

# *DNA-based Self-Assembly*

We've discussed inorganic self-assembly in forms such as crystal growth

We've described how organic self-assembly produces things like nanotubes and Buckyballs

Today we'll introduce the ultimate self-assembly process, that based on DNA

Note that as a non-molecular biologist, in preparing this lecture I've consulted texts including:

Molecular Biology of the Cell by Bruce Alberts et al.

The World of the Cell by W.M. Becker, J.B. Reece & M.F. Poenie

Biochemistry by D. Voet and J. Voet

Which I will credit in figures as: "Alberts," "World of the Cell" or "Voet & Voet"

(As always, uncredited figures are from the UVA Virtual Lab website)

## *Let's start at the beginning*

By early 1950's, scientists were hot on the trail of molecule that programmed cells:

- Segments of molecule were ~ cylindrical in shape, 2 nm in diameter

- X-ray diffraction (by Wilkins & Franklin of Kings College) indicated:

  - Periodic structures along the cylinder with spacings of 0.34 nm and 3.4 nm

  - Pattern of diffraction spots also suggested the presence of a spiral structure

- When chemically disassembled, it yielded large quantities of four bases:

  - Adenine (A), Cytosine (C), Guanine (G) and Thymine (T)

- Proportions varied, but Erwin Chargaff (Columbia University) noted:

  - Concentrations of A & T were always the same

  - Concentrations of G & C were always the same

## *Enter James Watson and Francis Crick of Cambridge University*

From diffraction data, they developed model with a pair of sugar-phosphate polymer helices

They assumed information was encoded by location of bases

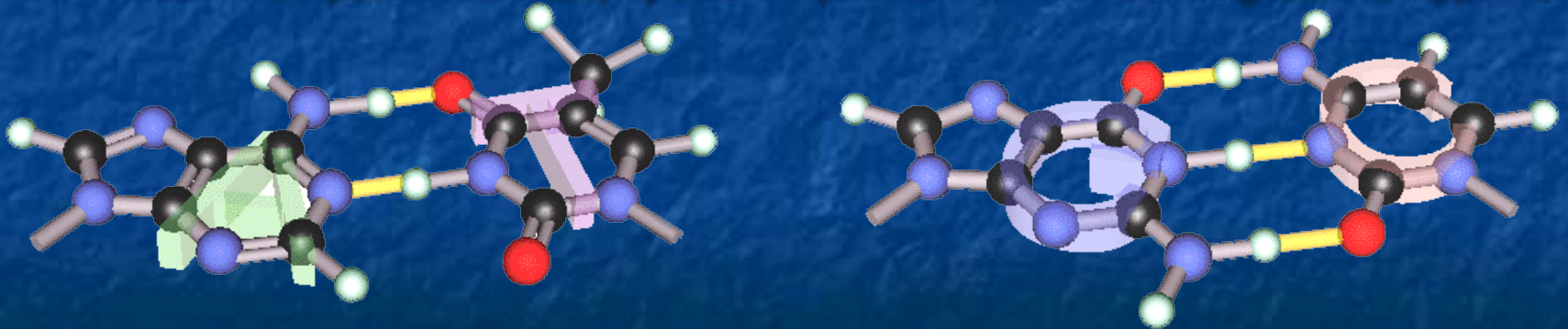
But they'd been working on ways of fitting **like pairs of bases** inside the helices

(Hadn't they paid any attention to Chargaff's papers!)

One morning, helped by lab mate Jerry Donohue, Watson noticed:

Hydrogen bonding could cause A & T bases to pair, and G & C bases to pair =>

Attachment points identically spaced & symmetrical => Interchangeable & flippable!



Figures: [https://WeCanFigureThisOut.org/VL/DNA\\_close\\_up.htm](https://WeCanFigureThisOut.org/VL/DNA_close_up.htm) or [DNA Lecture - Supporting Materials](#)

## *Then Watson and Crick . . .*

Worked through morning to fit base pairs inside their double helix model

When done, they had found that:

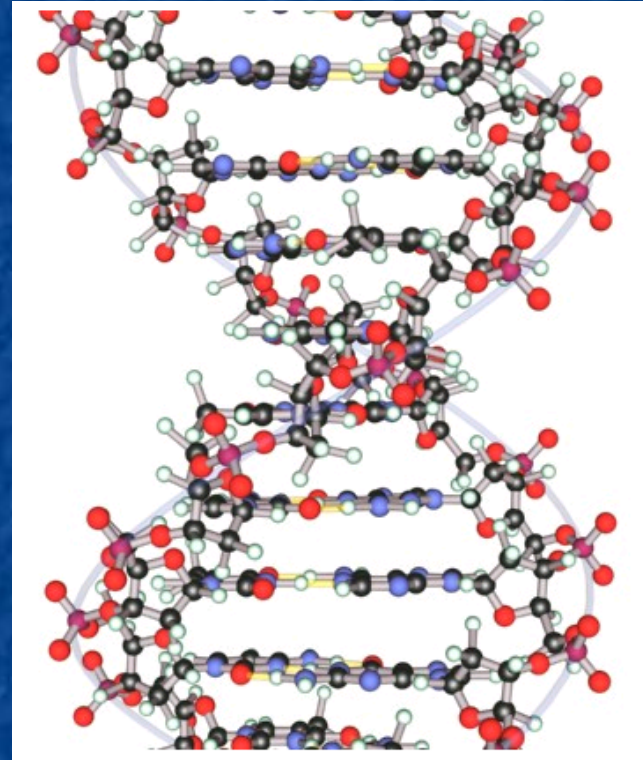
Helical diameter was right: 2 nm

Diffraction spacing of 0.34 nm =

Spacing between parallel base pairs

Diffraction spacing of 3.4 nm =

1 revolution = stack of 10 base pairs

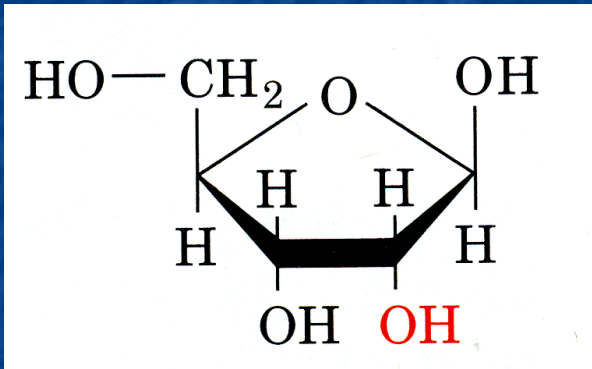


Wrote up a short two page journal article describing their finding

# "Deoxyribonucleic Acid" What does it mean?

Most of the name refers to the make-up of the spiral backbone

One of the most basic sugars is called ribose:



(from Voet & Voet p.18)

From class of "carbohydrates"

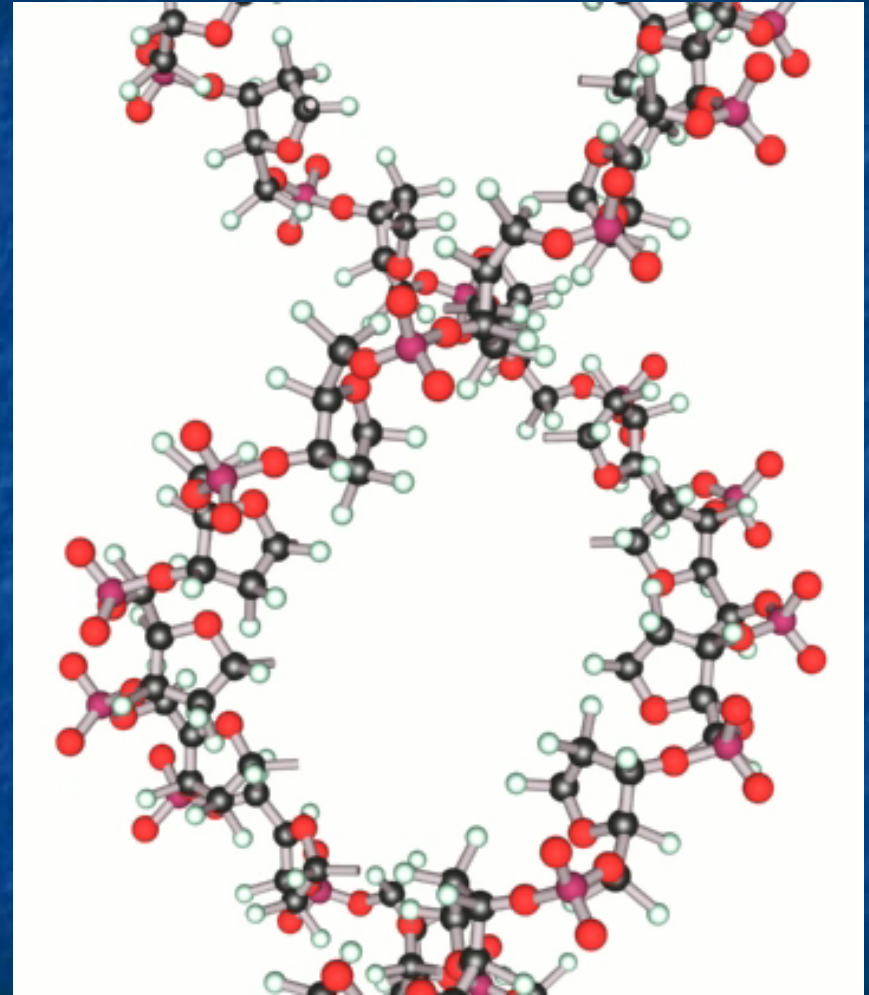
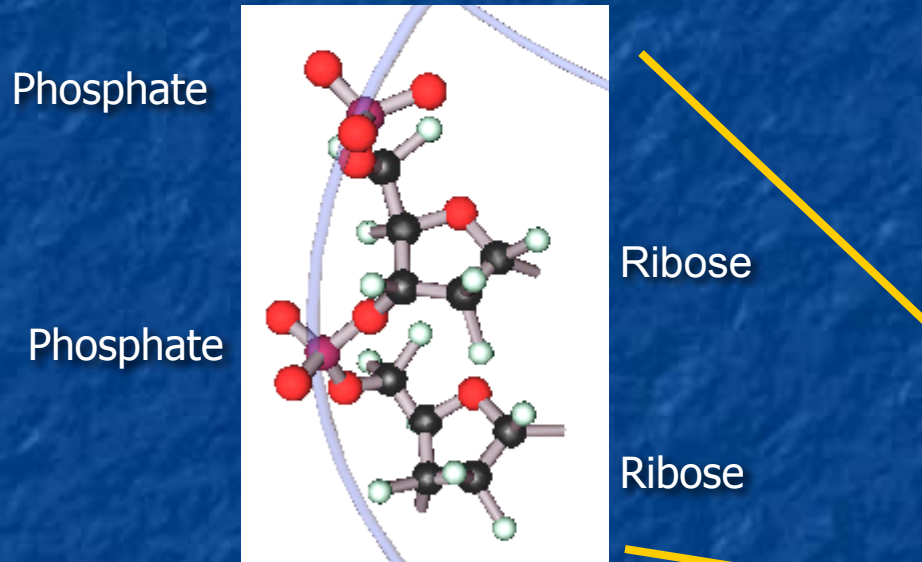
Named because once thought to consist of carbon structures surrounded by water

True structure **splits** water's H and OH

2 - deoxyribose is a variant where the **OH** group in red is replaced by just H

Then, when phosphate group replaces OH at upper left, this yields DNA's backbone:

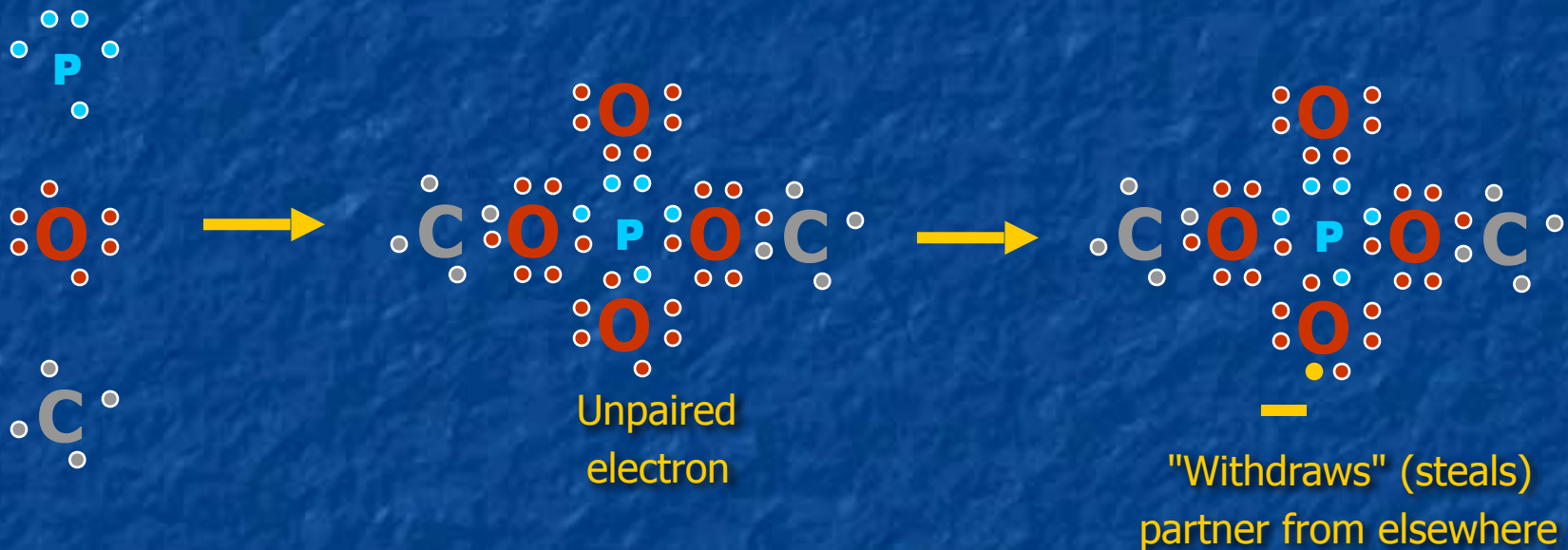
# DNA helical backbone of 2-Deoxyribose - Phosphate



Figures: [https://WeCanFigureThisOut.org/VL/DNA\\_big\\_picture.htm](https://WeCanFigureThisOut.org/VL/DNA_big_picture.htm)

# An important feature of the phosphate: It's charged!

As we learned in last lecture based on Lewis structure of phosphate group:



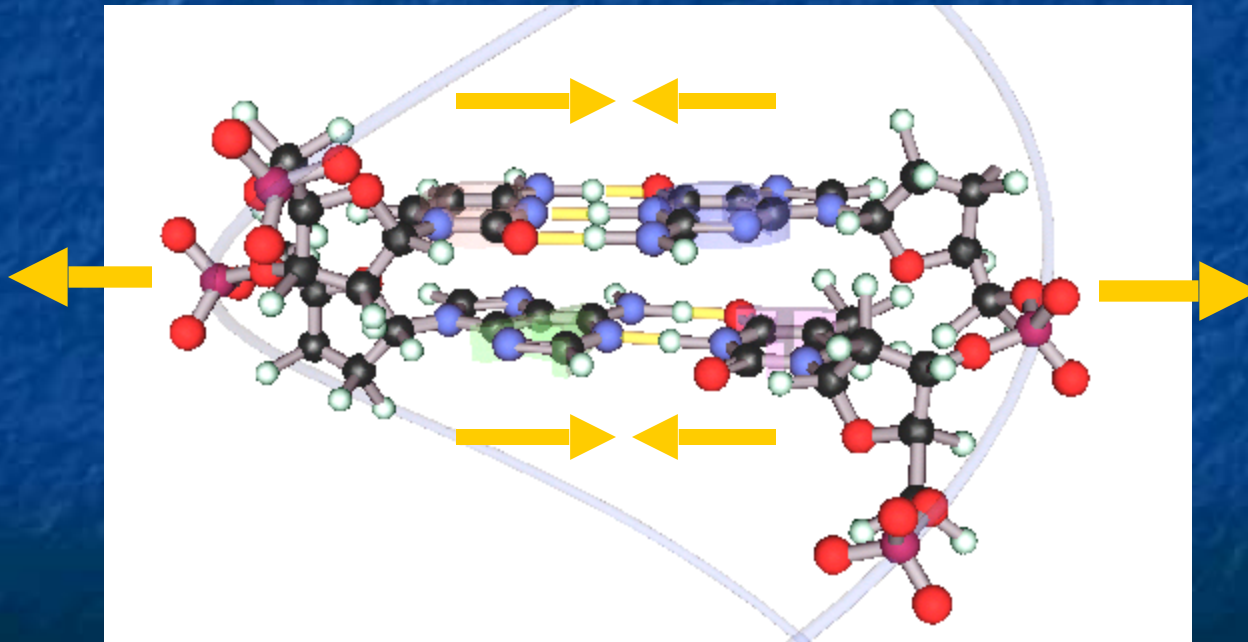
## *Balance of forces:*

Negative charge on backbones tries to push DNA strands apart

Offset by hydrogen bonding between bases

RESULT: DNA strands come apart easily - "denature" (e.g. upon heating to 95 C)

Which turns out to be part of their job!





# Terminology break:

## Nucleotide

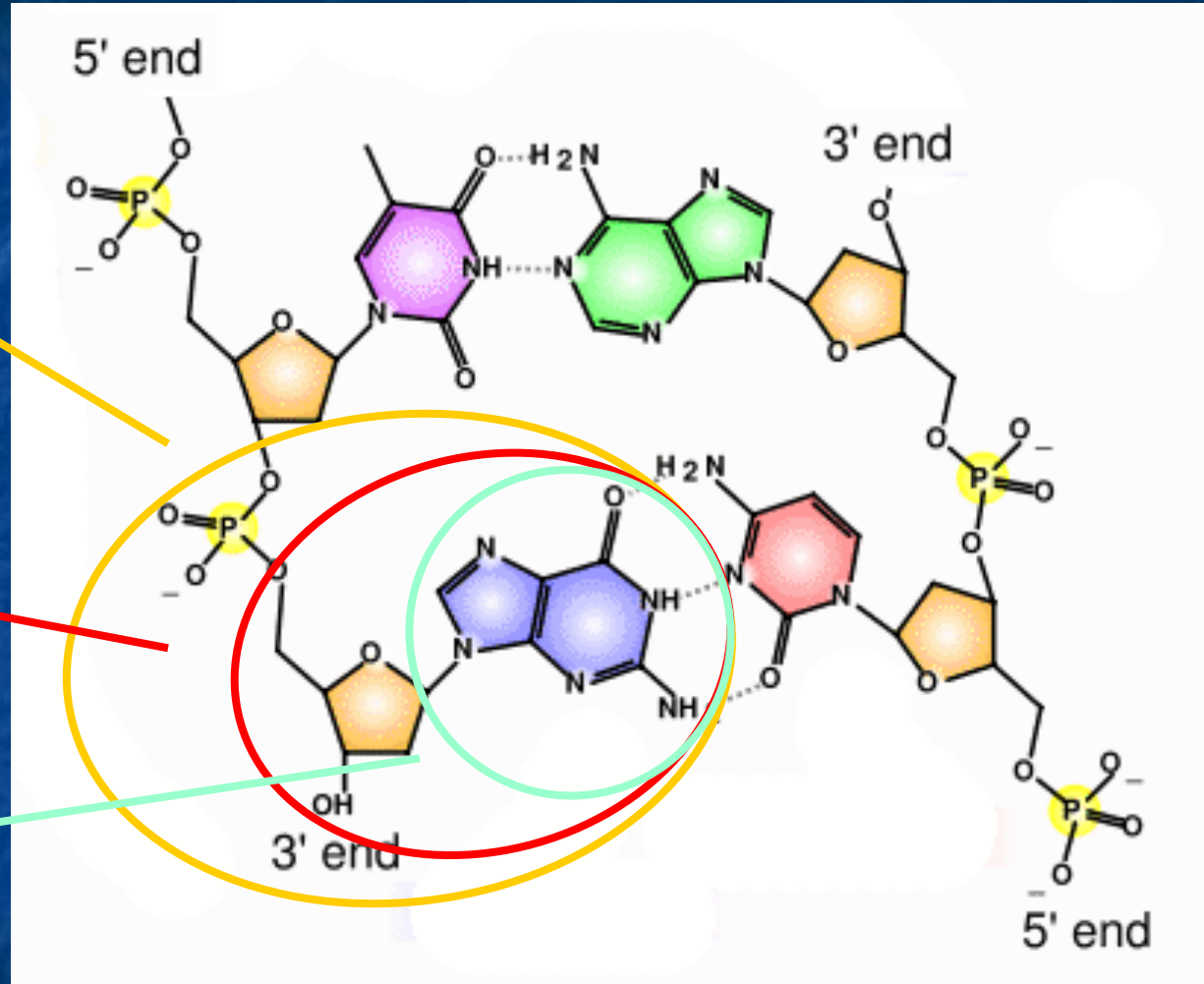
Complete building block of base, ribose and phosphate

## Nucleoside

Base + ribose

## Base

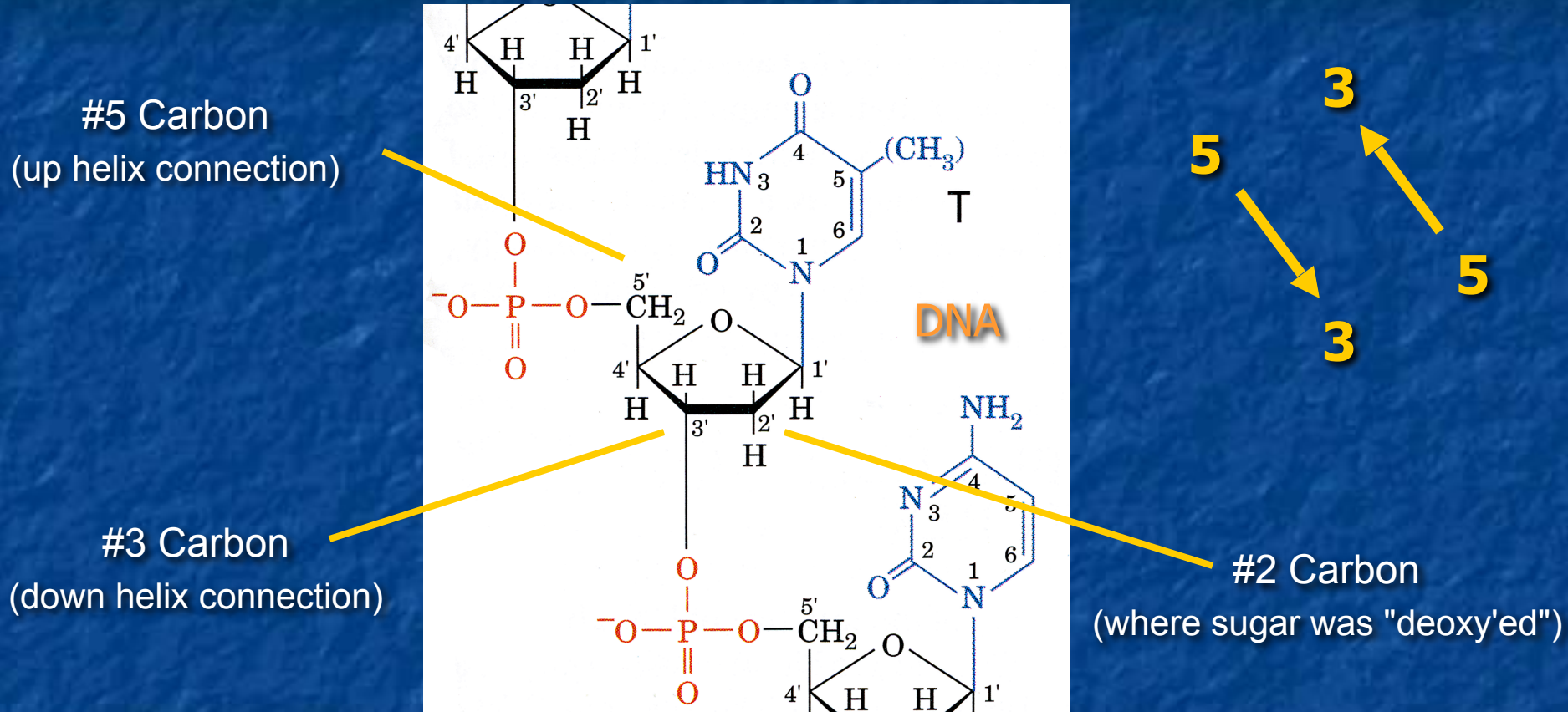
A, C, G or T



(Modified from Wikipedia page on DNA)

# And what about the 3' and 5' end labels?

Goes back to Chemist's numbering of carbon sites on the 2-deoxyribose:



(modified from Voet & Voet figure 28.1)

**DNA is built up by adding new "nucleotide" units to 3' ends of helices**

**Because the facilitating enzyme will ONLY bind itself to that end!**

## Information content of DNA?

At each helix attachment site, four alternatives: CG, GC, AT, TA base pairing

### Total number of base sites in DNA?

SV40 virus	$5 \times 10^3$
T2 virus	$2 \times 10^5$
Peas	$5 \times 10^9$
Fruit Fly	$2 \times 10^8$
Human	$3 \times 10^9$
Trillium lily	$1 \times 10^{11}$

### Total length of human DNA?

3.2 billion sites x 0.34 nm base spacing →

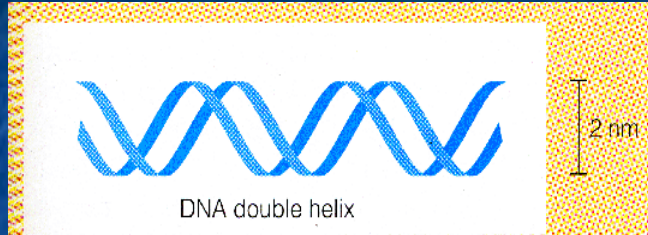
Total length of human DNA = 1.1 meter

# Divided between 23 chromosome pairs in humans:

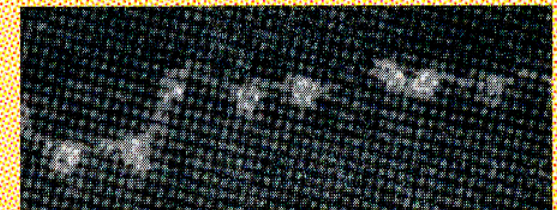
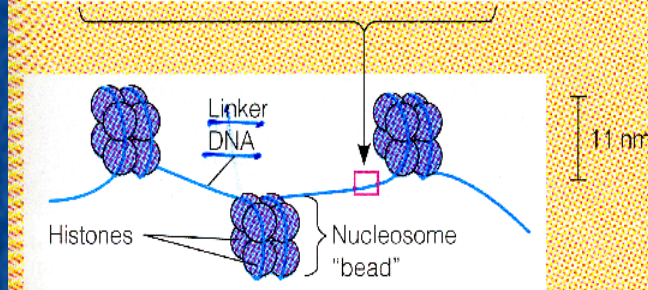
Schematic:

Electron Micrograph:

Take DNA double helix strands

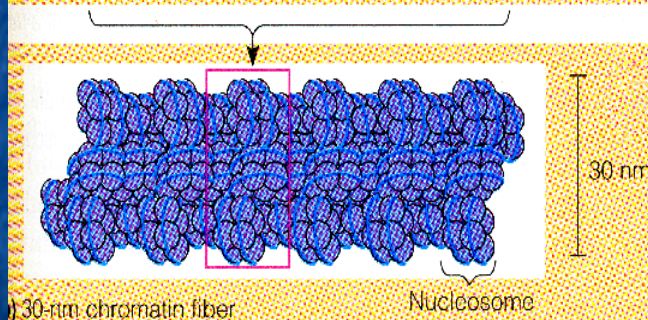


Wrap around **Histone** beads . . .



Nucleosomes ("beads on a string")

Coil necklace of beads into chains . . .

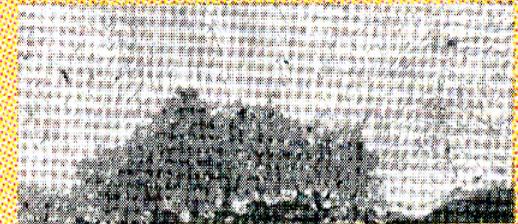
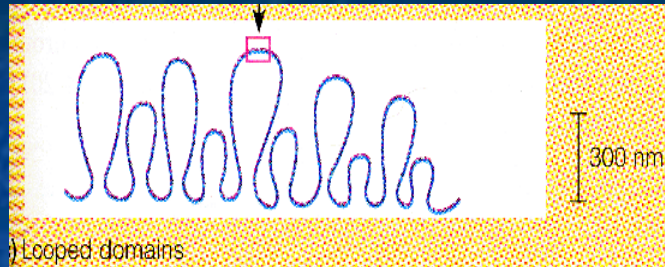


(from World of the Cell figure 14.8)

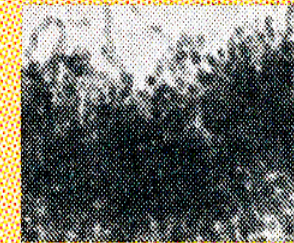
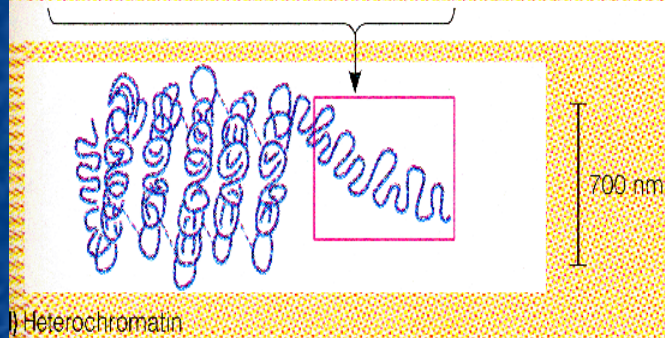
Schematic:

Electron Micrograph:

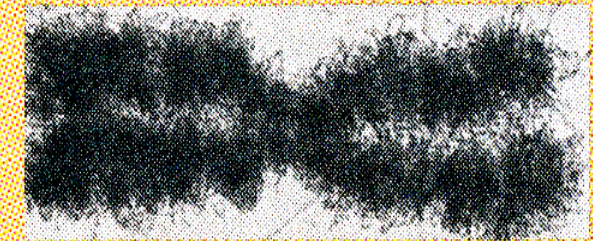
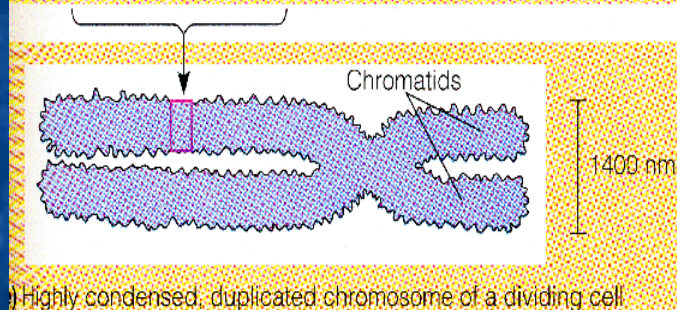
Loop those chains . . .



And then coil them . . .



To form commonly depicted form of **Chromosomes**:



(from World of the Cell figure 14.8)

*Or as an animation:*



From: [www.youtube.com/watch?v=5UoKYGKxxMI&feature=related](http://www.youtube.com/watch?v=5UoKYGKxxMI&feature=related)

(or for cached version of 1<sup>st</sup> part see supporting webpage: [DNA Self-Assembly - Supporting Materials – DNA\\_packing\\_and\\_replication](#))

*A Hands-on Introduction to Nanoscience: [WeCanFigureThisOut.org/NANO/Nano\\_home.htm](http://WeCanFigureThisOut.org/NANO/Nano_home.htm)*

## *But an important point of clarification!*

Most of the time chromosomes are **NOT** in final highly organized “H” like structure!

The final “H” shaped structure adopted **only** during cell division (mitosis)!!

When tight organization is required for division of DNA between daughter cells

Chromosomes spend **most** of their time in the form of simpler long threads

The non-cell-dividing state is called “interphase”

This normal strung out form makes DNA strand separation much easier

**As required in the DNA replication process described below!**

(From Wikipedia: "Chromosomes")

In human beings:

23 pairs of chromosomes:

Each set containing net ~ 1 meter of DNA

3 billion base pair long DNA strand:

Very different organization between animals:

Cat DNA is separated into 38 chromosome pairs!

Fine, but what is a "Gene" ?

Concept goes back to late 1800's

LONG before knowledge of DNA!

= That "thing" that programs one "inherited trait"

Chromosome	Genes	Total bases	Sequenced bases <sup>[54]</sup>
1	3,148	247,200,000	224,999,719
2	902	242,750,000	237,712,649
3	1,436	199,450,000	194,704,827
4	453	191,260,000	187,297,063
5	609	180,840,000	177,702,766
6	1,585	170,900,000	167,273,992
7	1,824	158,820,000	154,952,424
8	781	146,270,000	142,612,826
9	1,229	140,440,000	120,312,298
10	1,312	135,370,000	131,624,737
11	405	134,450,000	131,130,853
12	1,330	132,290,000	130,303,534
13	623	114,130,000	95,559,980
14	886	106,360,000	88,290,585
15	676	100,340,000	81,341,915
16	898	88,820,000	78,884,754
17	1,367	78,650,000	77,800,220
18	365	76,120,000	74,656,155
19	1,553	63,810,000	55,785,651
20	816	62,440,000	59,505,254
21	446	46,940,000	34,171,998
22	595	49,530,000	34,893,953
X (sex chromosome)	1,093	154,910,000	151,058,754
Y (sex chromosome)	125	57,740,000	22,429,293

To establish precise Gene to DNA link, first need to know more about what DNA DOES!

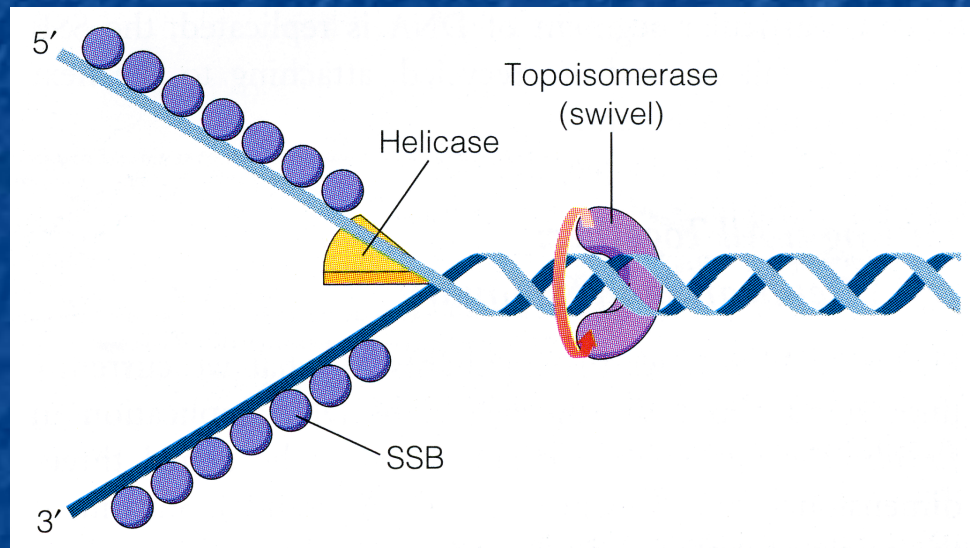


# DNA Self-Assembly Task #1: Replication

Replicate so that copies of DNA can go into nucleus of every cell:

Helicase = enzyme (catalyst) that splits apart DNA strands

SSB = single strand binding proteins that stabilize strands once separated

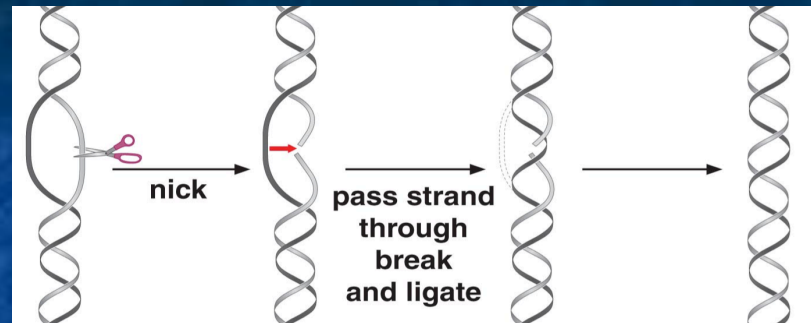


(from World of the Cell  
figure 15.10)

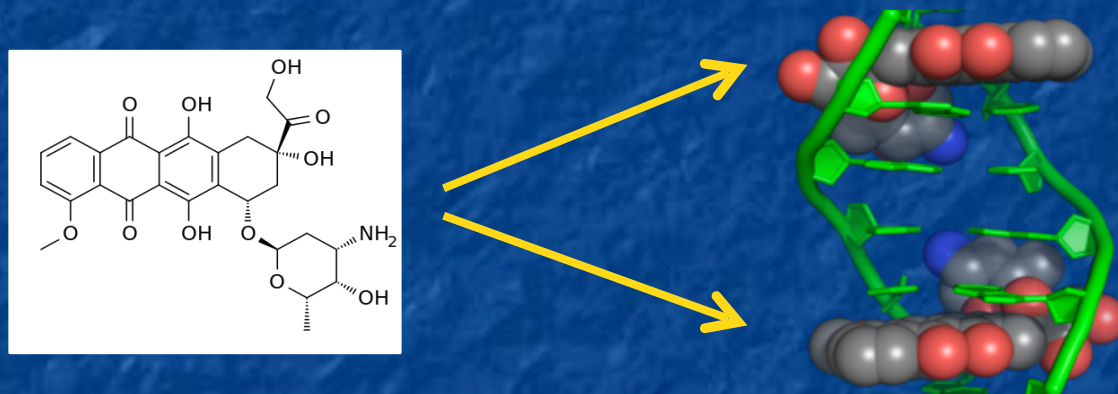
But my favorite is Topoisomerase:

To separate strands (like those a of rope) one MUST untwist the strands

*Topoisomerase does this by cutting, then reconnecting, one DNA strand:*



**DIGRESSION:** The chemical **doxorubicin** can **block** the DNA's reconnection:



Leading to doxorubicin's use in chemotherapy to stop cancer cell reproduction

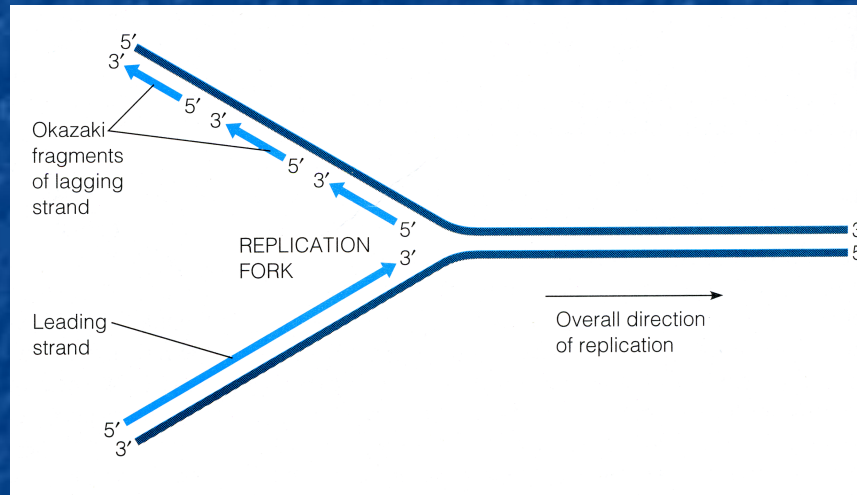
Which would be fantastic if it didn't kill off healthy cells the same way!

# Once separated, complementary strands can be assembled

New strand has complementary bases: G's opposite old strand's C's, A's opposite T's . . .

However, new strands **MUST** grow by adding nucleotide units to their 3' ends

One of the new strands **MUST** end up facing wrong way:



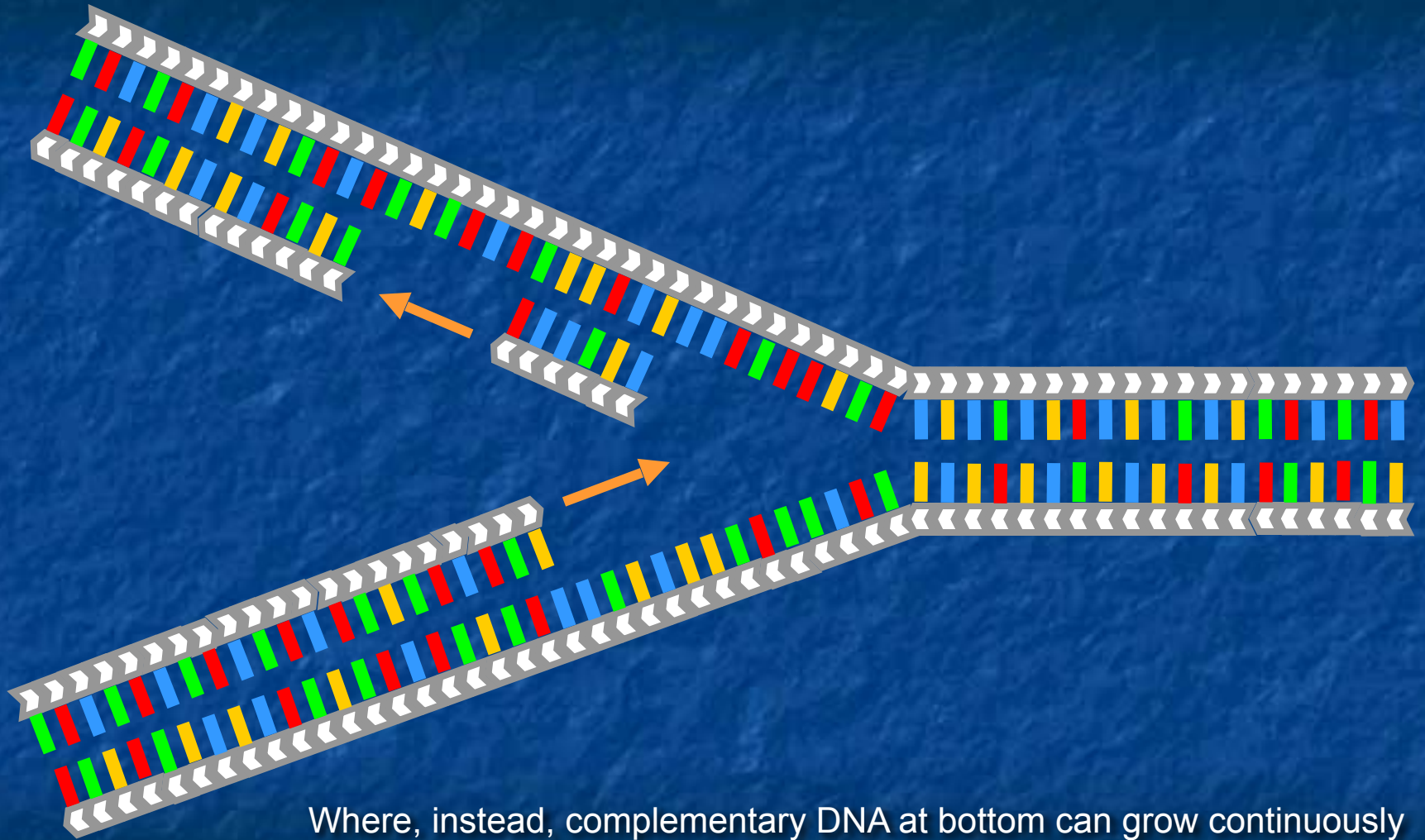
(from World of the Cell  
figure 15.8)

**BOTTOM:** No problem, new complementary strand steadily adds nucleotides to 3' end

**TOP:** Problem, want new complementary strand to grow on its 5' end (which nature won't do)

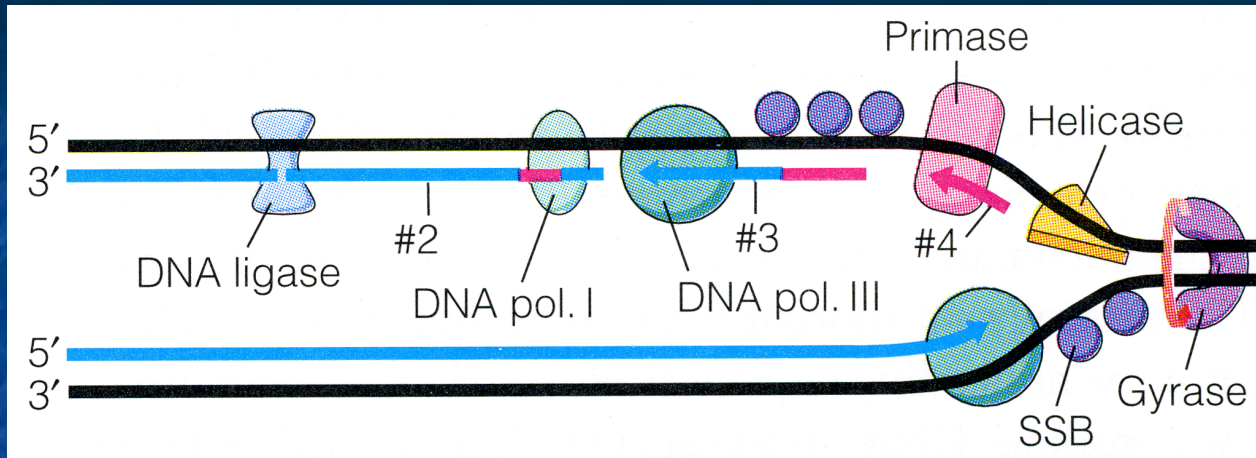


Complementary DNA at top, must grow as sequence of wrong-way “Okizaki” segments:



Where, instead, complementary DNA at bottom can grow continuously

## Yielding final more complex procedure:



(from World of the Cell  
figure 15.1)

**Gyrase** (Topoisomerase) to untwist incoming DNA

**Helicase** to split strands

**SSB** (single strand binding proteins) to stabilize separate strands

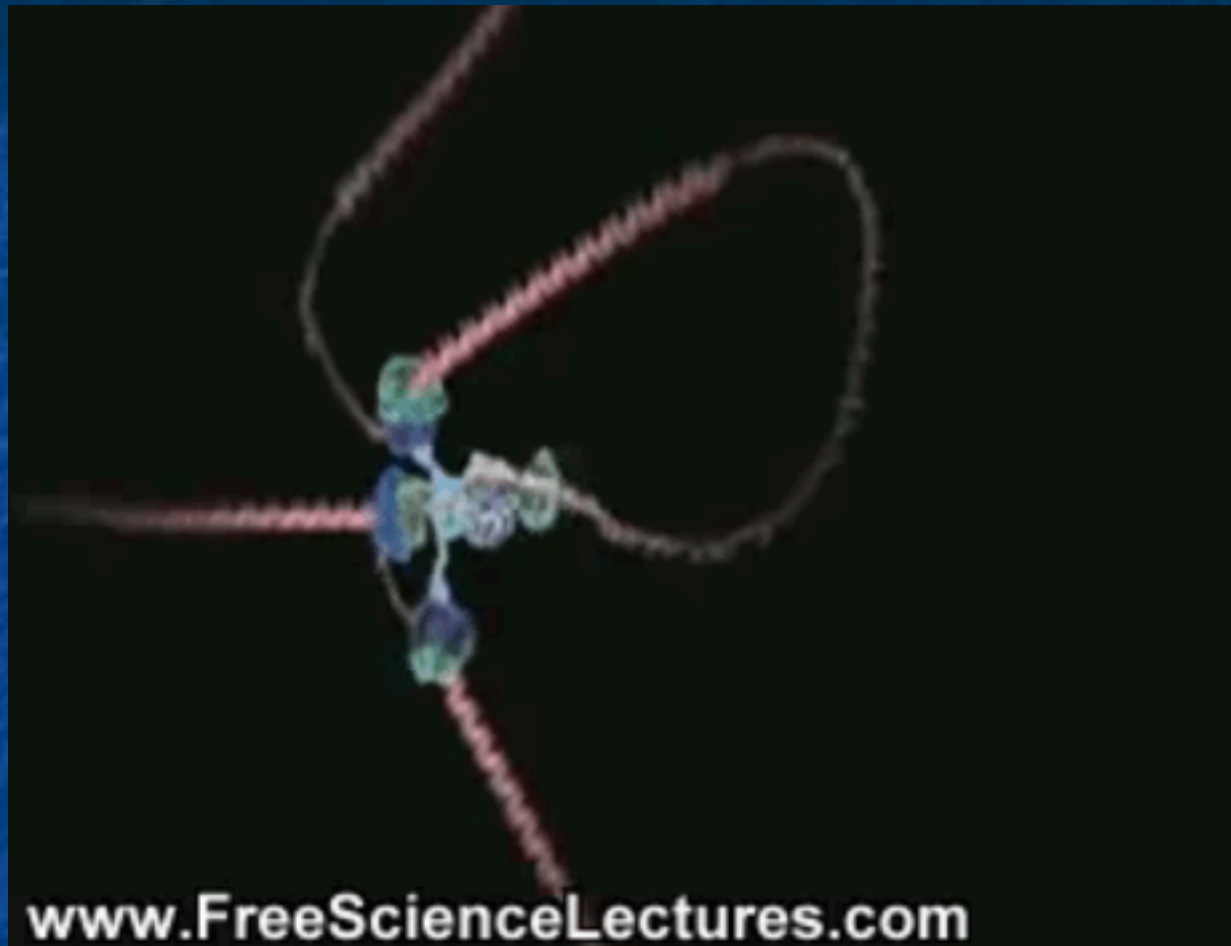
**Primase** to initiate growth of new strand segment (provides template / attachment point)

**DNA polymerase III** to add nucleotide units at growing end of new strands

**DNA polymerase I** to remove Primase where Okazaki segments grow together

**DNA Ligase** to then knit together seam between Okazaki segments

*Or as an animation:*



From: [www.youtube.com/watch?v=5UoKYGKxxMI&feature=related](http://www.youtube.com/watch?v=5UoKYGKxxMI&feature=related)

(or for cached version of 2<sup>nd</sup> part see supporting webpage: [DNA Self-Assembly - Supporting Materials – DNA packing and replication](#))

*A Hands-on Introduction to Nanoscience: [WeCanFigureThisOut.org/NANO/Nano\\_home.htm](http://WeCanFigureThisOut.org/NANO/Nano_home.htm)*

# DNA Self-Assembly Task #2: RNA "Transcription"

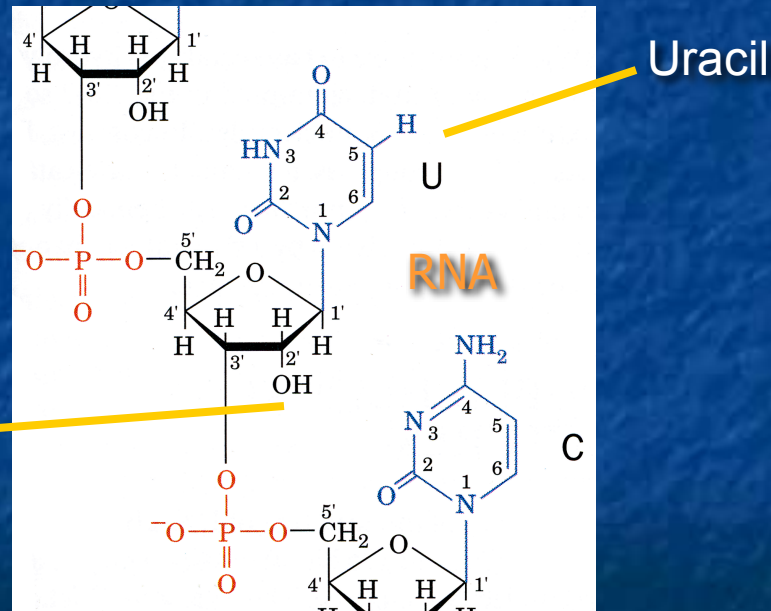
Here goal is to build complementary single strand of "Ribonucleic" acid

Difference between Deoxyribonucleic acid (DNA) and Ribonucleic (RNA)?

- 1) "Deoxy" - That is, in ribose of RNA the OH group is NOT removed
- 2) One of the four bases, Thymine, is replaced by Uracil (U)

Compare with DNA  
figure in slide 10!

OH group at ribose site 2



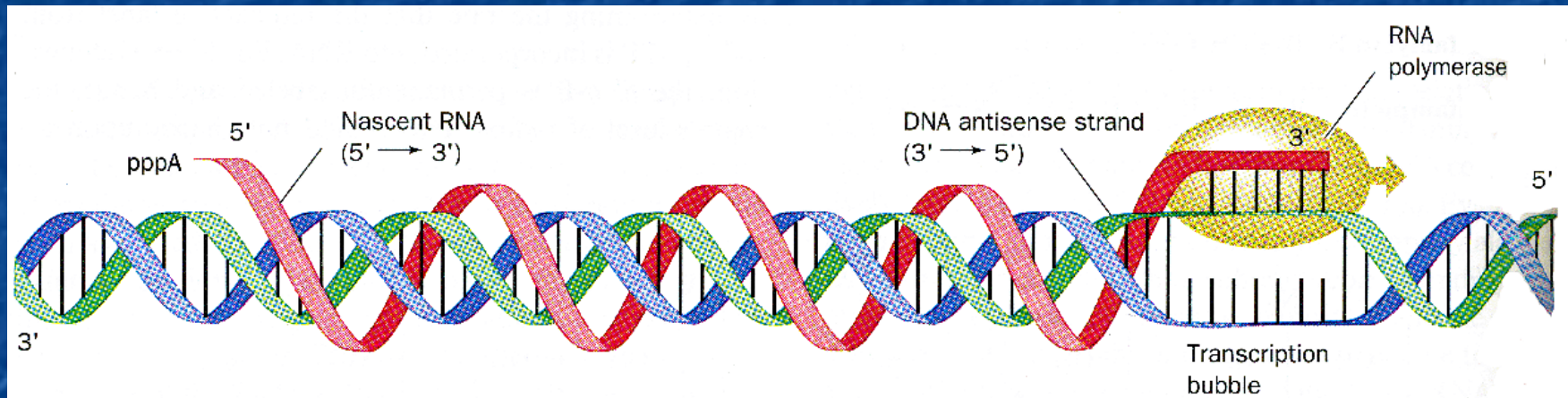
(modified from Voet & Voet figure 28.1)



# *Easier to synthesize single RNA strand:*

RNA polymerase just untwists DNA for about one turn

In that space, it assembles leading end of new RNA strand:



(from Voet & Voet figure 29.12)

Blue / Green ribbons = DNA being "expressed"

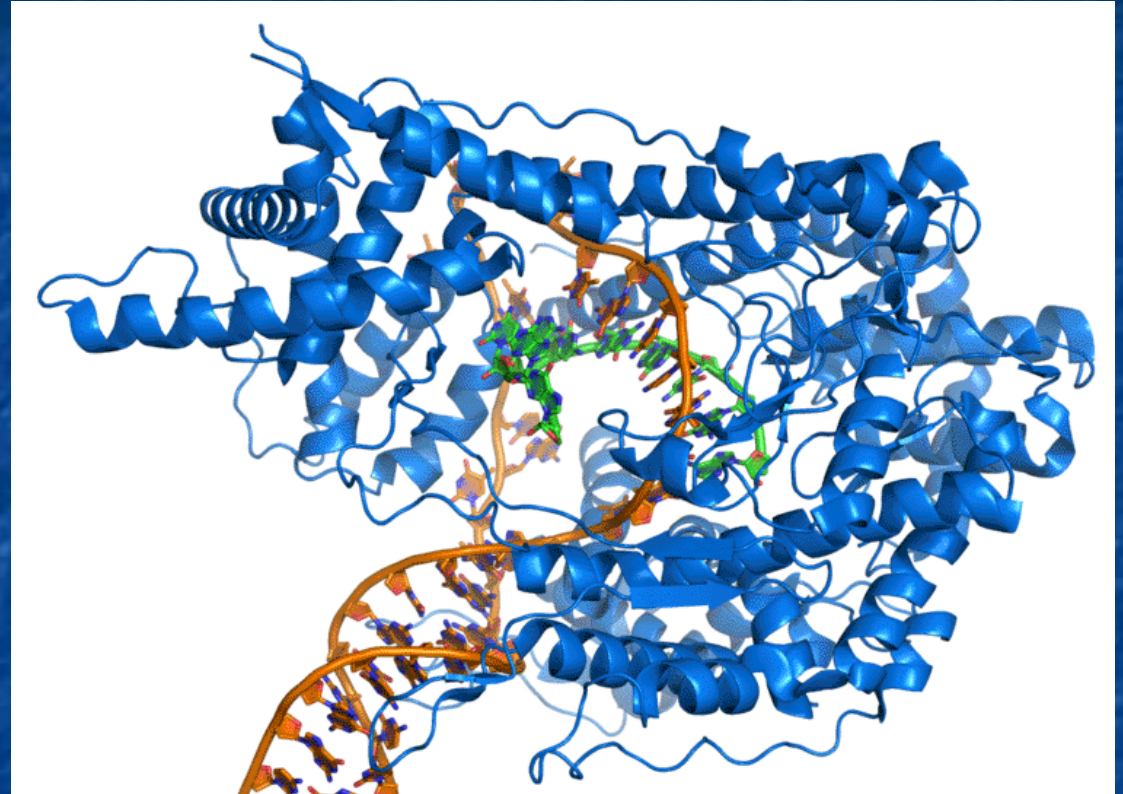
Red ribbon = RNA complementary to DNA's green strand

# Reality is not quite so neat:

Linked DNA strand (gold)  
entering from bottom

RNA polymerase (blue)  
at center unwinding DNA

Newly created single  
strand of RNA (green)



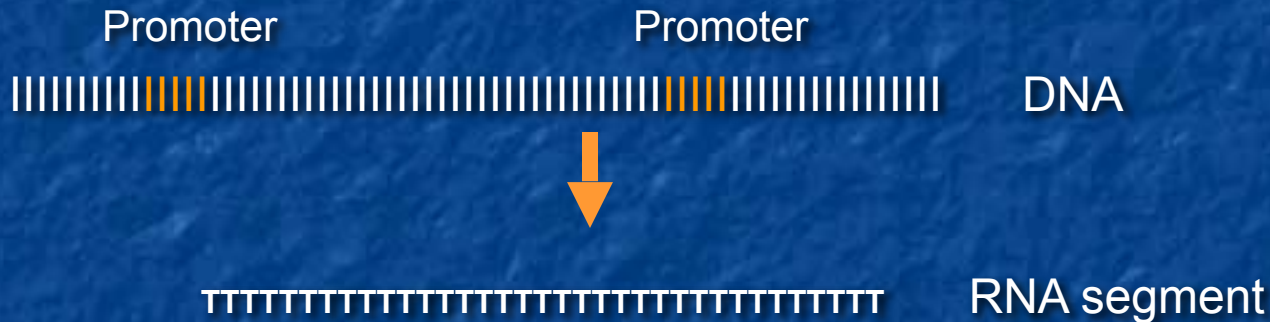
(From Wikipedia page on DNA)

Determination of RNA Polymerase's structure (from X-ray diffraction data) earned Roger Kornberg the 2006 Nobel Prize!

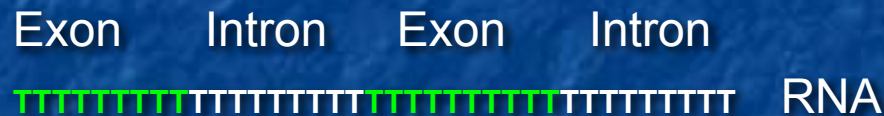
# Replicate *WHOLE* length of DNA in RNA?

No, would end up with unimaginably complex tangle

Instead, RNA “transcription” initiated at certain DNA “Promoter” base sequences



But the RNA segment contains further divisions based on its coding:



Only “Exon” segments have base sequences that will initiate next step:

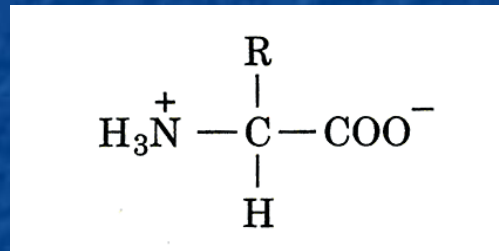
# DNA Self-Assembly Task #3: RNA Protein "Translation"

JOB of RNA Exon segments is to build proteins

**Proteins = basic cellular building material**

Proteins are based on chains of amino acids (or "peptide units")

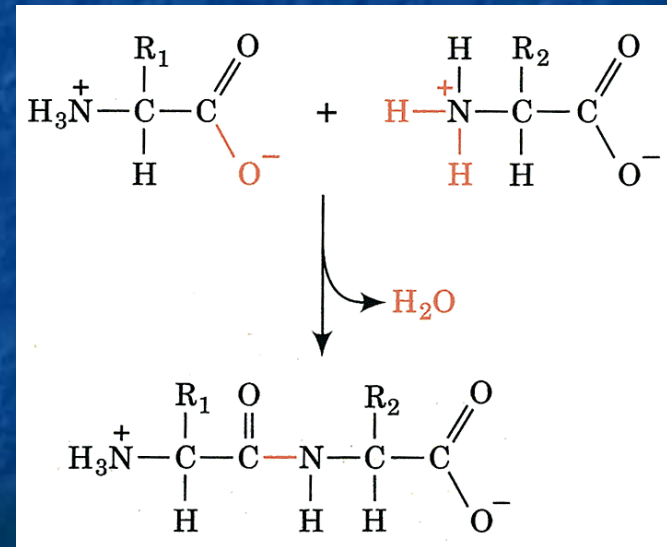
Single Amino Acid:



"R" = Chemist's shorthand for organic unit

That is: "Something more is attached here"

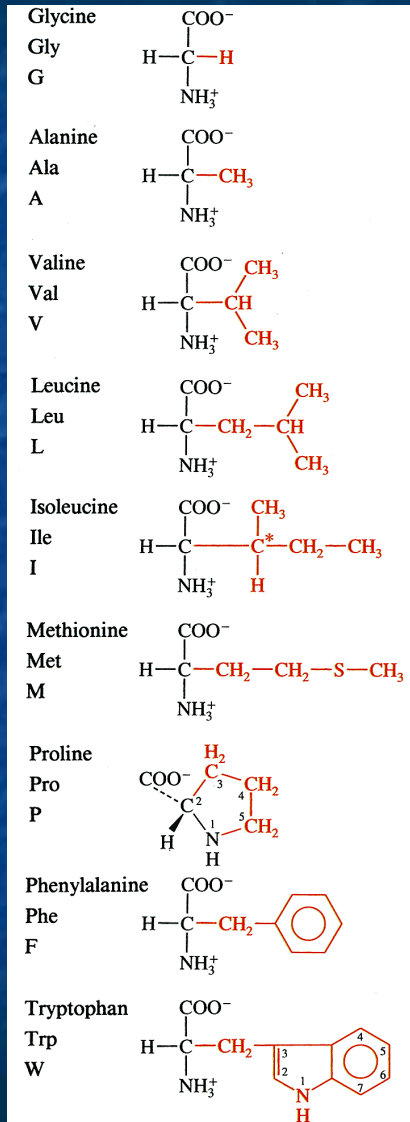
Linked together as polypeptide chains:



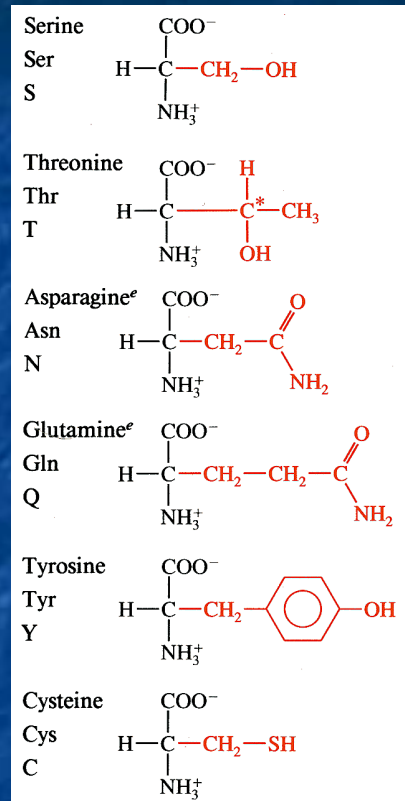
Variation of amine + carboxylic acid reactions of last lecture

# Twenty different amino acids based on choice of R group:

Non-charged, non-polar:

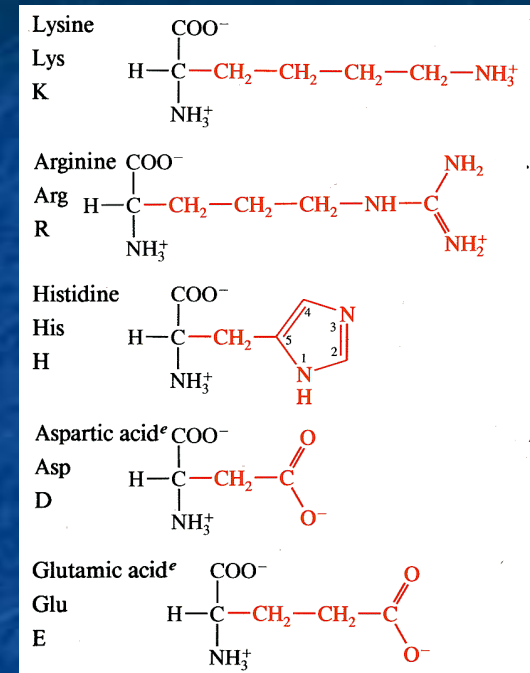


Non-charged, but polar:



Hydrophillic

Charged and polar:



Strongly Hydrophillic

Strongly Hydrophobic

Red parts = R's of previous page

(from Voet & Voet table 4.1)

## *Amino acid sequence along peptide chain can:*

Exploit differences in water attraction

Tune which sections of chain flee from water (coiling back upon themselves)

Tune which sections of chain seek water (trying to branch outward)

Pull distant segments together through attraction of polarized side "R" groups

Slightly negative oxygens or nitrogens on one group may be

attracted to slightly positive hydrogens on a distant group

Why is all of this important?

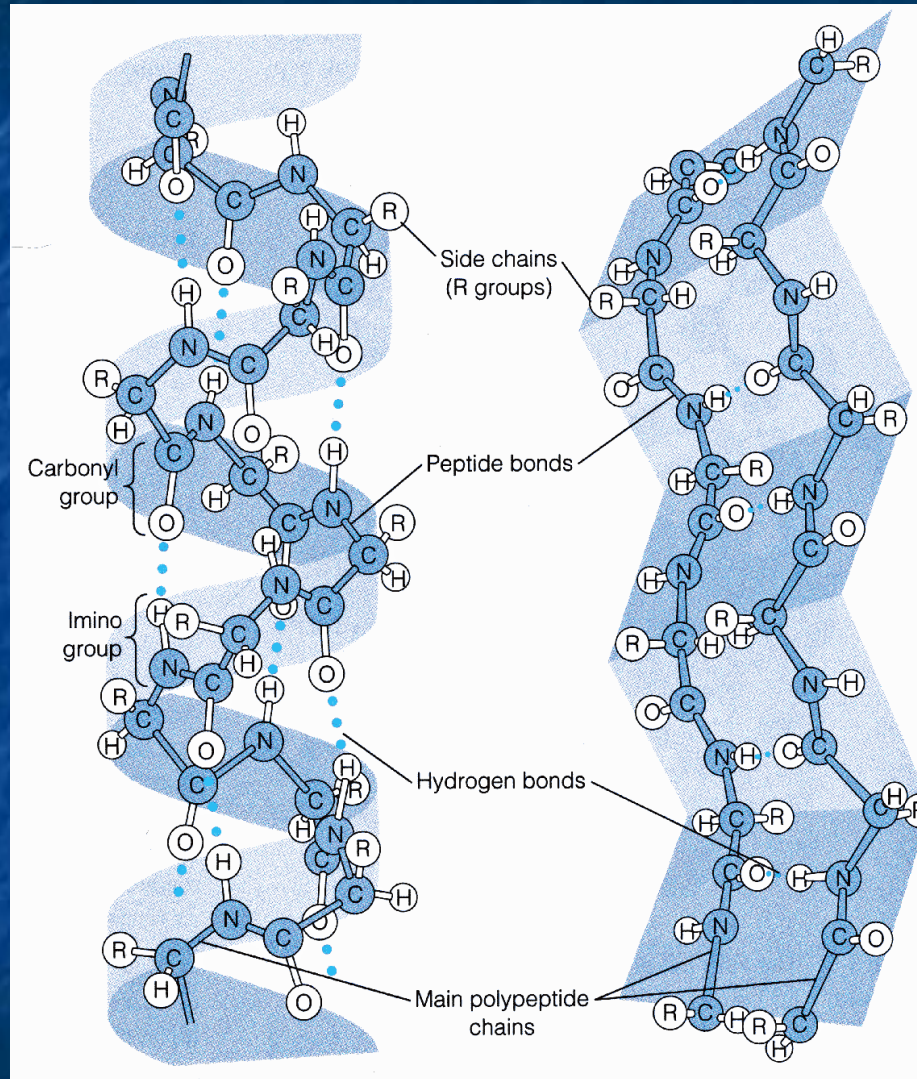
**It offers INCREDIBLE CONTROL OVER 3D STRUCTURE!**

# Two most basic polypeptide structures:

## "Alpha Helix"

O- on one peptide  
attracted to H+ on  
another

RESULT: Polypeptide  
spirals back upon itself



## "Beta Sheet"

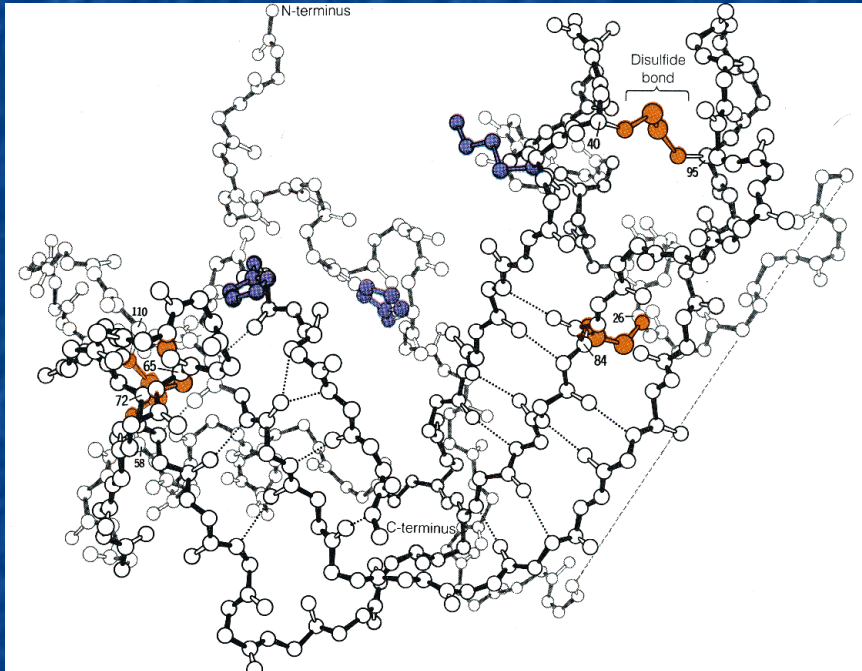
O- 's and H+ 's on  
adjacent polypeptides  
attracted to one another

RESULT: Parallel  
polypeptide chains  
linked together

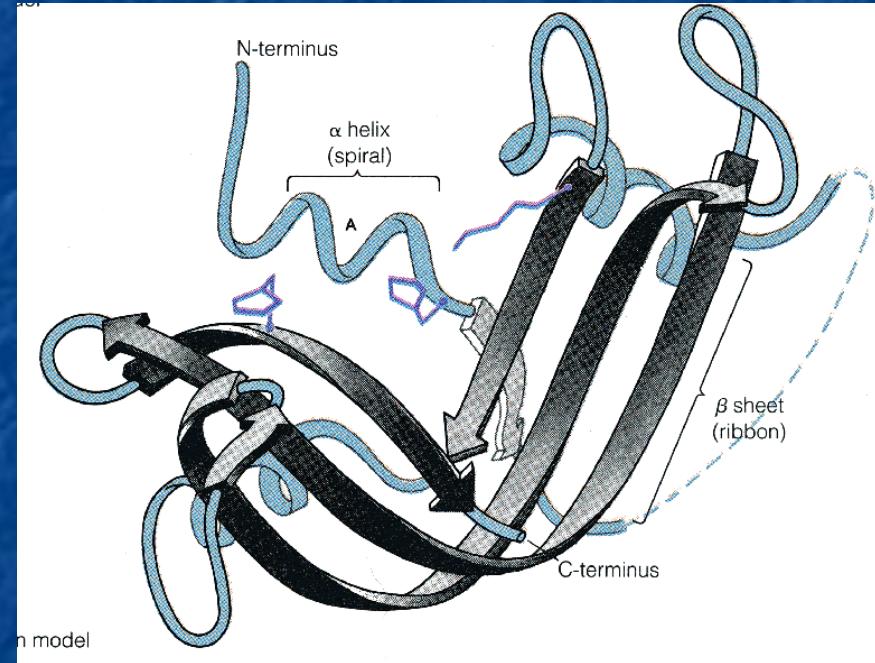
(from World of the Cell, figure 3.5)

# Which can be combined . . .

Ball and stick model:



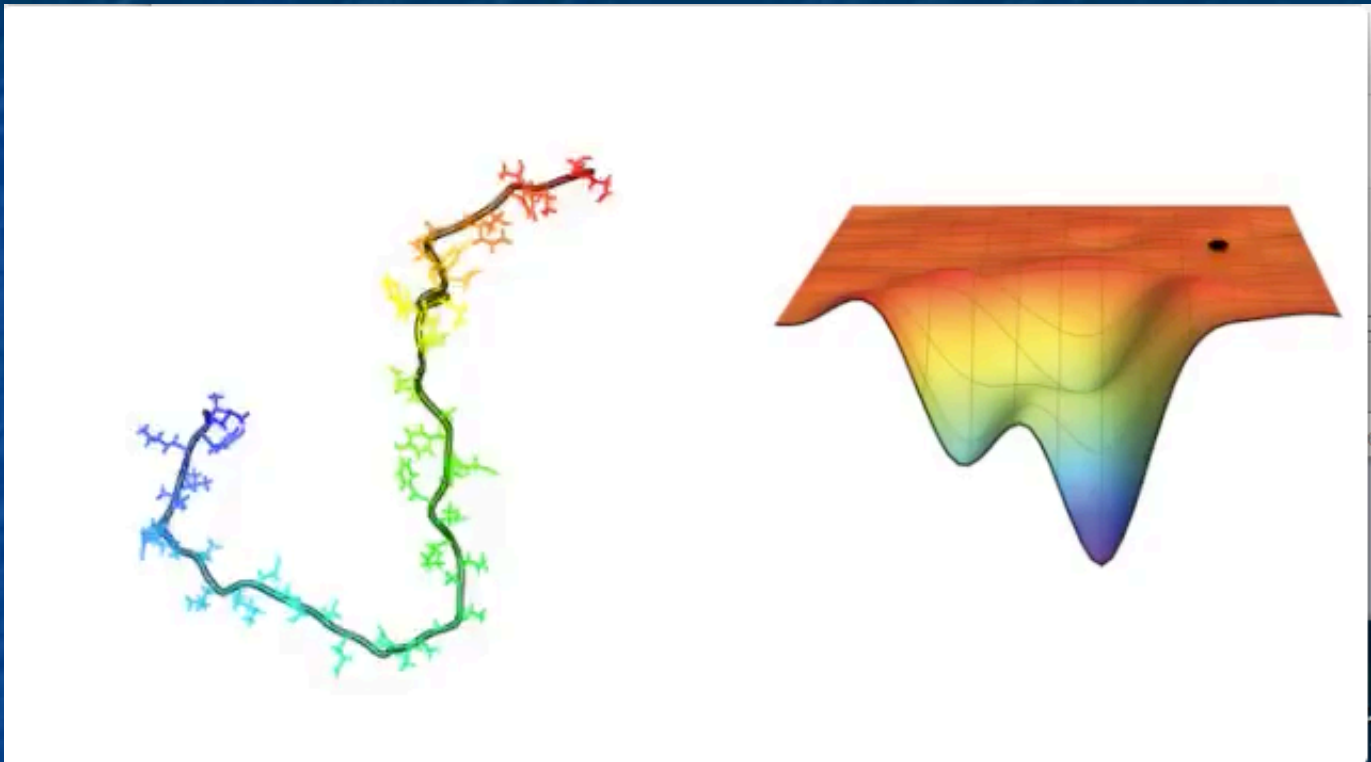
Spiral and ribbon model:



("Ribonuclease" structure from World of the Cell, figure 3.9)



*Which actually assemble via a trial and error energy minimization process*



Supporting webpage with embedded movie: [DNA Self-Assembly - Supporting Materials - Folding](#)

from: [www.youtube.com/watch?feature=player\\_embedded&v=YANAso8Jxrk](http://www.youtube.com/watch?feature=player_embedded&v=YANAso8Jxrk)  
(original source believed to be U. Stonybrook)

# *Final 3D groupings of polypeptide(s) $\equiv$ "Protein"*

**But how do RNA segments program Protein assembly?**

Exon RNA sub-segments separate to form either:

## **1) Messenger RNA (mRNA):**

Long chains with base sequence equaling recipe for the protein to be created

## **2) Transfer RNA (tRNA):**

Hunter units that go out into cellular soup to collect a required amino acid

Different type to gather each of the 20 different types amino acid

Each with 3 unpaired bases (a "codon") to dock at certain points on mRNA

## **3) Ribosomal RNA:**

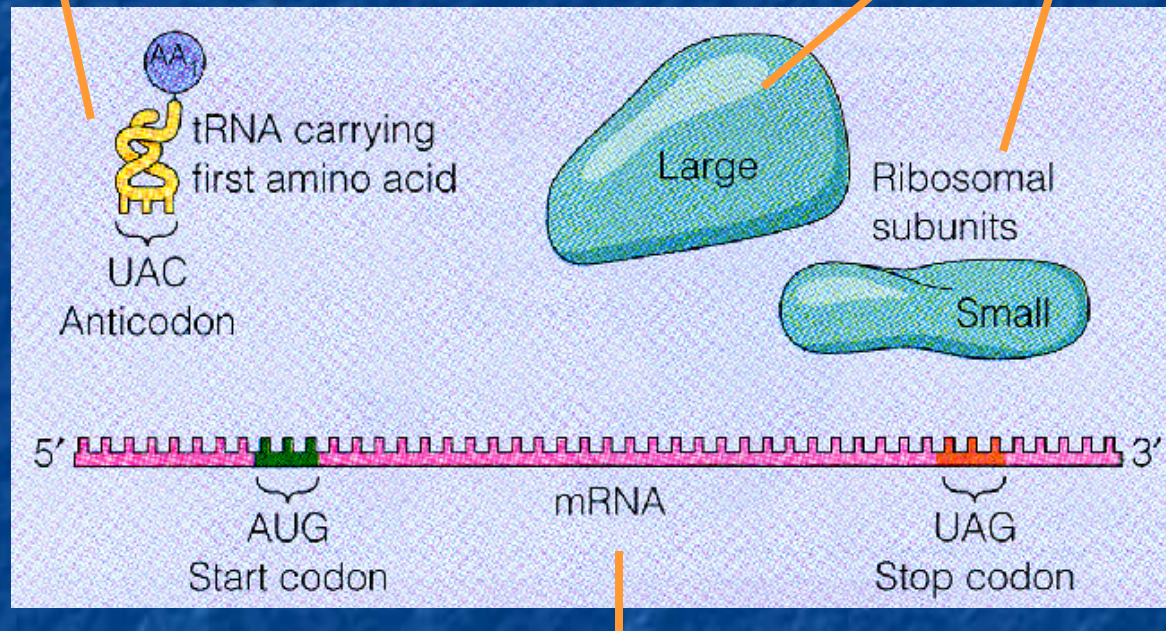
Forms ribosome templates to bend and coil protein chains into desired 3D form

# Somewhat easier in pictures:

Ribosomal RNA segments coil into shapes used as assembly templates

One of the 20 types of tRNA with its captured amino acid

This one has an exposed UAC base sequence (anticodon)



(From World of the Cell figure 18.7)

Long mRNA programming chain

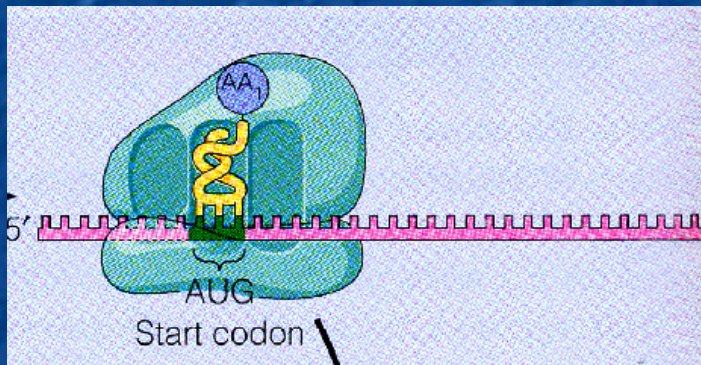
Including a three base sequence (codon) AUG complementary to the tRNA

# Ribosomal RNA units clamp onto mRNA chain:

Think of Ribosomal RNA unit as the thing you pull along a zipper

One side of the zipper is the **Messenger RNA (mRNA)** chain

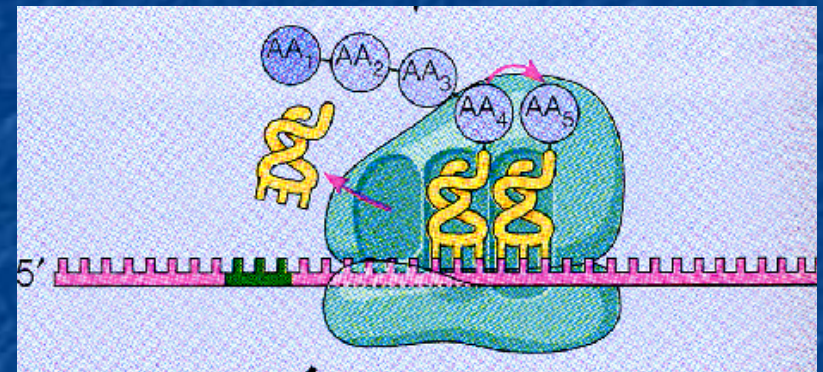
Other side is assembled BY the Ribosome using amino acids supplied by tRNAs:



1st tRNA unit with its amino acid

Snuggles into Ribosome

At its programmed place on mRNA



Add 2nd & 3rd tRNA's with amino acids

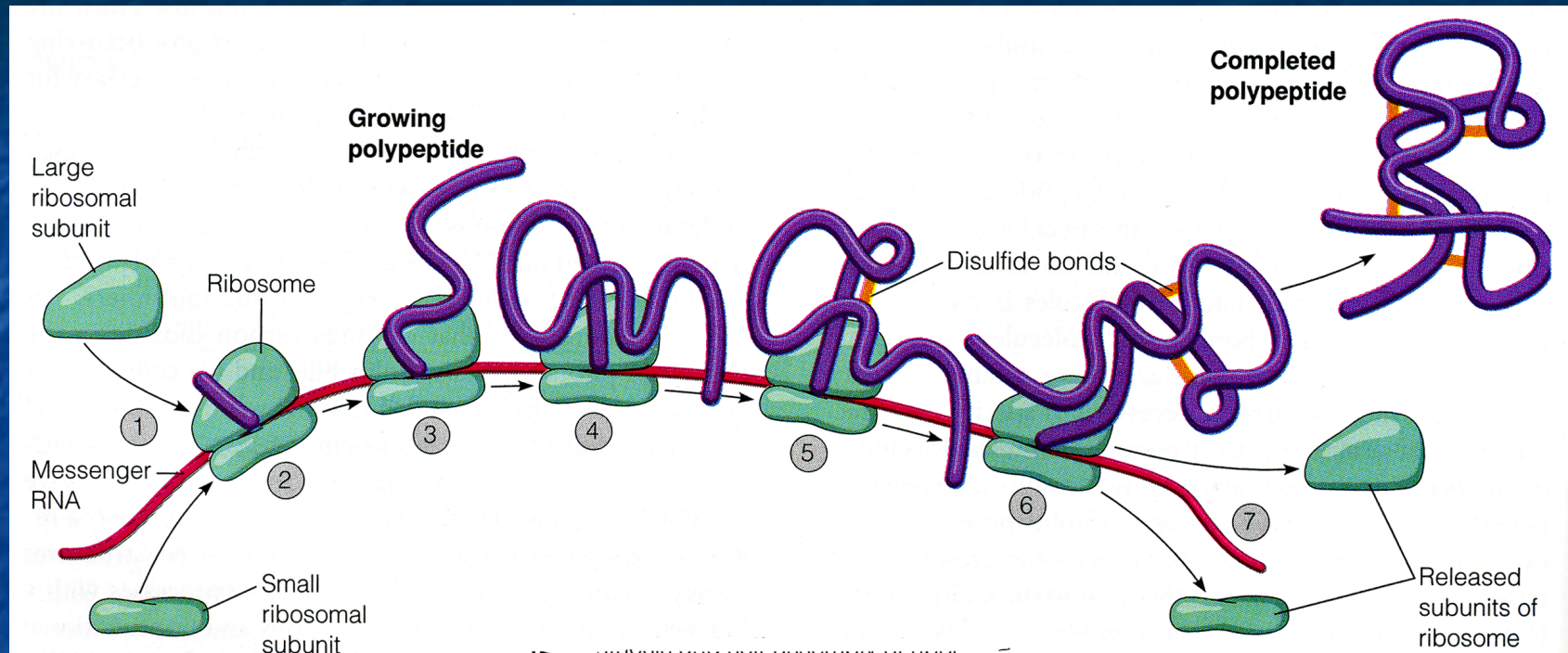
Amino acids start to link up

Ribosome moves right

Spits out 1st tRNA and

Growing chain of "polypeptide" (blue)

# Leading to creation of 3D structure of polypeptide(s) $\equiv$ "Protein"



Ribosome's SHAPE helps to program protein's (polypeptide's) SHAPE

WITHOUT ribosome, protein might have coiled differently => Different functionality

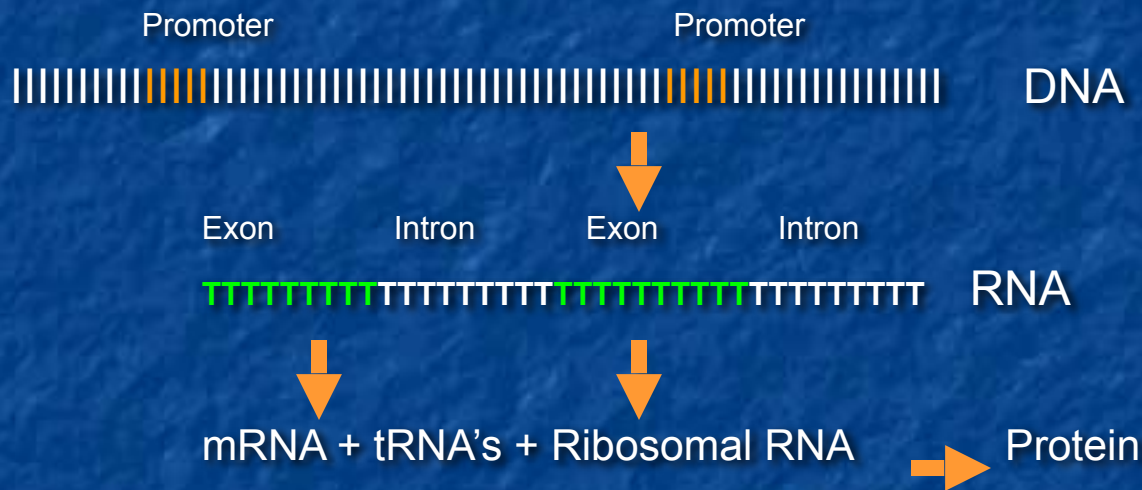
**Protein (polypeptide) Origami! => Building materials of cells**

## *But back to earlier question of “What is a Gene?”*

Functionally, it is the segment of DNA that programs “one characteristic”

Once thought that one protein might => one characteristic

Then “gene” would be length of DNA that programmed one protein:



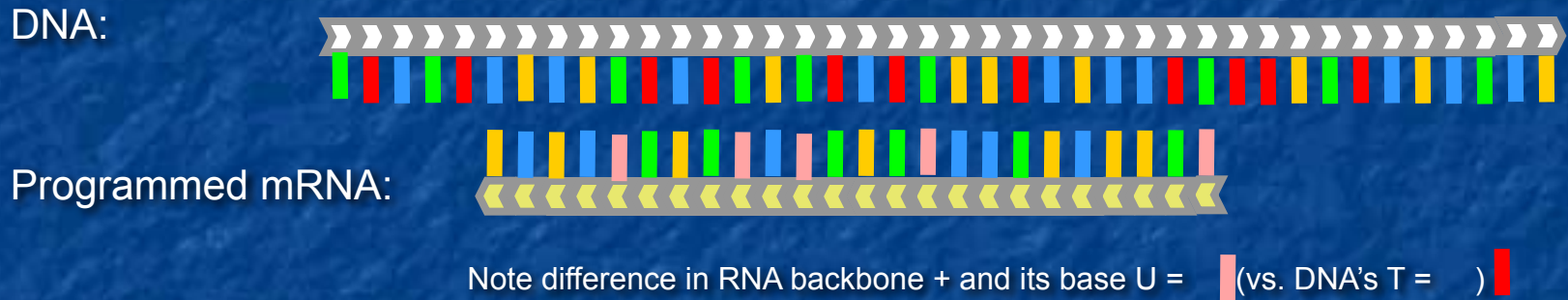
Now thought programming of single “characteristic” may require many proteins

So generalized “gene” involves longer DNA segments

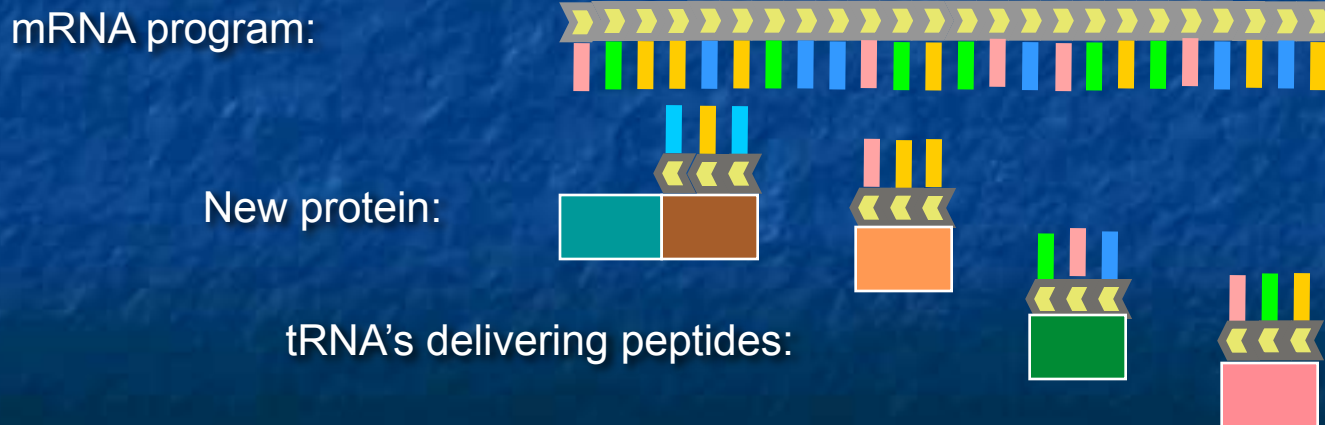
*But we are learning that it is even more complicated:*

**DNA can block the function of DNA!**

Normal process:



mRNA, per earlier slides, then wanders off and programs the synthesis of a protein



# What if complementary RNA subunit is programmed elsewhere on genome?

DNA:

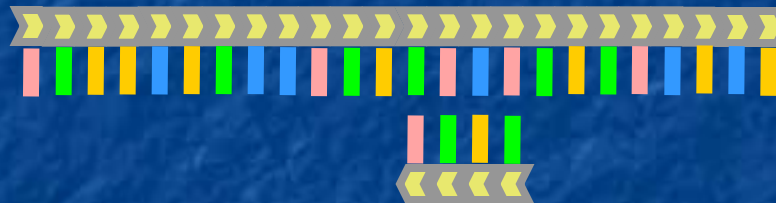


protein mRNA:

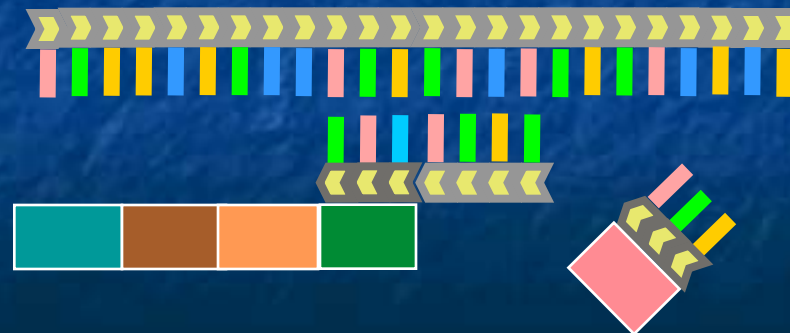


:Other RNA

RNA subunit from elsewhere can mate with first long segment of mRNA:



Stopping tRNA from delivering its programmed peptide to that site, halting protein's synthesis





## *This “microRNA” = another form of gene*

First DNA sequence programmed protein => inherited characteristic = “gene”

Second shorter microRNA DNA sequence did not program a protein

But it did stop the production of that first protein

So its presence did CHANGE an inherited characteristic

Thus, by definition, the DNA segment programming microRNA is also a “gene”

Some now suggest:

As many as 25% of proteins programmed by DNA, are blocked by microRNA

So revised view: Long genes (~10,000 base pairs) program proteins

Shorter genes (~ 100 base pairs) can block protein synthesis

(for more information see, for instance, Smithsonian Magazine article, July 2009)

## *This trick is now being investigated in Nano Medicine*

In the above, our **OWN** DNA is programming this "anti-gene"

But we can also easily (and cheaply) **synthesize** such interfering RNA segments:

By splicing required DNA sequence into bacterial DNA ("**recombinant DNA**")

Then letting bacteria reproduce, harvesting and purifying the results

To produce "**small interfering RNA**" (**siRNA**) segments

Which can then be used as a drug to "silence" (turn off) a **SPECIFIC** DNA segment!

Preventing the corresponding protein from being synthesized:

Which could kill or inhibit a cancer cell that depends on that protein

Or suppress a disease induced by over-production of that protein

*But genetic programming has GOT to be even MORE complicated:*

1) ALL cells contain identical (complete) DNA “program” yet cells of our body are different!

Different PARTS OF THE GENOME MUST BE TURNED OFF in different cells

2) Also observed that disease vulnerability can switch back and forth over generations!!

Swedish laplanders are very vulnerable to crop failure: Males experiencing famine in their late teens had grandchildren with decreased longevity due to diabetes

If grandchildren did not experience famine, change is reversed in their offspring

Timing (famine in late male adolescence) suggests alteration of sperm’s genetic program

And because change is SUDDEN (on evolutionary timescale) and REVERSIBLE, this implies:

Single gene must be capable of storing ALTERNATE PROGRAMS and  
ALTERNATE PROGRAMS must be selectable by environmental factors

To learn more about the Swedish study, see this excellent Time Magazine article ([link](#))

## *Further:*

### 3) Twins who start out identical don't stay identical!!!

Again suggesting environmental selection between alternate genetic programs

Understanding of multiple selectable genetic programs = emerging field of

## **"EPIGENETICS"**

To learn more, see Time Magazine article (previous page), this New York Times story ([link](#))

and PBS Nova / Nova Science Now shows, including:

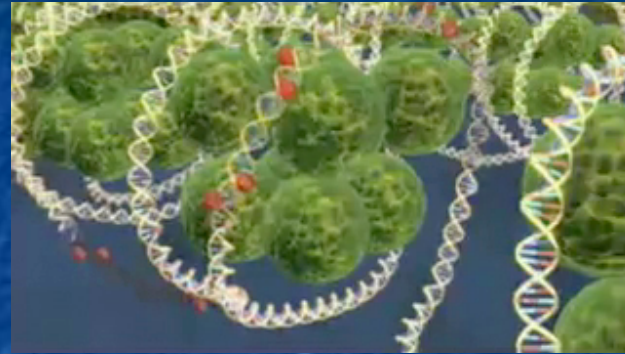


Supporting webpage with embedded Nova clip: [DNA Self-Assembly - Supporting Materials - Nova](#)

*A Hands-on Introduction to Nanoscience: [WeCanFigureThisOut.org/NANO/Nano\\_home.htm](http://WeCanFigureThisOut.org/NANO/Nano_home.htm)*

To “express” a gene, have got to undo that wonderful DNA packing:

One deactivation mechanism thus locks DNA to the “histone” beads it is wound around



From PBS's  
Nova Science Now 2008

Another adds chemicals to portions of DNA double helix impeding its separation

Which in rats can be initiated/reversed by maternal diet including consumption of vitamins

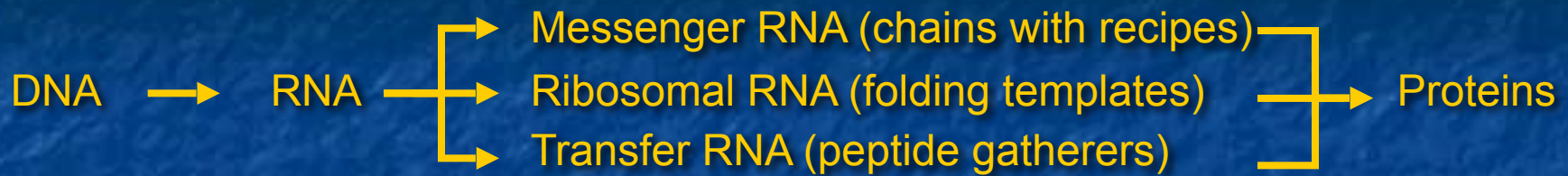
But, from what I am reading, this is likely just the tip of the iceberg:

And all that “junk DNA” may not be so useless/inert as once (arrogantly) thought!

With rapidly expanding field of epigenetics instead suggesting:

**DNA = Palette of recipes (slowly evolved) now selectable by environment**

*In conclusion, we've now learned about the "ultimate" in self-assembly:*



OR Simply: DNA → RNA → Proteins = 3D Molecular Origami

Makes manmade self-assembly processes of preceding lectures look sort of lame

Scientists now understand many of the processes involved

Also know how to instigate many of the steps

Suggests opportunity for developing non-cellular DNA-based self-assembly

(as will discuss further in lecture 11)

# *Credits / Acknowledgements*

Funding for this class was obtained from the National Science Foundation (under their Nanoscience Undergraduate Education program).

This set of notes was authored by John C. Bean who also created all figures not explicitly credited above.

Copyright John C. Bean

(However, permission is granted for use by individual instructors in non-profit academic institutions)