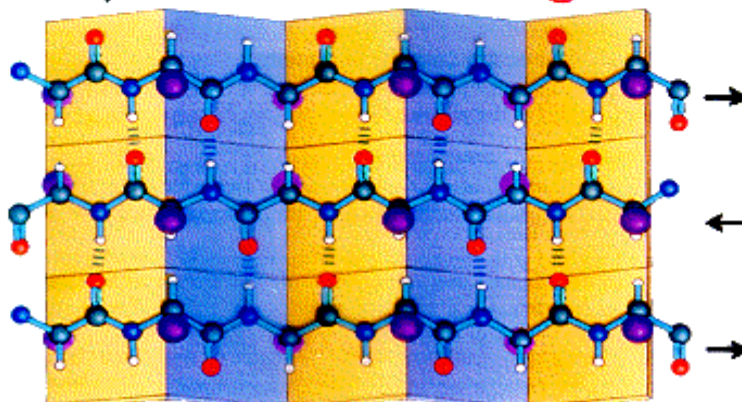
$C=O$ of one a.a. can H-bond to NH group of a.a. 4 a.a. away.

Stabilized by H-bond (about 2 H bond per a.a.)

α -helix



β -sheet - stab. by H-bonds.



α -helix, β -sheets depends on specific amino acids

Typical size of genes & proteins

Length in # a.a. of protein?

average mol. weight a.a. = 110

small Protein 100 a.a. \approx 10,000 m.w.

Large protein 1000 aa \approx 100,000 m.w.

antibodies - 200,000

Cystic fibrosis protein \approx million m.w. (?)

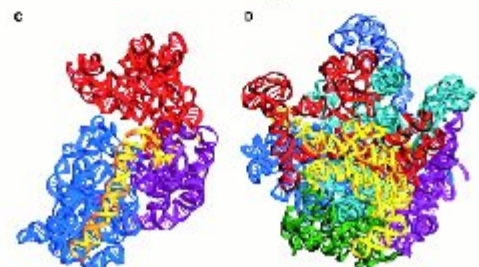
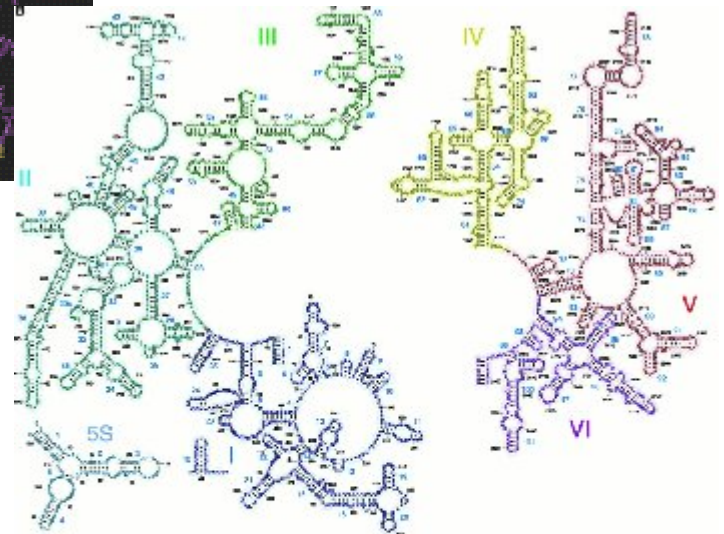
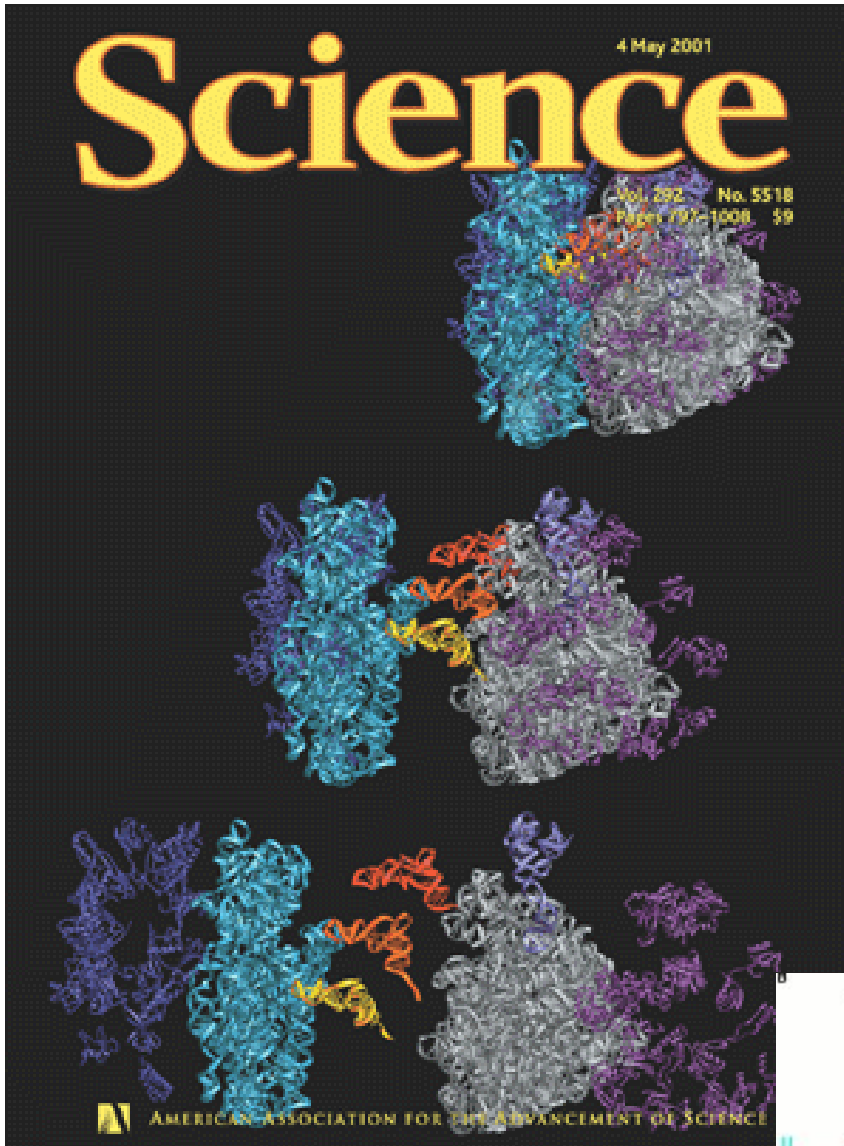
3 bases in DNA codes for 1 a.a.

Length of gene \approx 3 x # a.a. protein

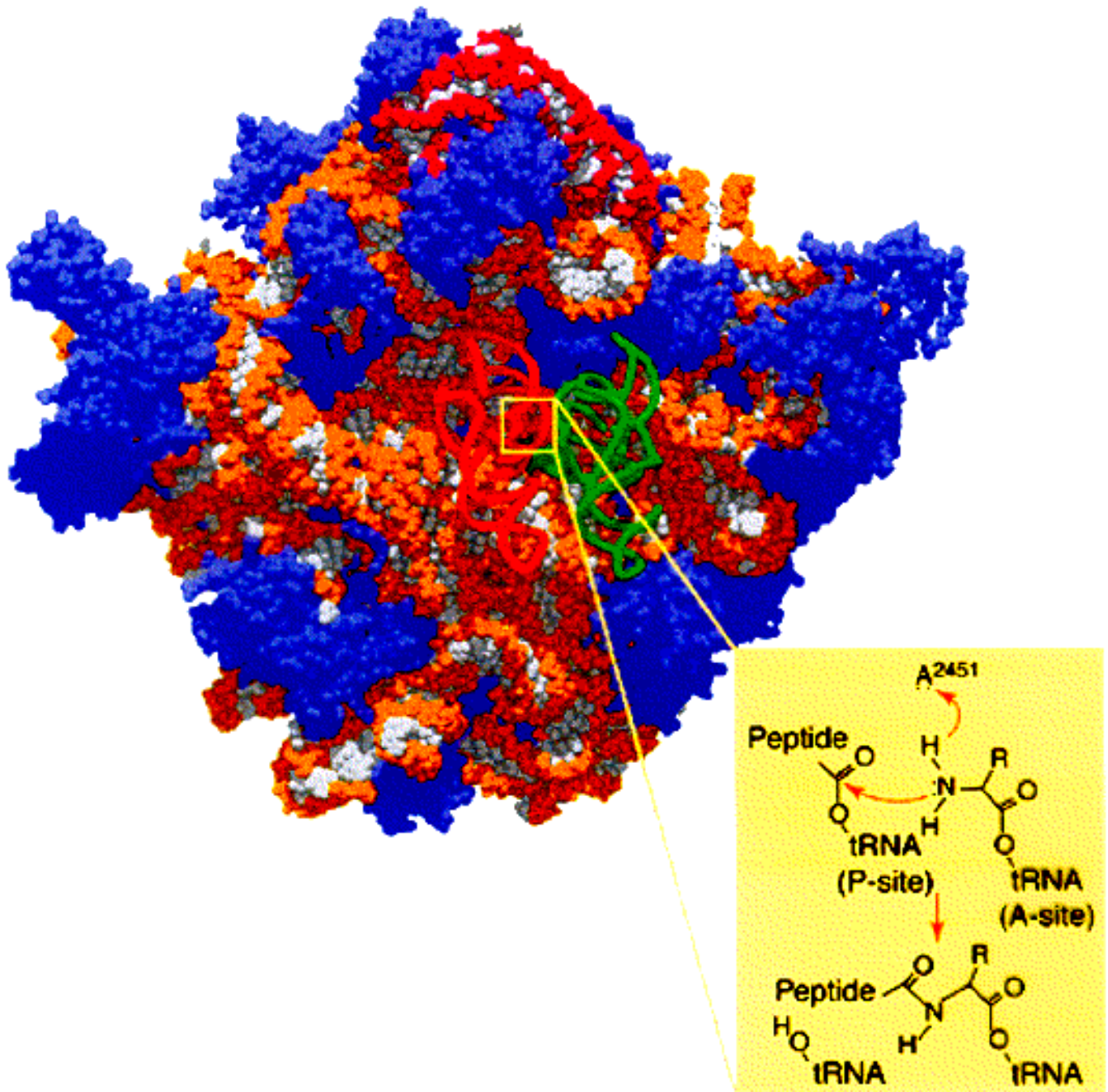
Typ. gene few ~~hundred~~ thousand bases

Cys fibrosis - random mutation - more likely to happen in big gene (just by chance)
→ diseases often happen in large genes/proteins.

Ribosome (2001)



Structure and function of a Ribosome



The large subunit of the ribosome ; proteins in purple; rRNA in orange, white, burgundy; tRNA in green red.

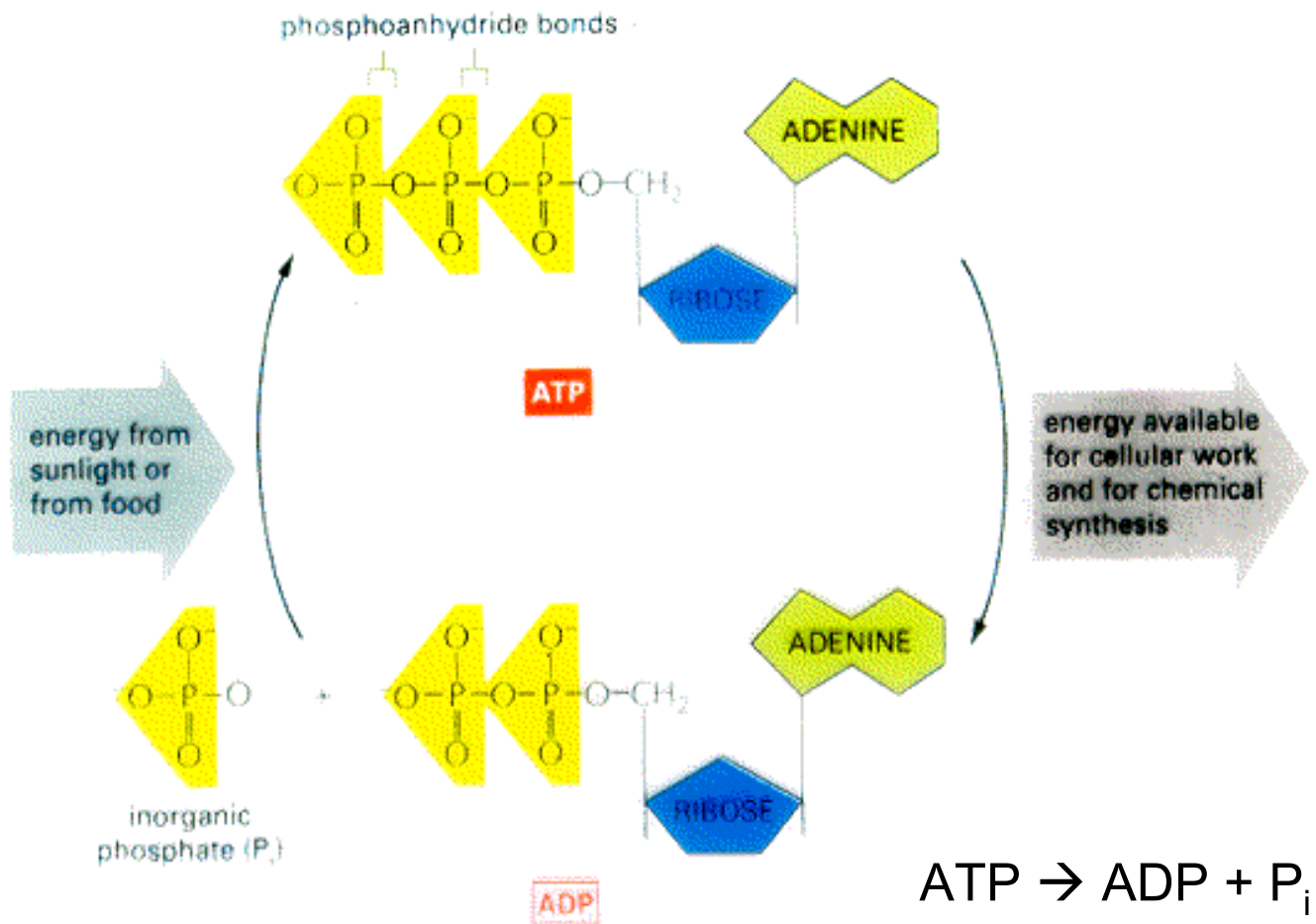
rRNA takes Nitrogen's proton, making reaction forming peptide bond easier.

RNA is an enzyme!

ATP is the universal food currency of all cells

Nucleic acids perform several roles

1. Immediate source of energy in cell (ATP)



ATP is high energy because of electrostatic repulsion of negatively charged oxygens.

(Entropy is also generally increased because 1 molecule \rightarrow 2 molecules.)

Adenosine rings used as recognition/binding site by enzymes

ATP, your “just-in-time” energy source

1/2 dance! Amazing facts about ATP

- Only about 100g of ATP in body, yet it's main source of immediate energy for most chemical reactions.

How possible?

- Because ATP is turned over (used and restored) very quickly.
- Once ATP made, stays around in cell about 1 minute.
- Typically use about 40kg (90 lbs!) of ATP in 24 hrs.
- In 2 hr run can use 60kg (132 lbs!).

How can use 90 lbs?

$$\text{Net weight} = W_{\text{ATP}} - W_{\text{ADP}}$$

Energetics of ATP

1 ATP = 80-100 pN-nm of energy at 37 °C

= 20-25 kT of energy

(much more than $kT = 4$ pN-nm)

A lot of energy

Why do I say 80 to 100 pN-nm? Why not an exact amount?

What counts is ΔG , not ΔE , where $\Delta G = \Delta E - T\Delta S$

$\text{ATP} \rightarrow \text{ADP} + \text{P}_i$; depends on $[\text{ADP}]$ & also $[\text{P}_i]$ concentration

Class evaluation

1. What was the most interesting thing you learned in class today?
2. What are you confused about?
3. Related to today's subject, what would you like to know more about?
4. Any helpful comments.

Answer, and turn in at the end of class.