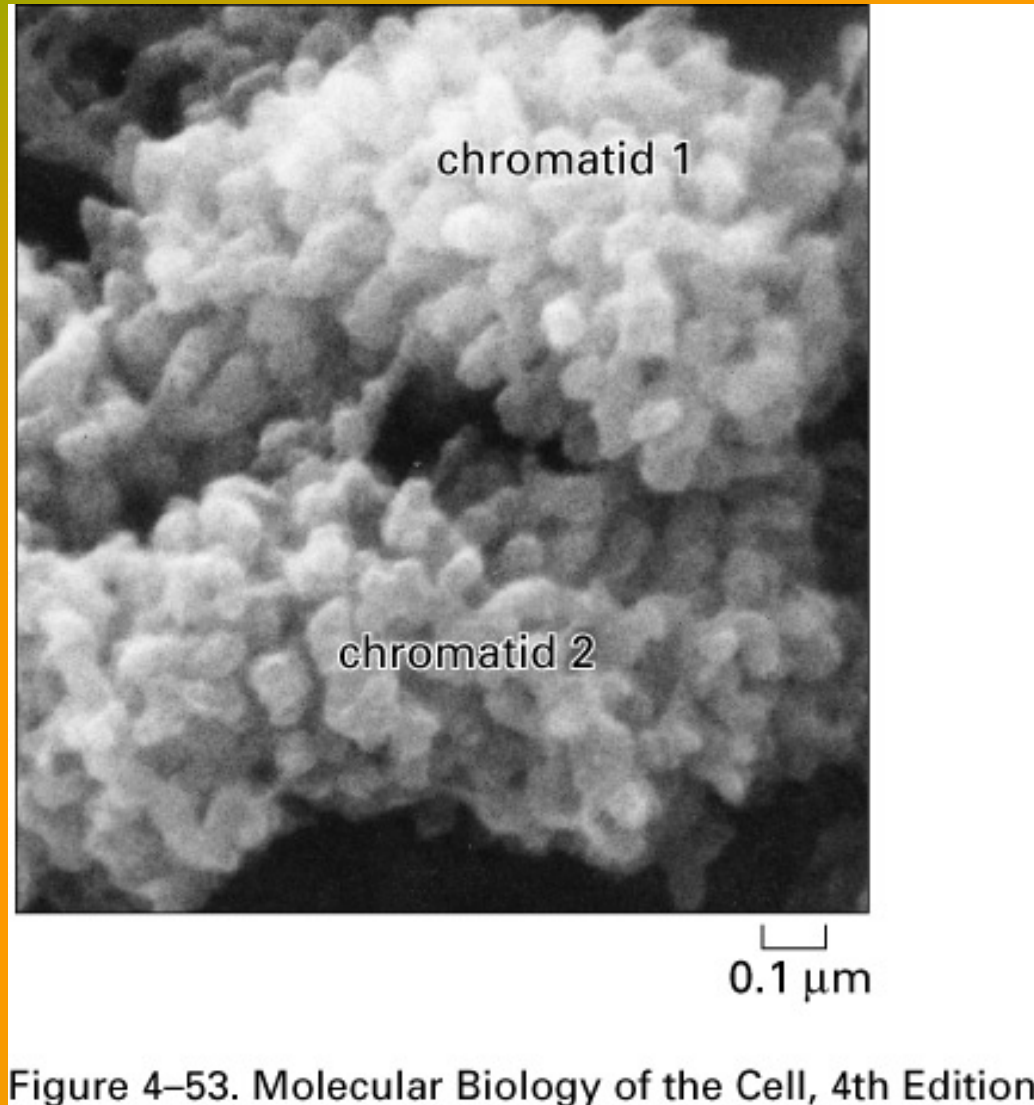


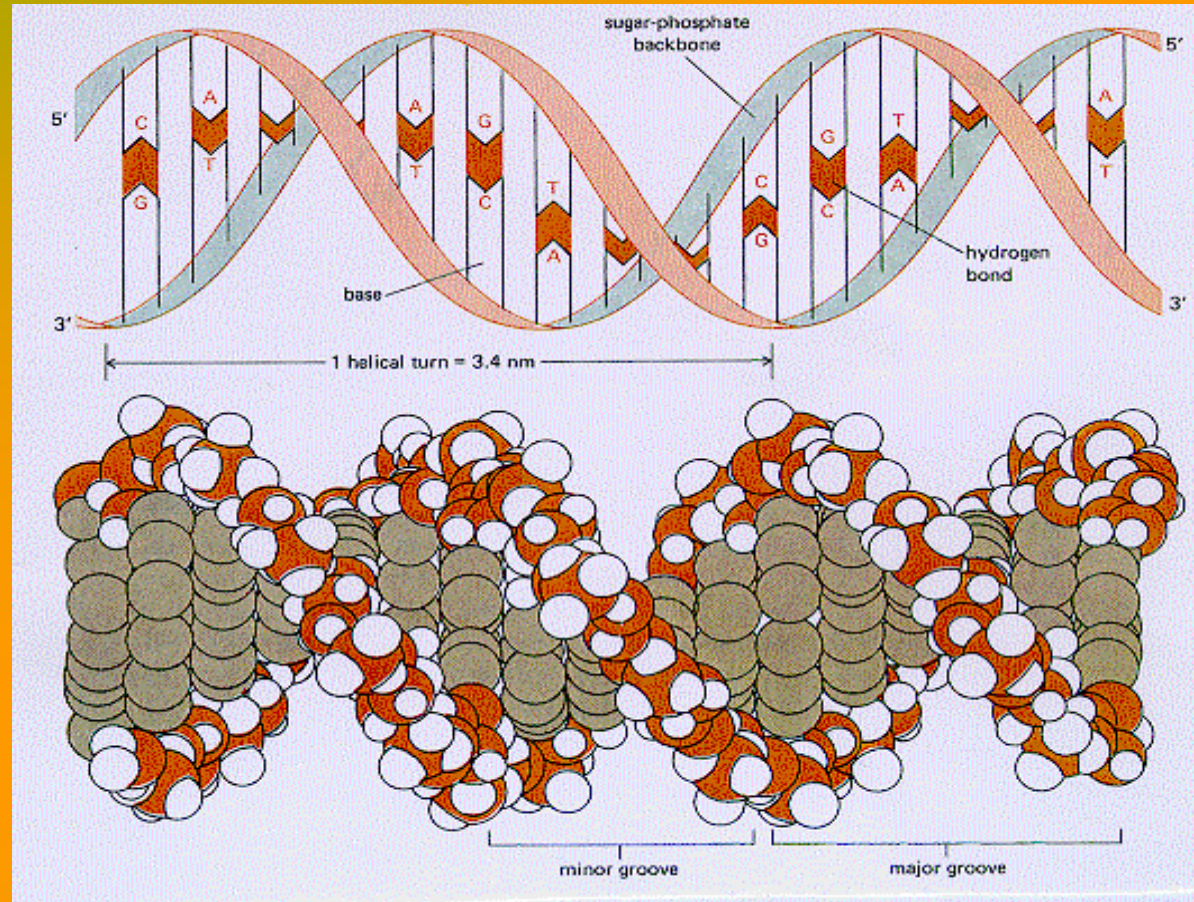
# Contents

- **Introduction: biology and medicine, two separated compartments**
- **What we need to know:**
  - boring basics in DNA/RNA structure and overview of particular aspects of molecular biology techniques
  - How DNA is organized and differs in every individual
- **Molecular diagnostics of cardiovascular diseases**
  - Mutations in Factor V
  - Mutations in Factor II
  - Mutations in MTHFR gene
- **Breast cancer and BRCA1 and 2 genes**
  - Breast cancer in the industrialized countries
  - Breast cancer genes
  - sequence in selected areas
  - p53 and breast cancer
- **Pharmacogenomics: finding the right drug for a patient**
  - ADR: an emerging problem
  - structure of cytochromes
  - Example 1: TPMT-enzyme and the metabolism of azathioprine
  - Example 2: Clozapine in the treatment of psychiatric diseases
  - CYP3A4 and the metabolism of anti-coagulant drugs

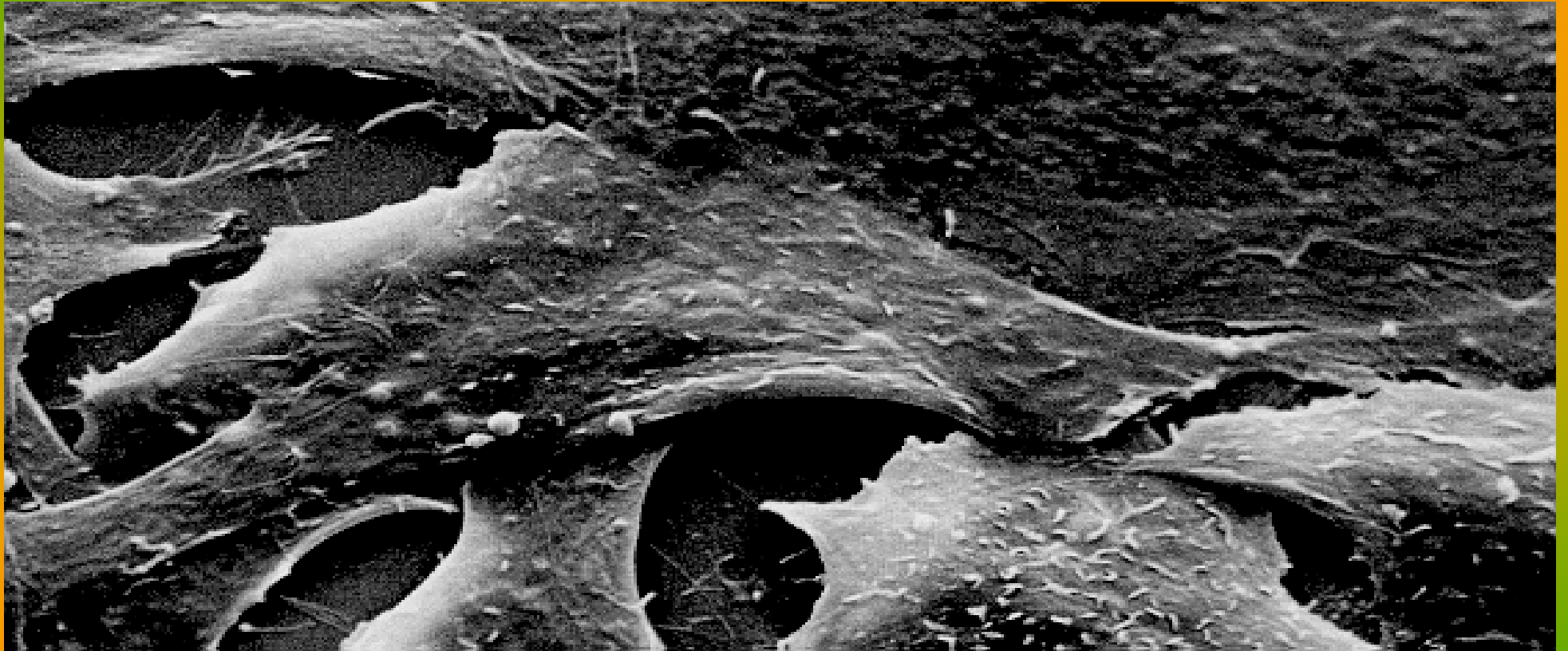
# What is about molecular biology?



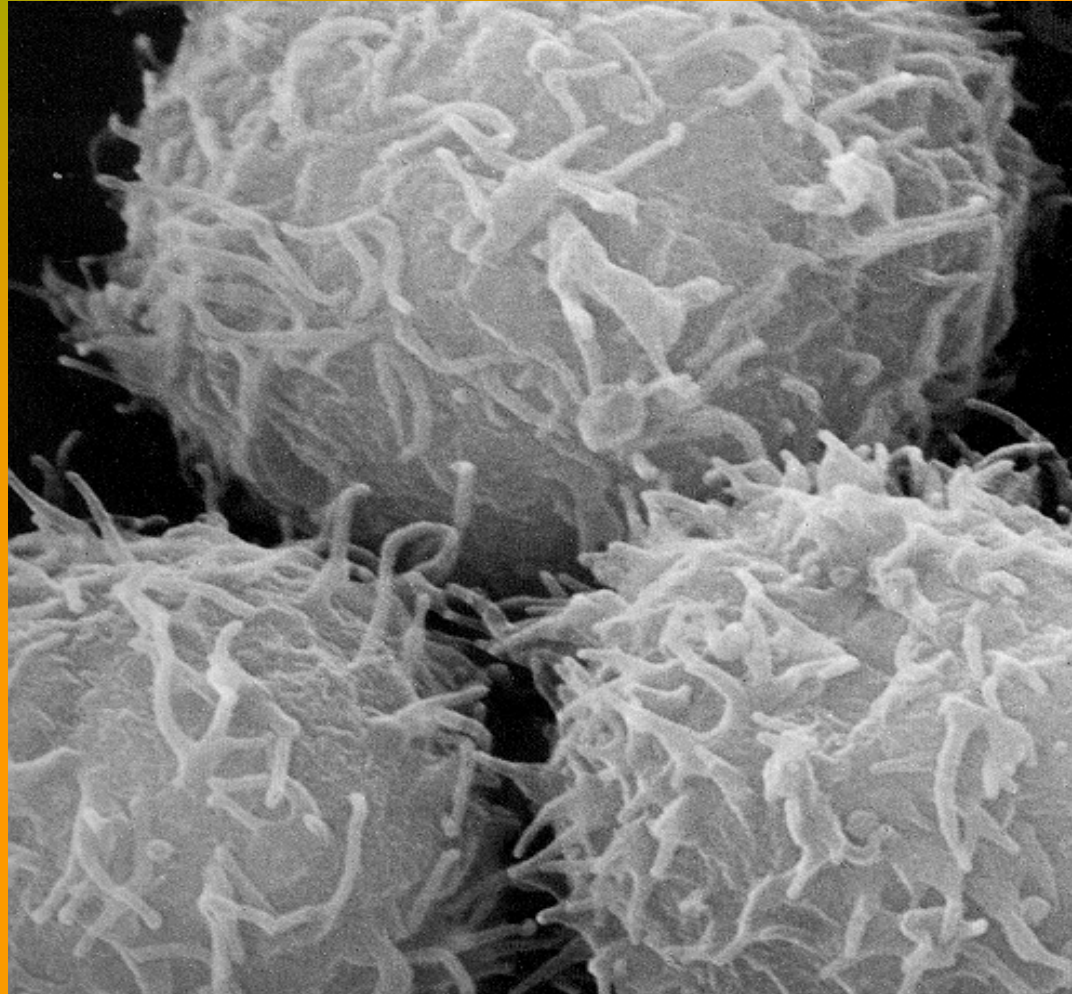
# Molecular biology is about DNA



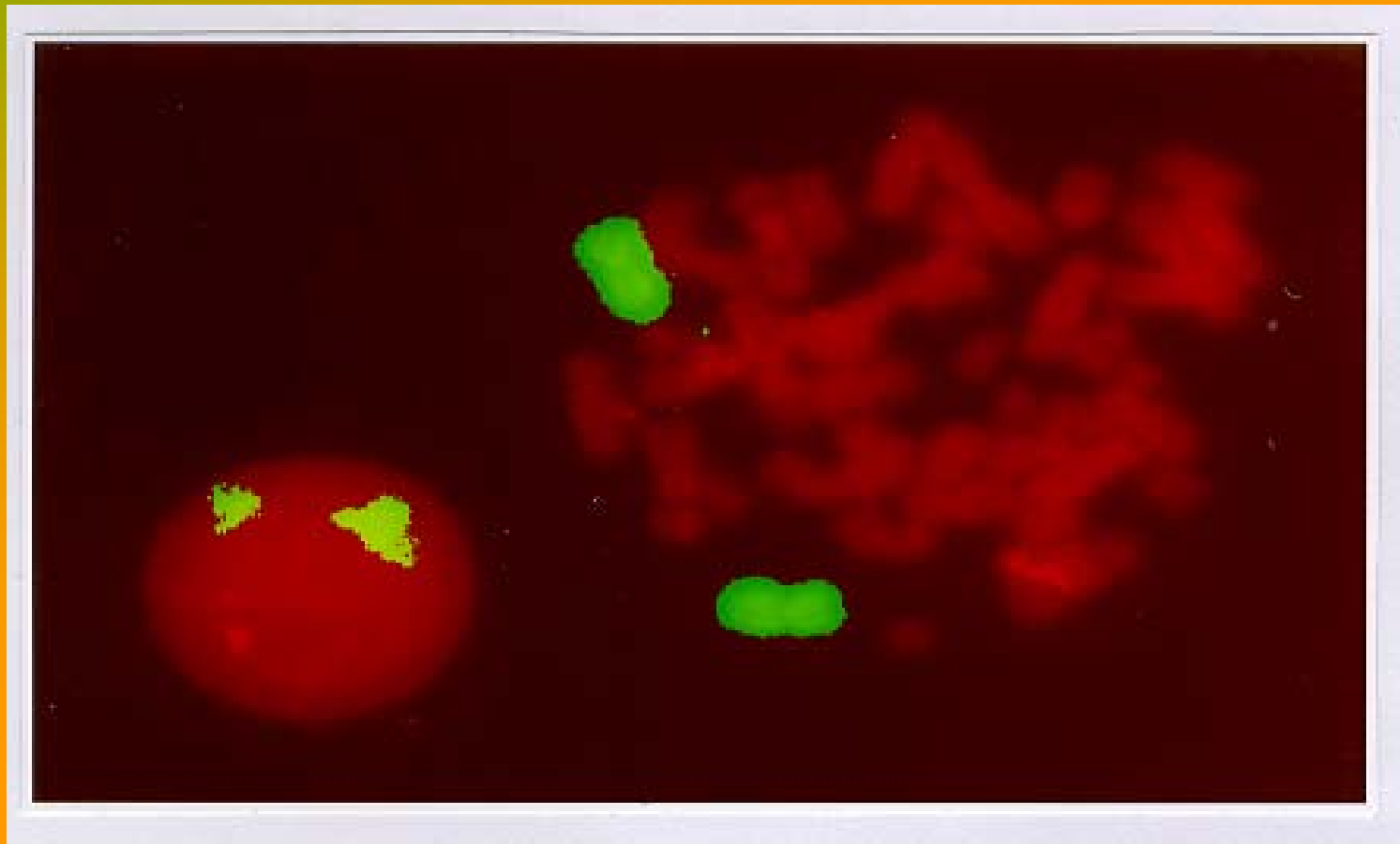
# DNA containers



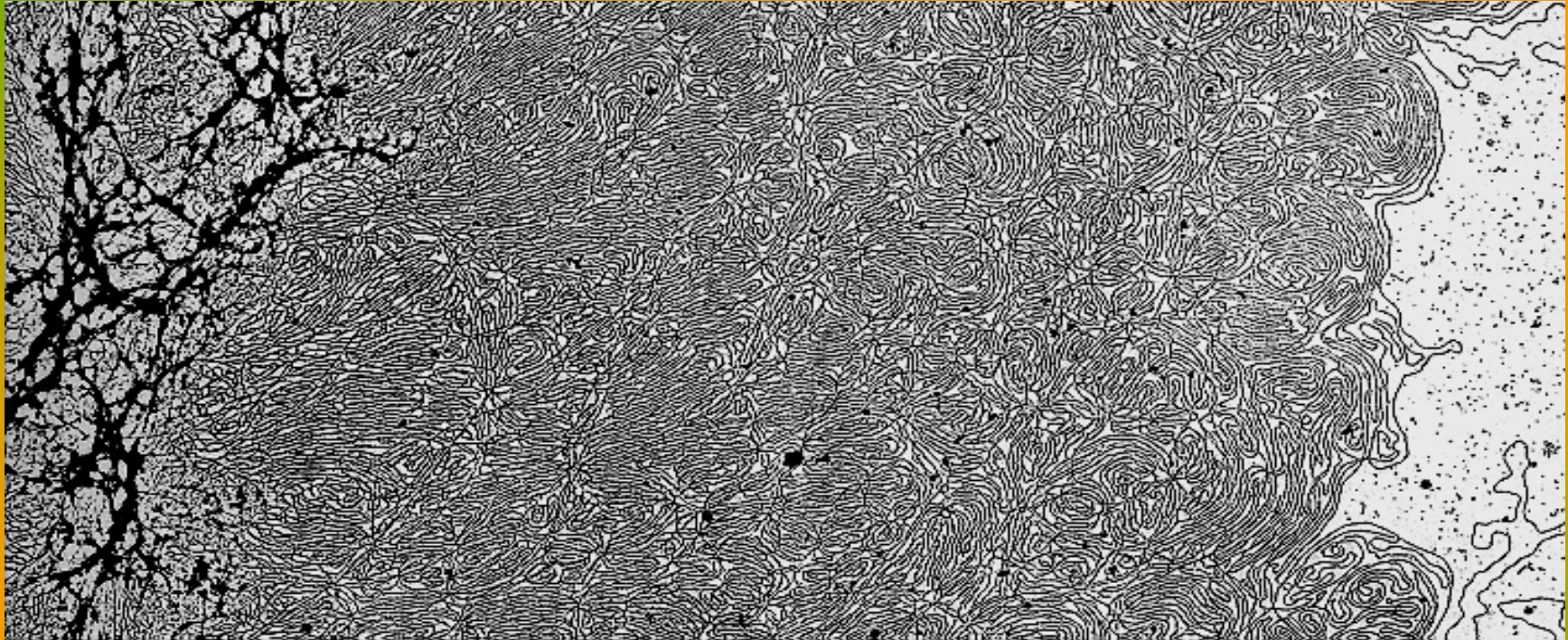
# And DNA containers



And inside cells...nuclei



# And inside nuclei...DNA



Inside every cell we can find more than 2 meters of DNA

# Chromosomes

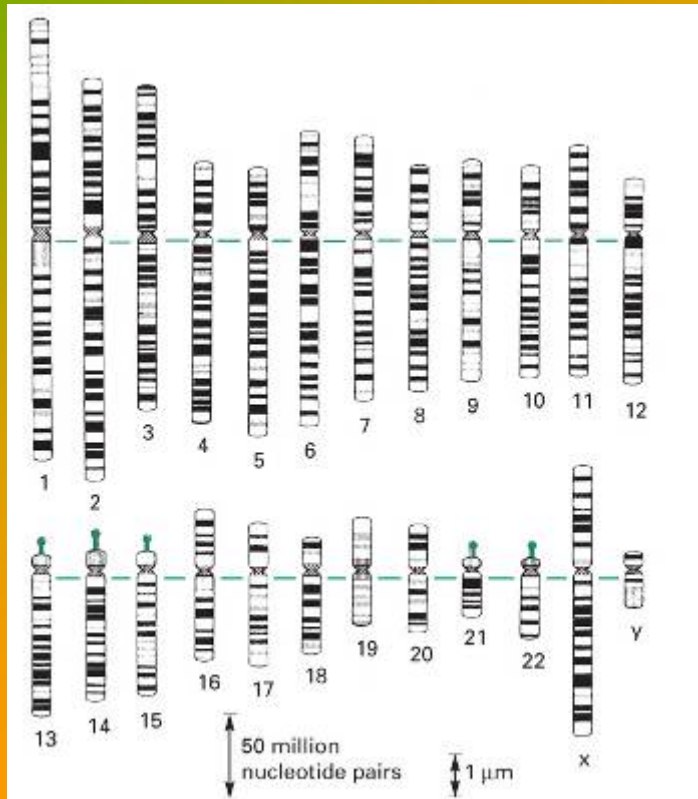


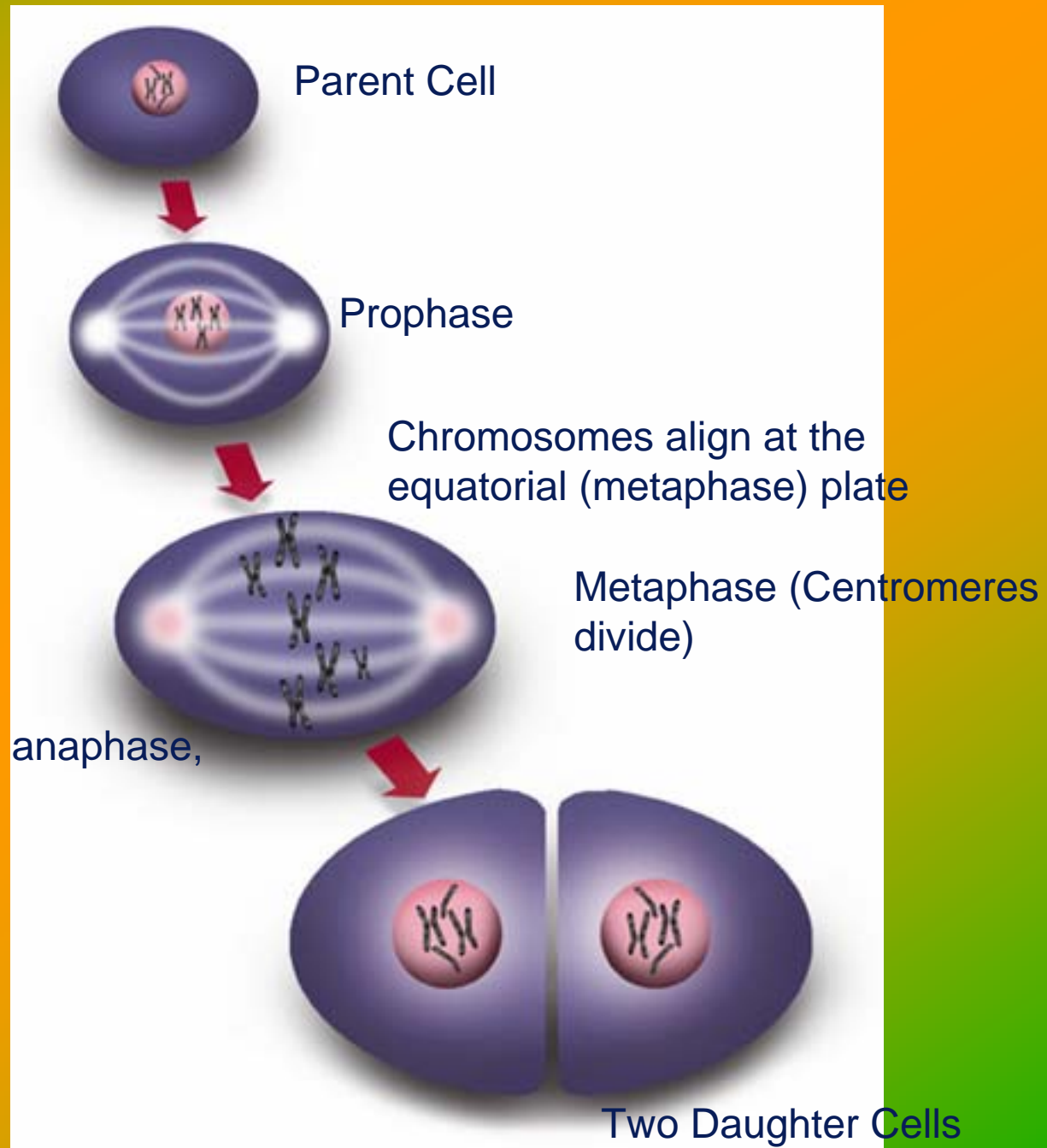
Figure 4-11. Molecular Biology of the Cell, 4th Edition.





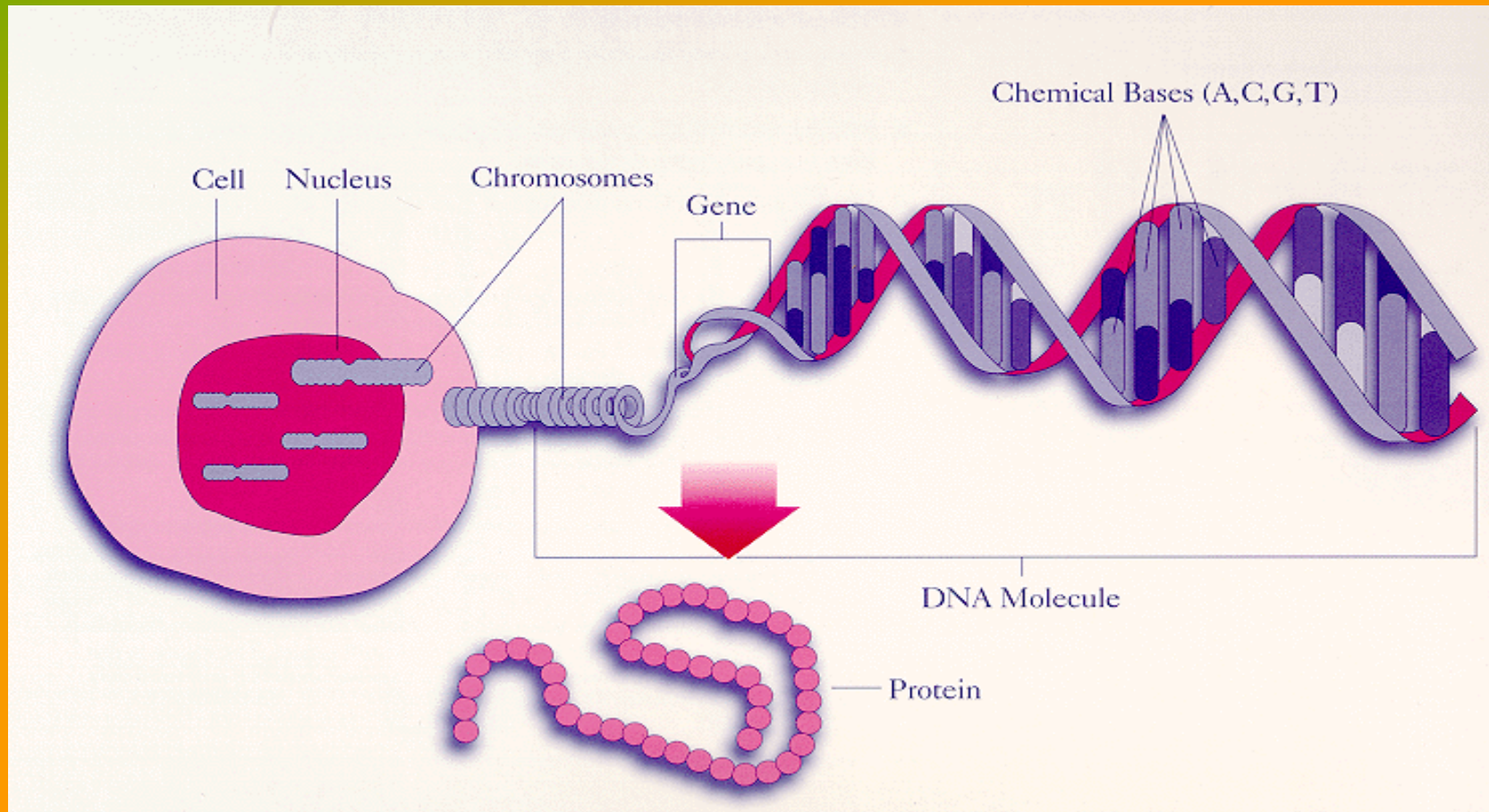
Every time a cell divides its DNA is duplicated

Sister chromatids separate during anaphase, becoming chromosomes



Two Daughter Cells

And here is how we explain it



# DNA structure

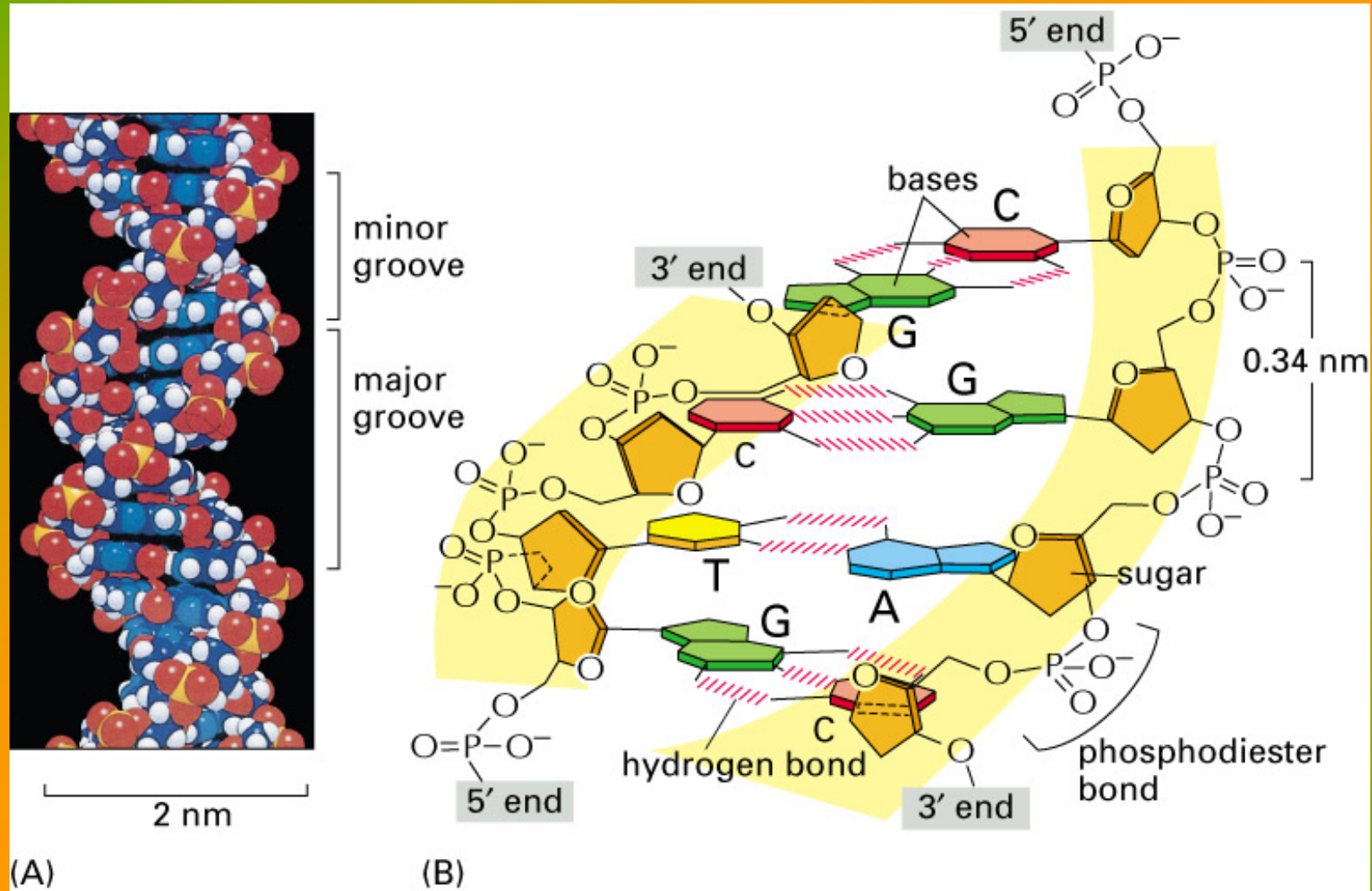
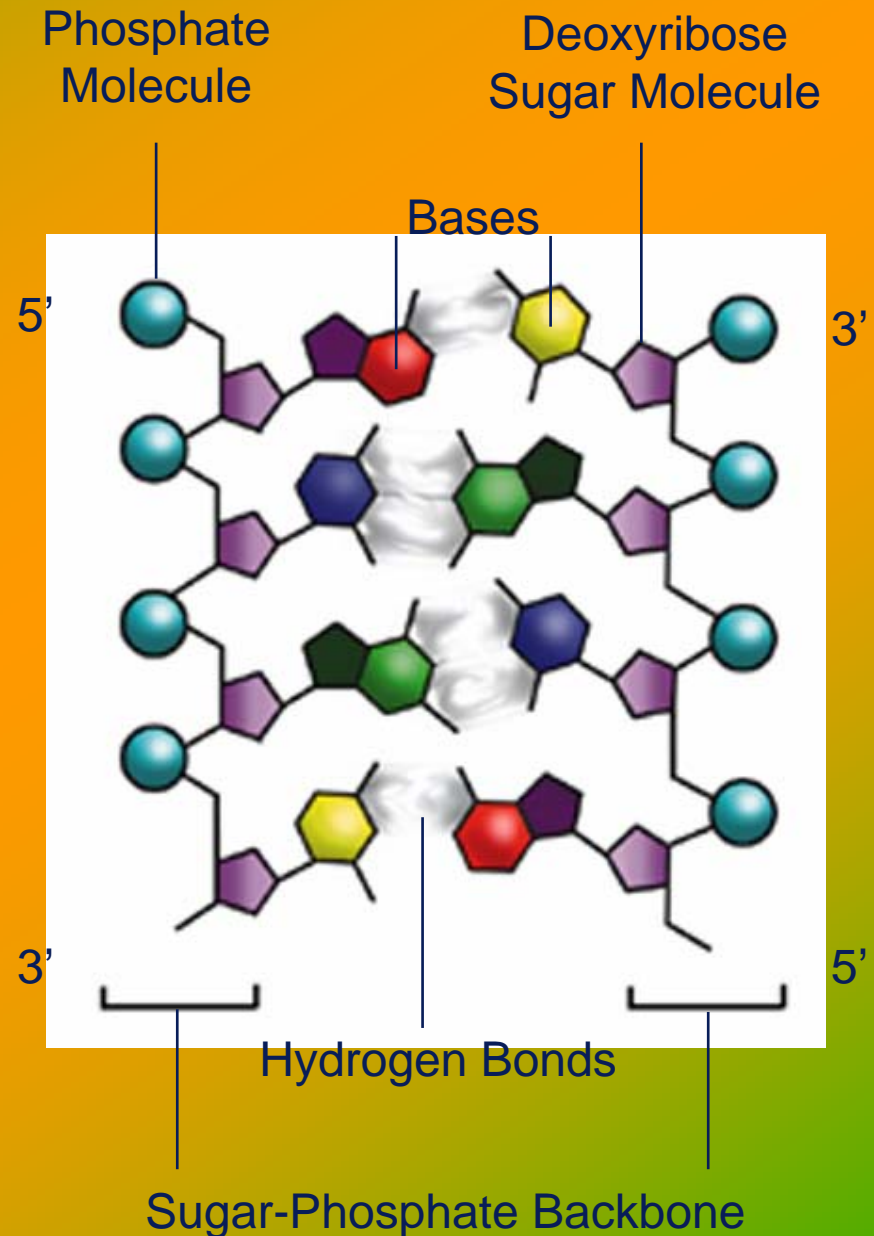


Figure 4-5. Molecular Biology of the Cell, 4th Edition.

# Bonding of Bases, Sugar and Phosphate Groups



# Hydrogen bonding

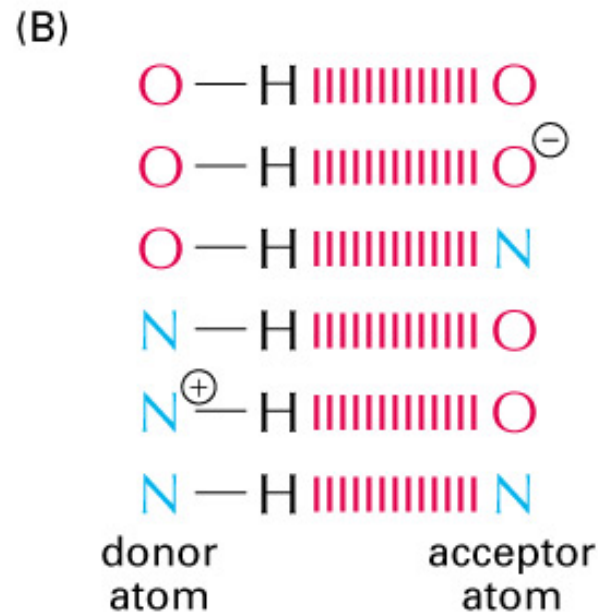
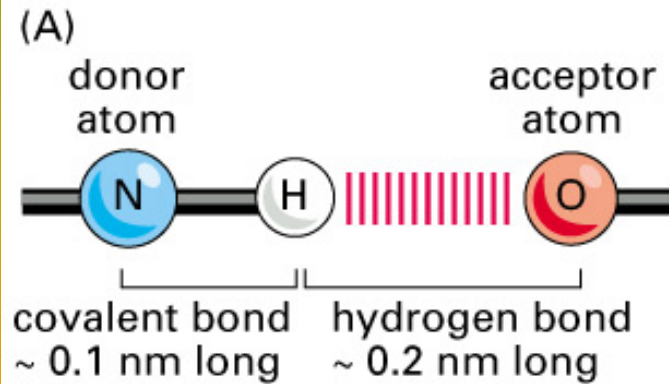


Figure 2-15. Molecular Biology of the Cell, 4th Edition.

# Hydrogen bonding in DNA

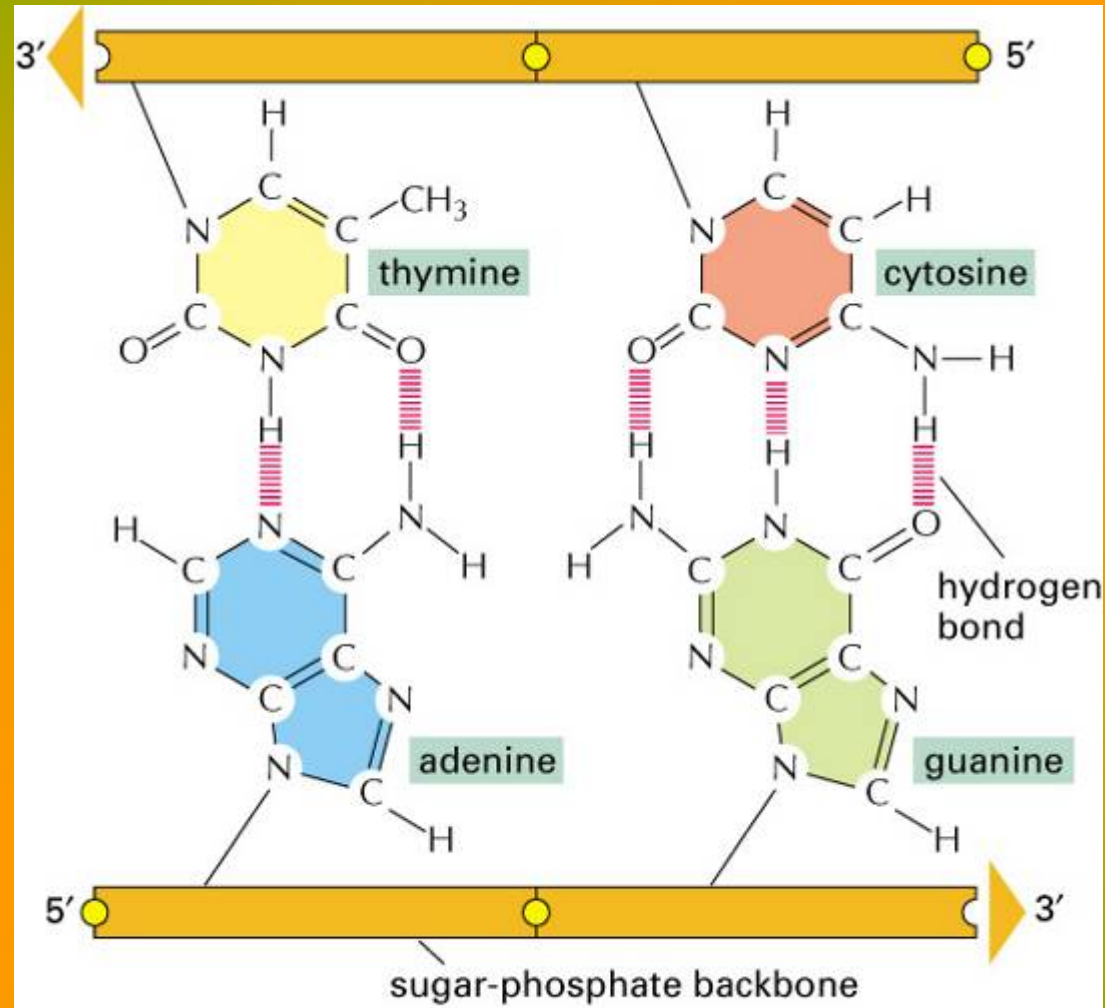
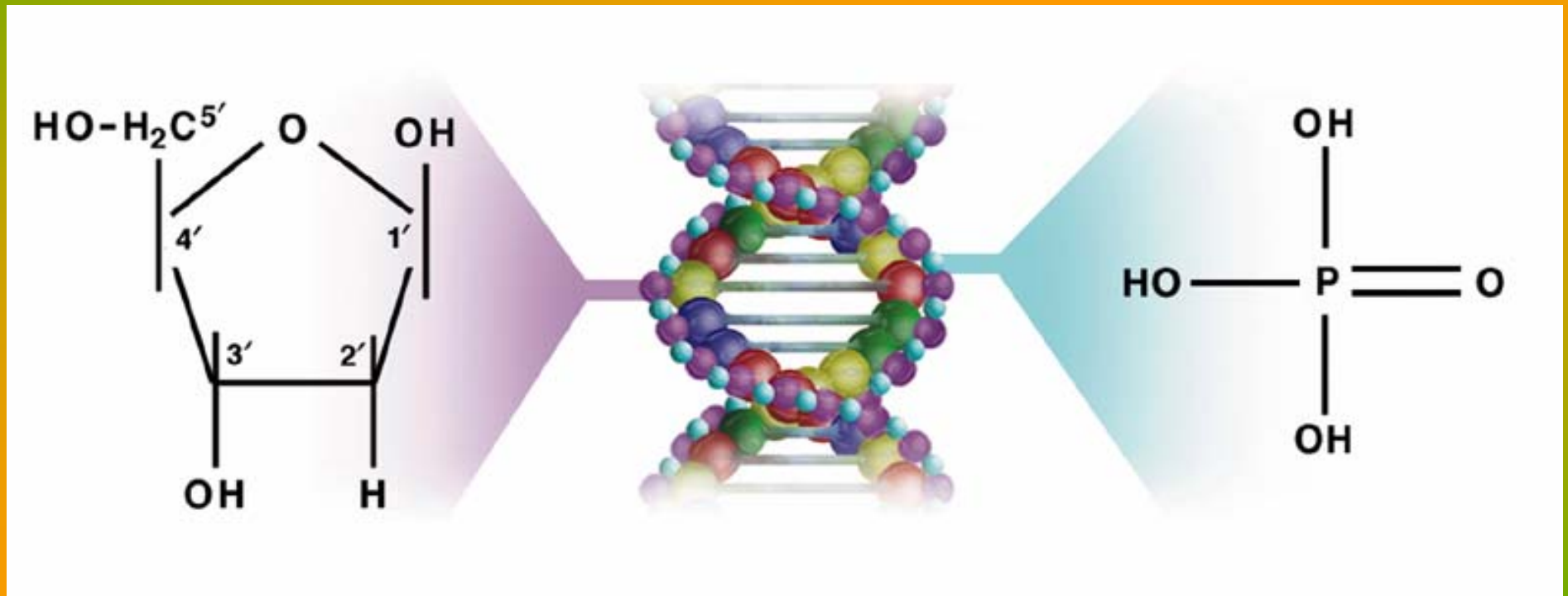


Figure 4-4. Molecular Biology of the Cell, 4th Edition.

# Deoxyribose and Phosphoric Acid



Deoxyribose

Phosphoric Acid

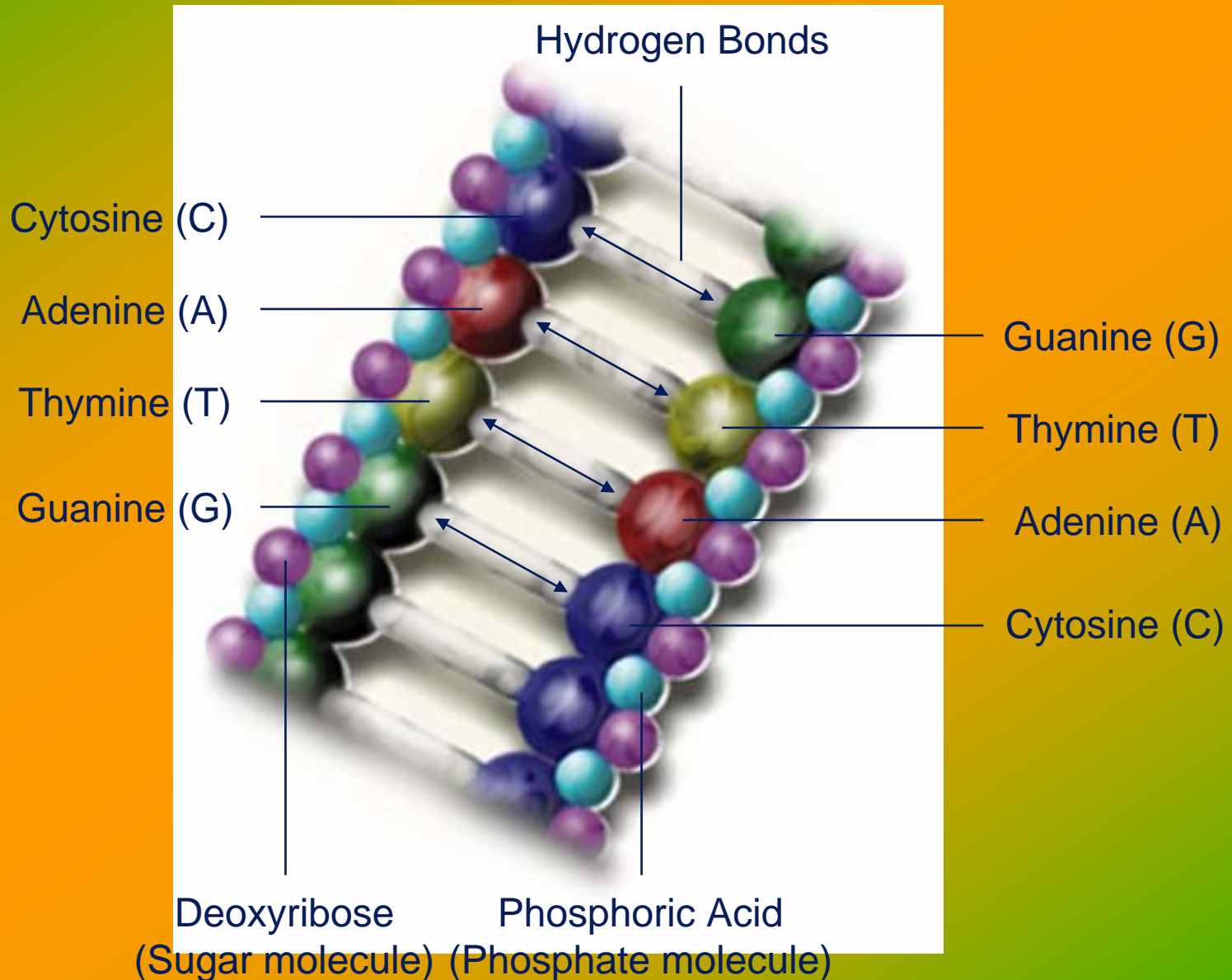
# DNA Base Nomenclature

## DNA Base Nomenclature

Base	Nucleoside	Nucleotide	Abbreviation	Base Ring Structure
Adenine (A)	Adenosine	Adenosine Triphosphate	dATP	Purine
Guanine (G)	Guanosine	Guanosine Triphosphate	dGTP	Purine
Thymine (T)	Thymidine	Thymidine Triphosphate	dTTP	Pyrimidine
Cytosine (C)	Cytidine	Cytidine Triphosphate	dCTP	Pyrimidine



# The Nucleotide Sequence



# Genomes are very similar but very different

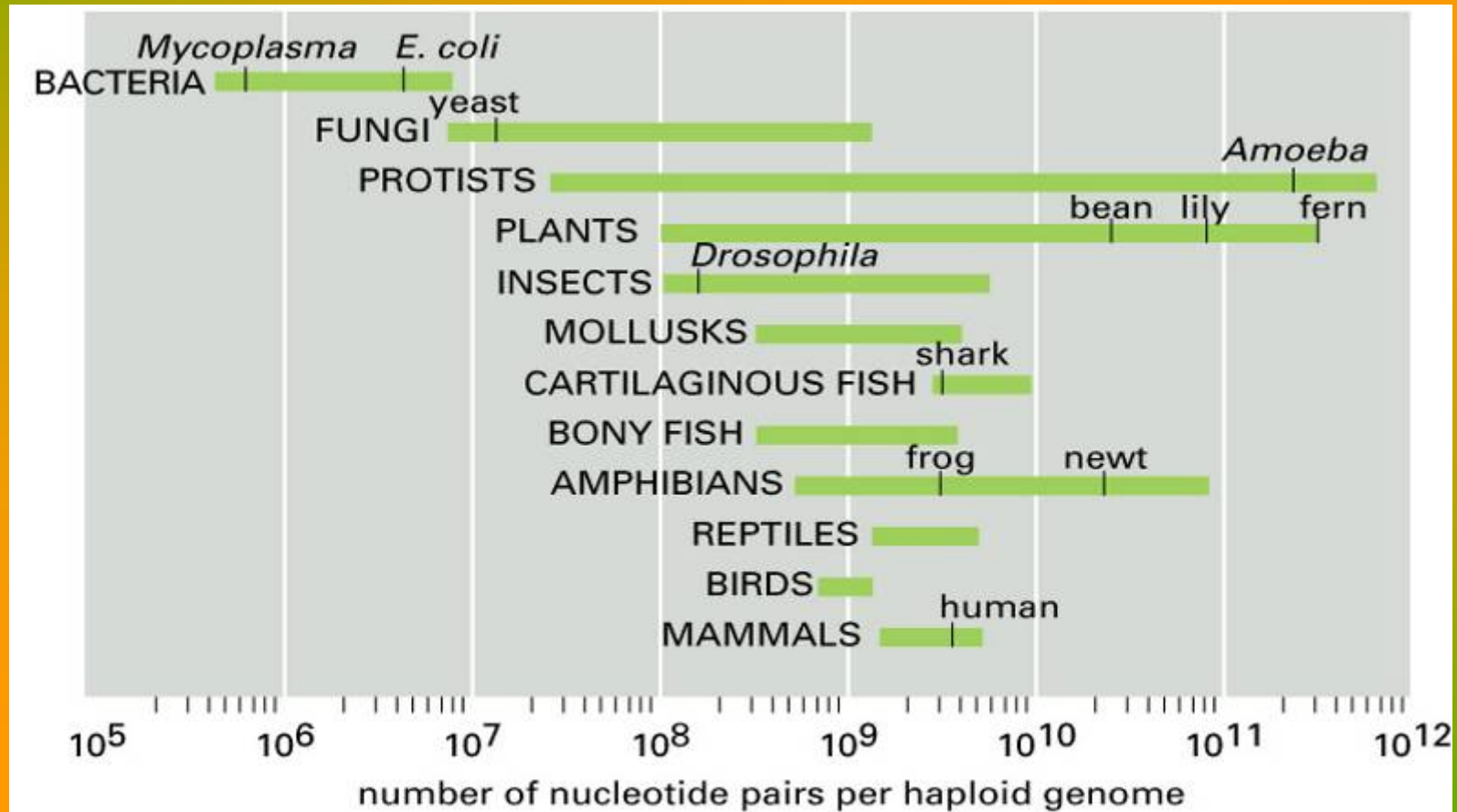


Figure 1-38. Molecular Biology of the Cell, 4th Edition.

# How DNA is organized

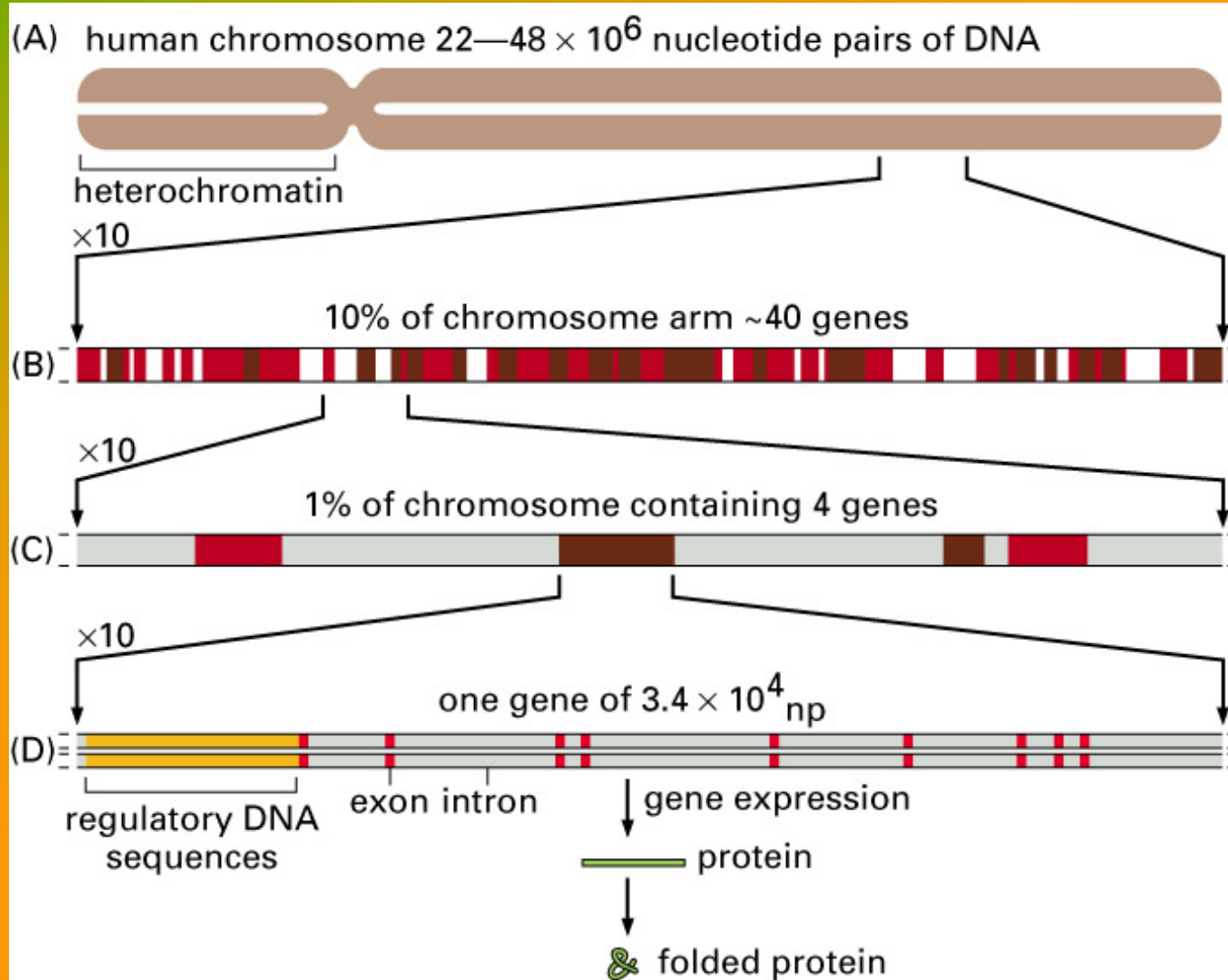


Figure 4-15. Molecular Biology of the Cell, 4th Edition.

# How DNA is organized

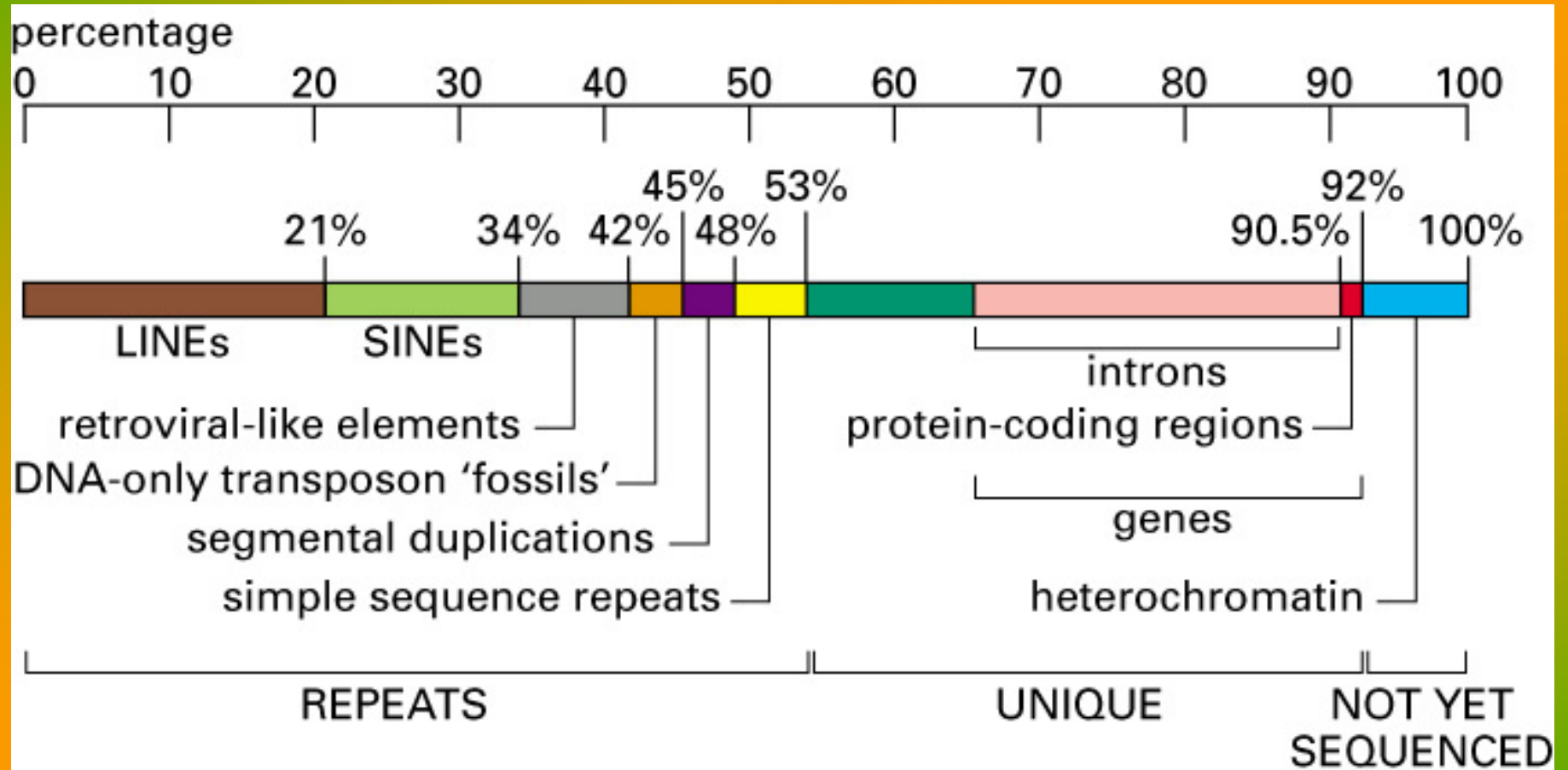
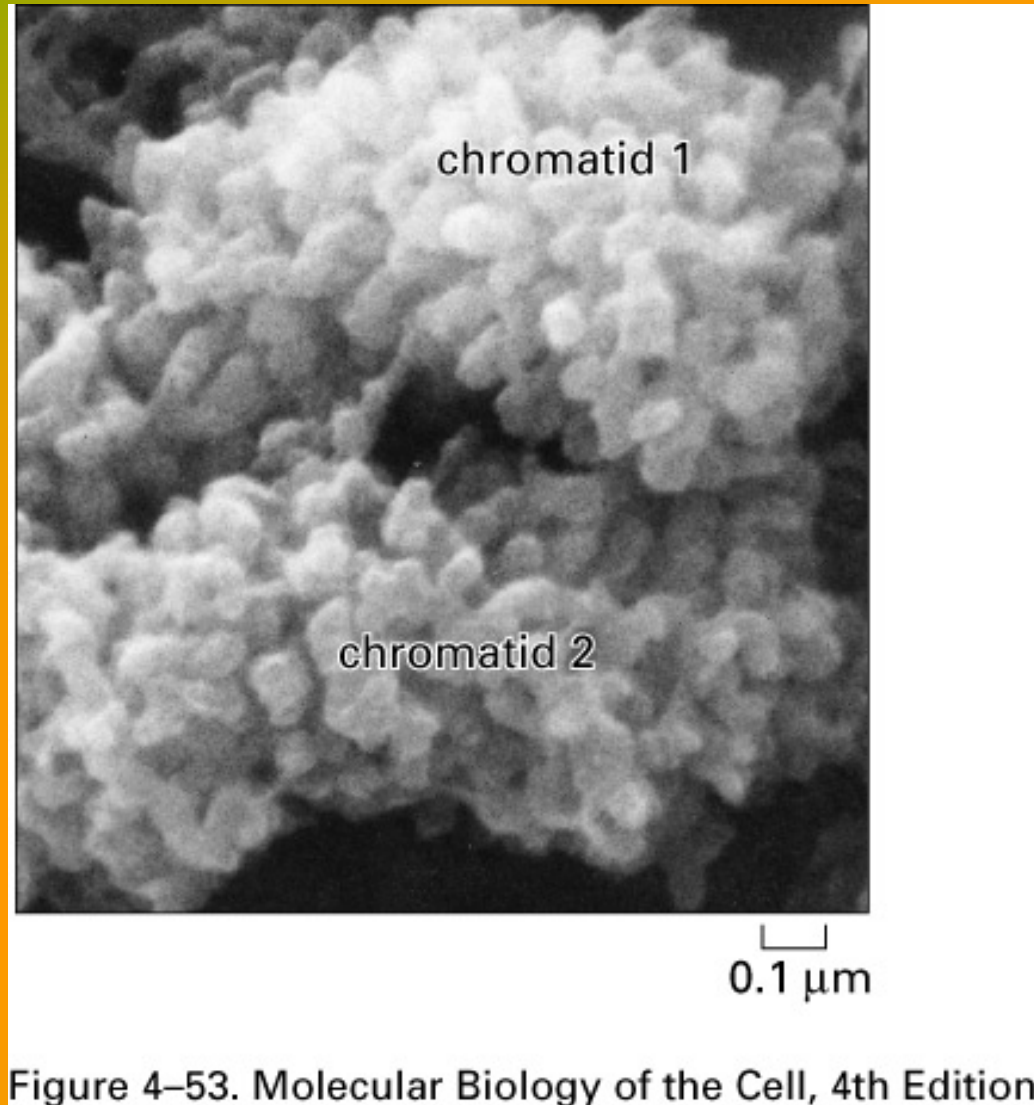
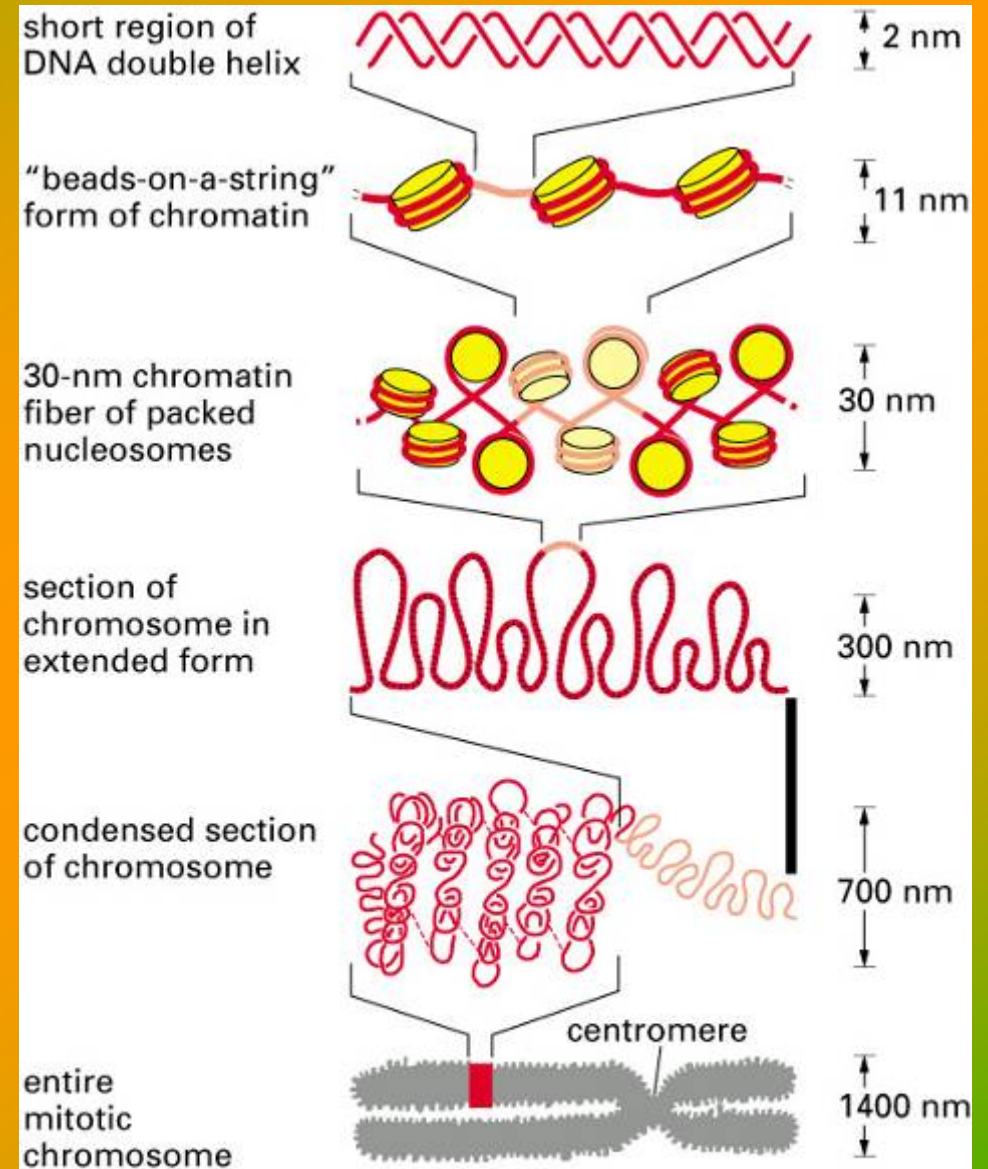


Figure 4-17. Molecular Biology of the Cell, 4th Edition.

# DNA is compacted into chromatin



# DNA package



NET RESULT: EACH DNA MOLECULE HAS BEEN PACKAGED INTO A MITOTIC CHROMOSOME THAT IS 10,000-FOLD SHORTER THAN ITS EXTENDED LENGTH

Figure 4-55. Molecular Biology of the Cell, 4th Edition.

# From chromatin observation to modeling

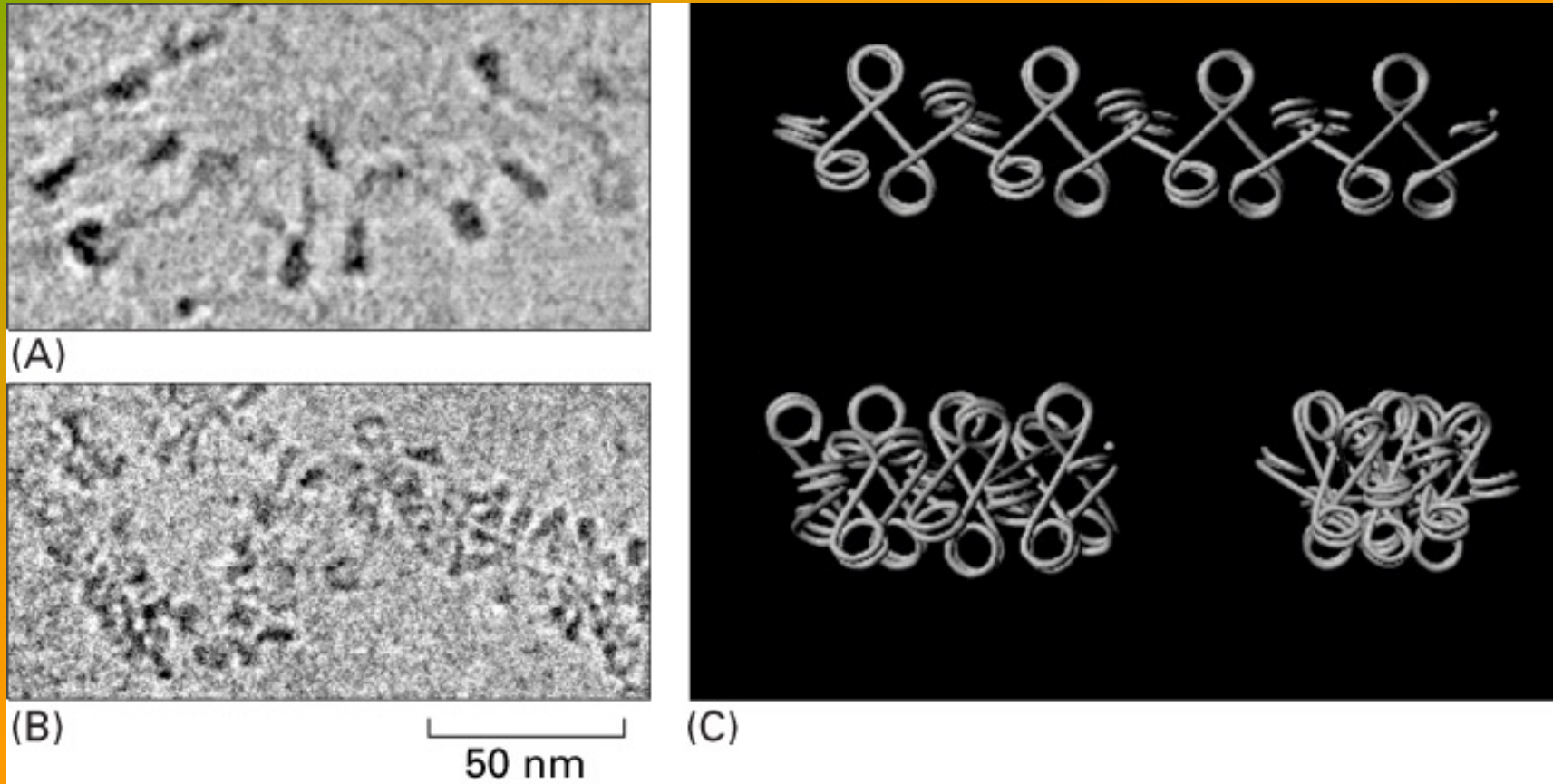


Figure 4-29. Molecular Biology of the Cell, 4th Edition.

# DNA is packaged in nucleosomes

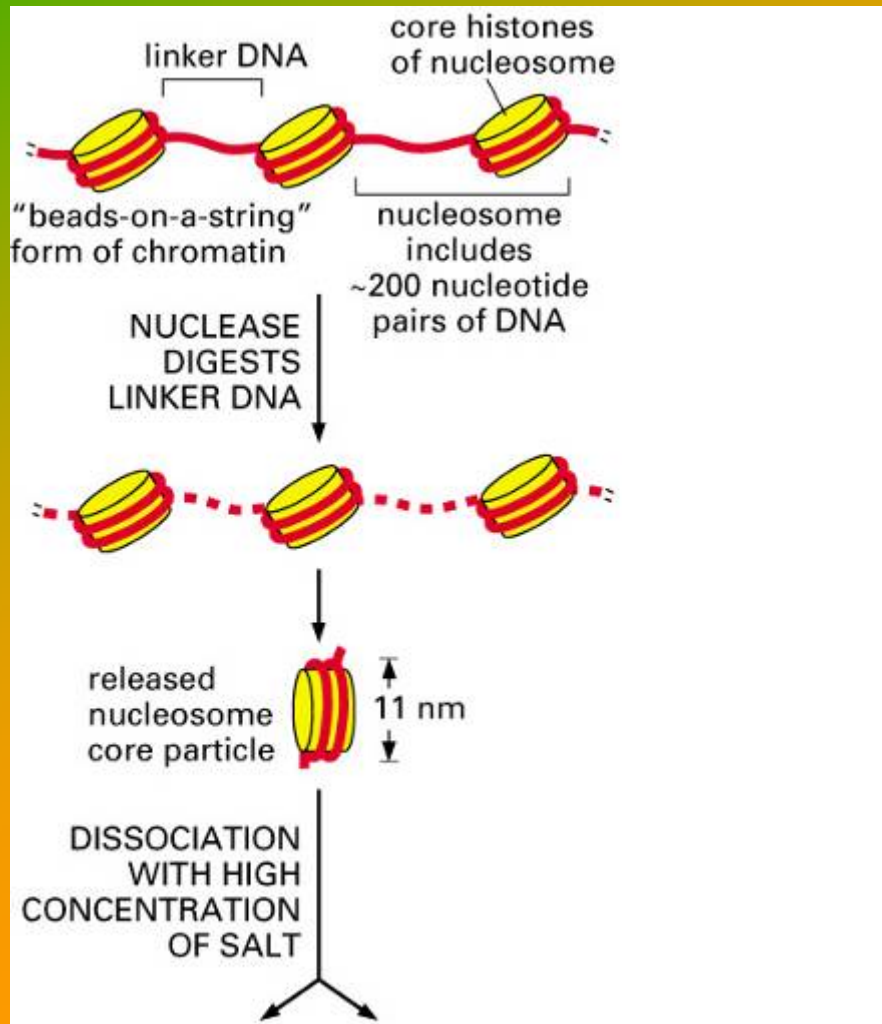


Figure 4-24 part 1 of 2. Molecular Biology of the Cell, 4th Edition.

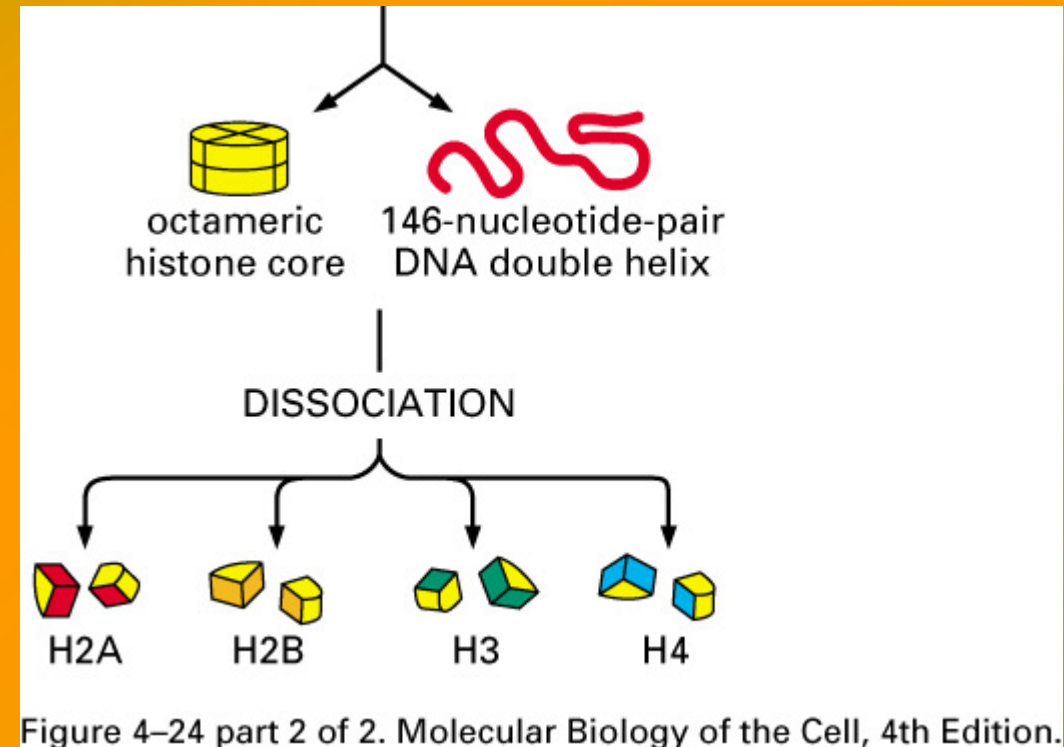
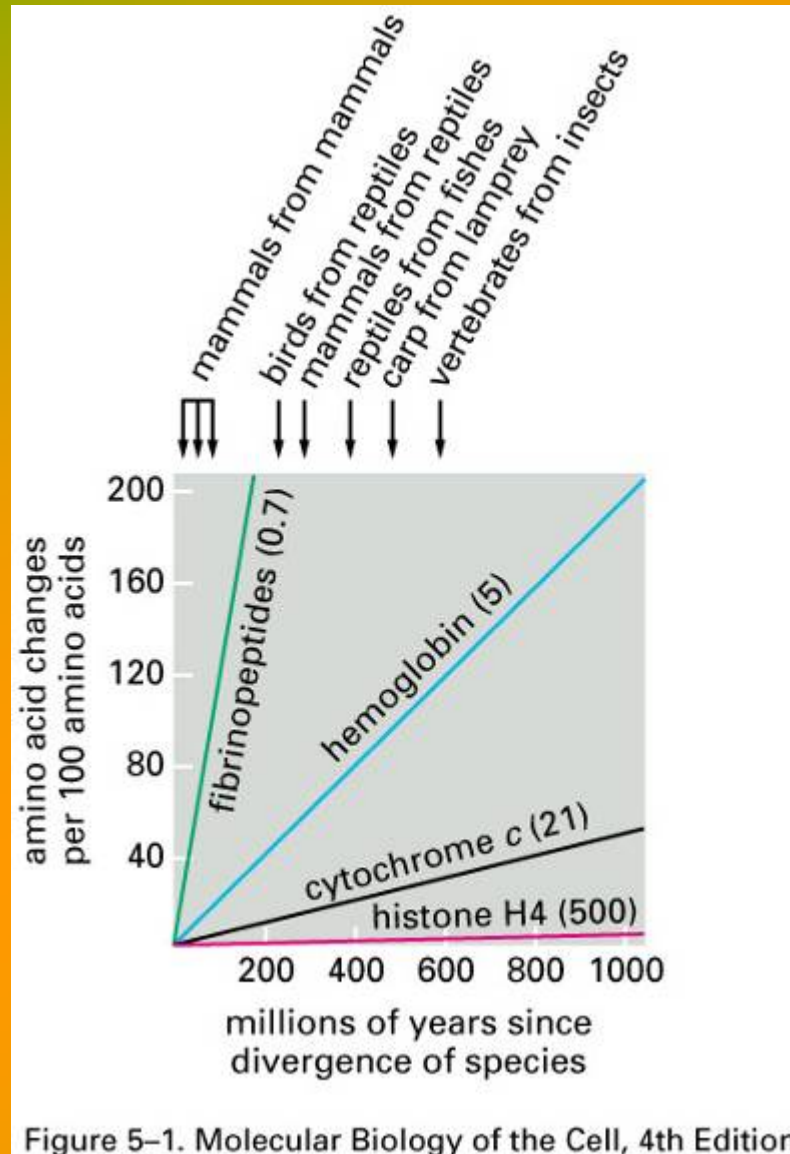


Figure 4-24 part 2 of 2. Molecular Biology of the Cell, 4th Edition.



# Histones are very conserved proteins



# The central dogma

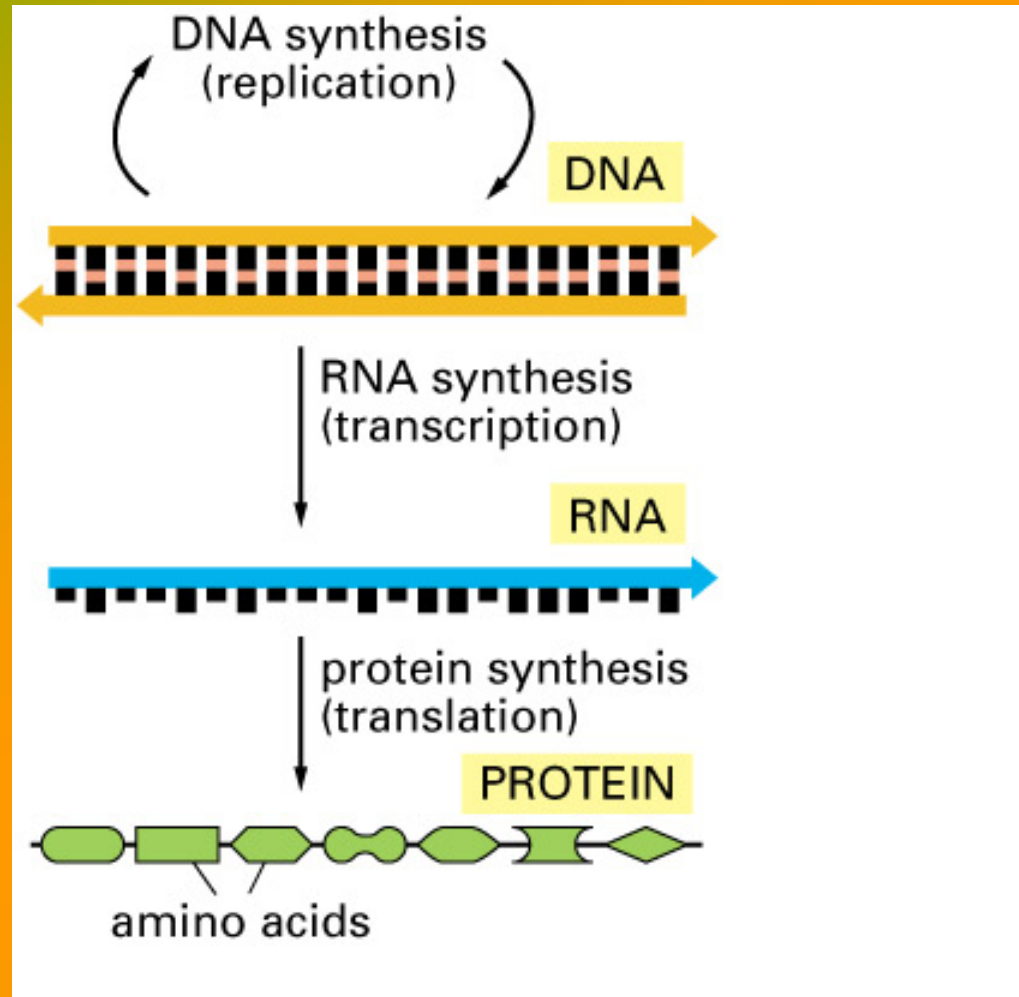


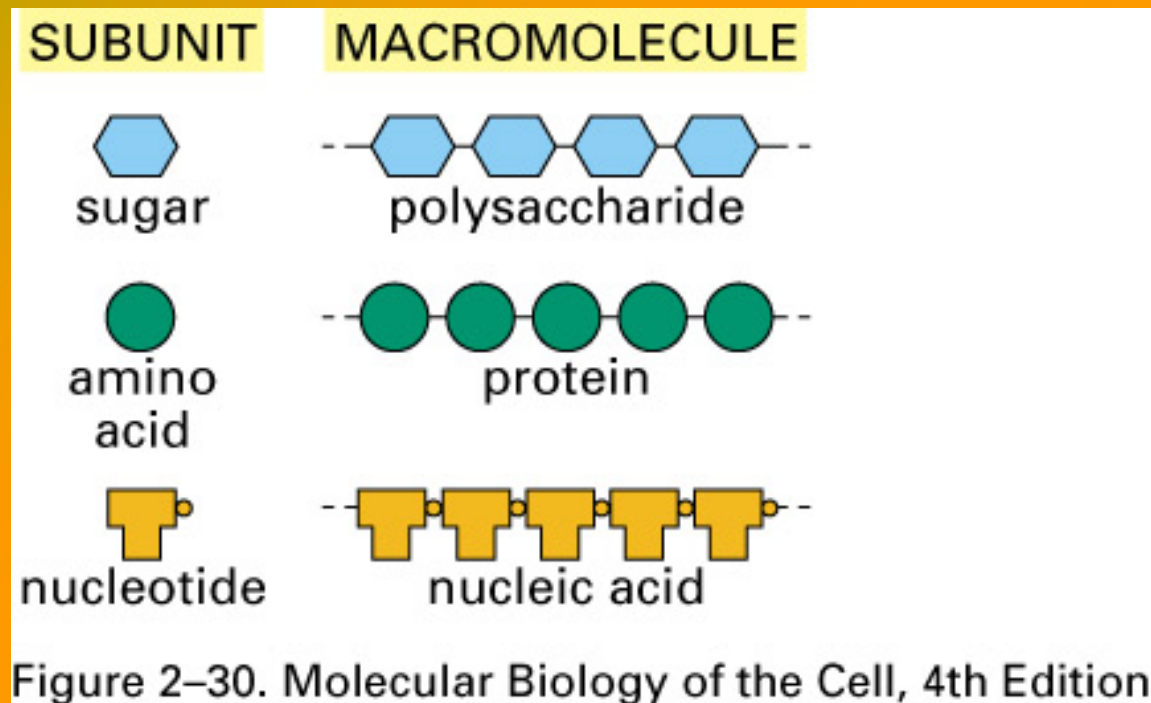
Figure 1-4. Molecular Biology of the Cell, 4th Edition.

Where protein translation is the final aim

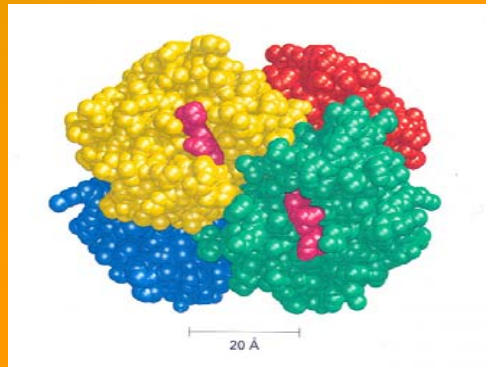


To build a new organism

Like many other molecules DNA works on a polymeric basis



There is a direct relationship between DNA sequence and protein



DNA

Protein

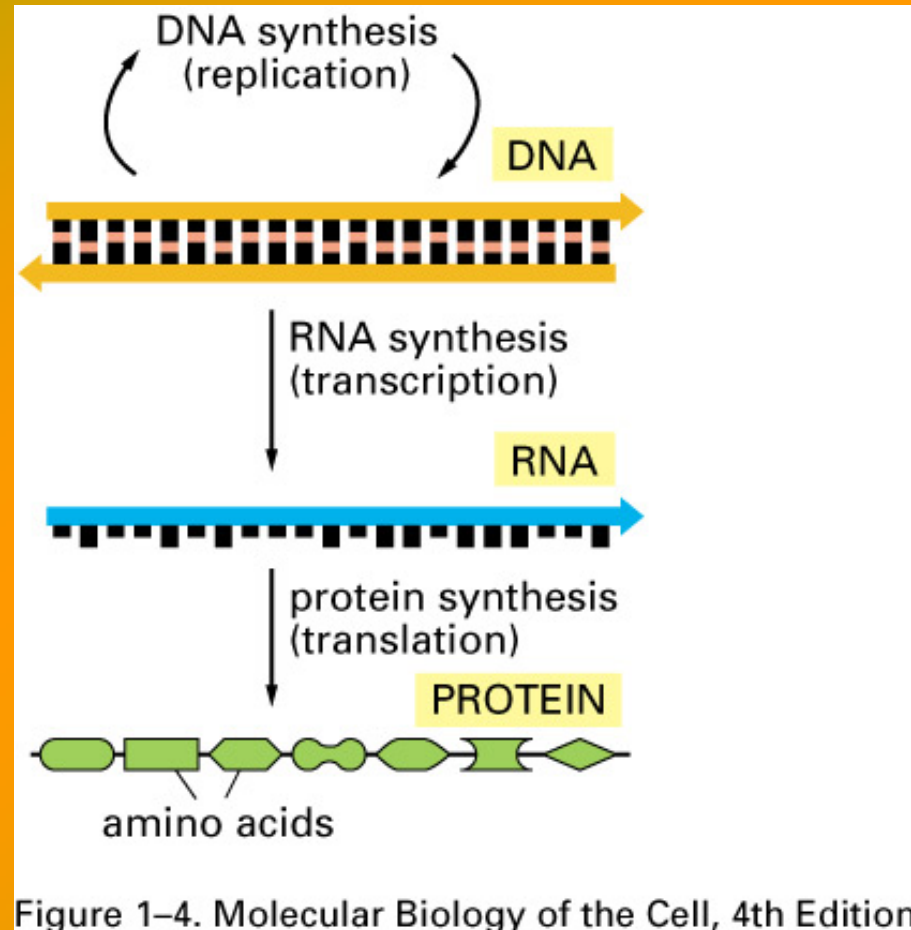
Organisms

# This direct relationship is the genetic code

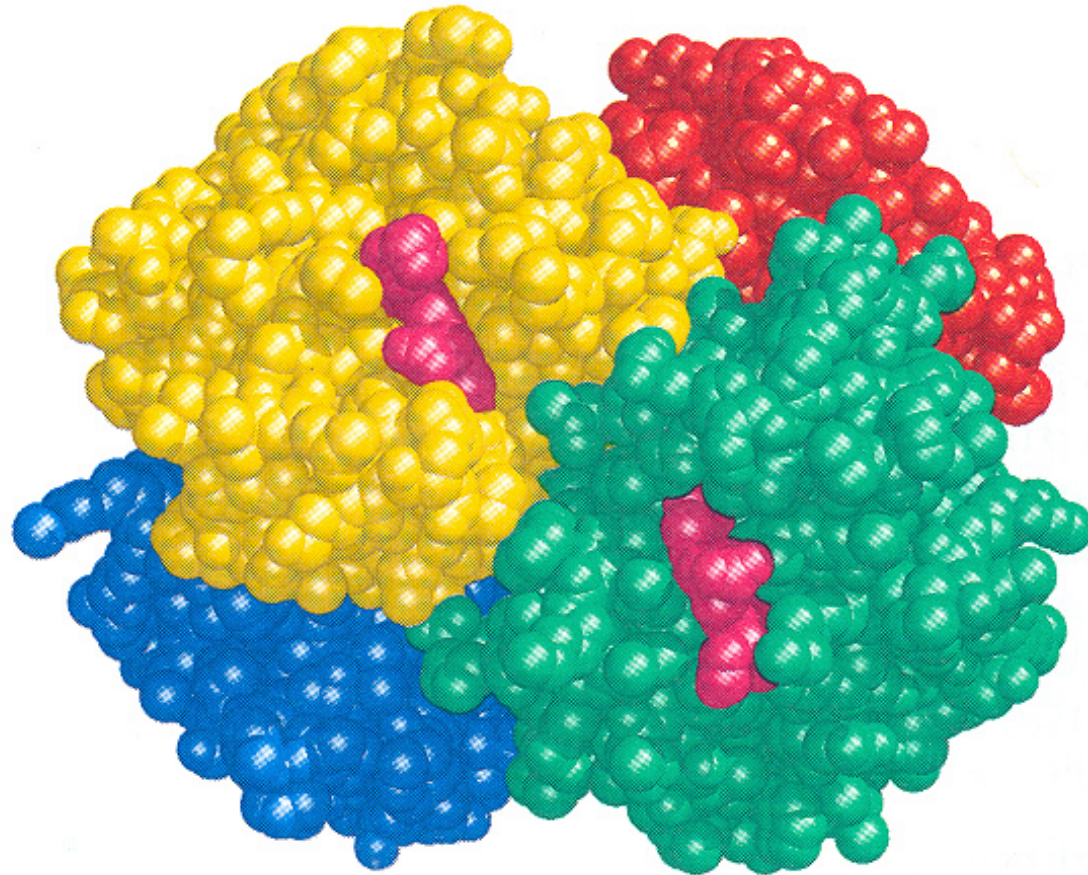
GCA	AGA									
GCC	AGG									
GCG	CGA						GGA			
GCU	CGC						GGC		AUA	
	CGG	GAC	AAC	UGC	GAA	CAA	GGG	CAC	AUC	
	CGU	GAU	AAU	UGU	GAG	CAG	GGU	CAU	AUU	
Ala	Arg	Asp	Asn	Cys	Glu	Gln	Gly	His	Ile	
A	R	D	N	C	E	Q	G	H	I	
UUA							AGC			
UUG							AGU			
CUA				CCA	UCA	ACA			GUA	
CUC				CCC	UCC	ACC			GUC	UAA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

Figure 6-50. Molecular Biology of the Cell, 4th Edition.

# The genetic code allows the translation of the DNA sequence into functional proteins



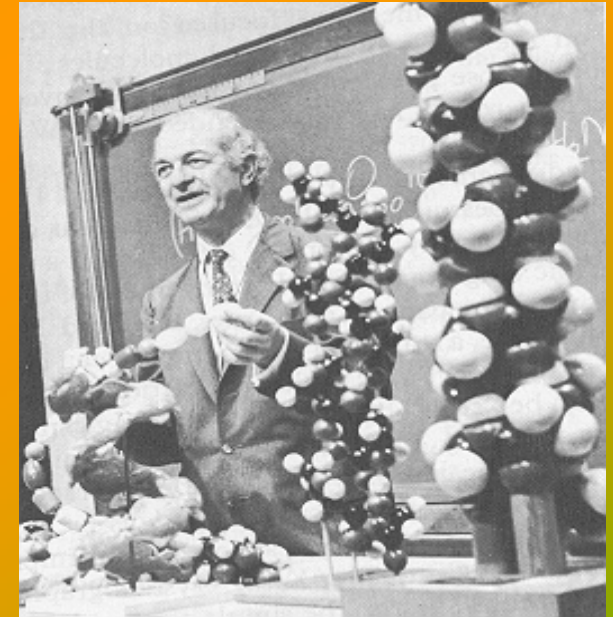
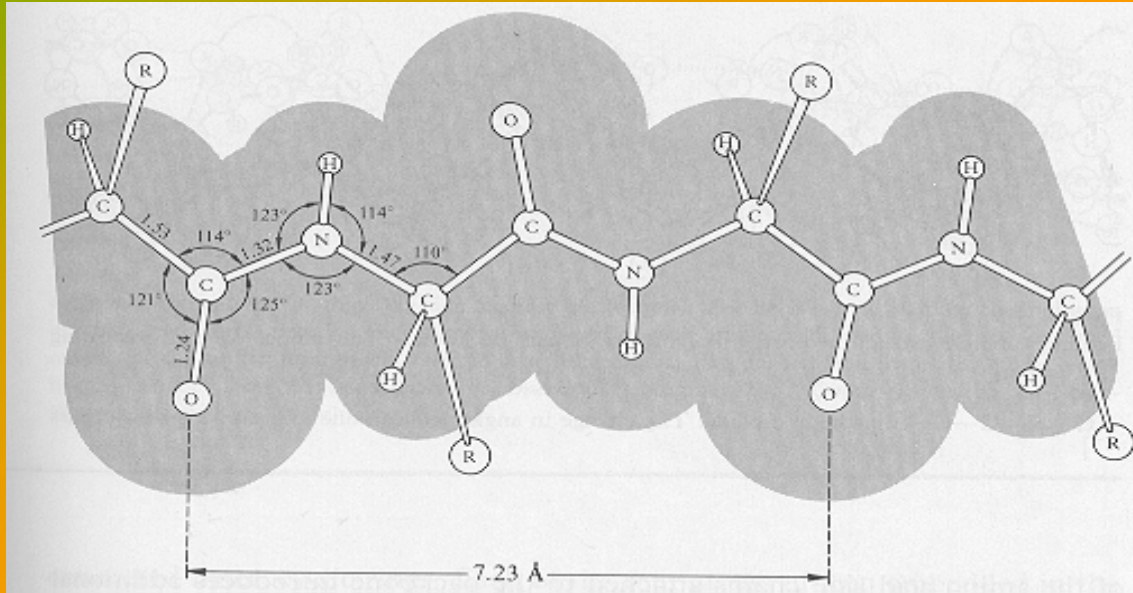
# Proteins



20 Å



# 1951: Linus Pauling and the secondary structure of proteins



Pauling worked out the secondary structure of proteins by crystallographic analysis. From these data he constructed a model of a regular peptide backbone: **the alpha-helix**.

# Secondary structure of proteins

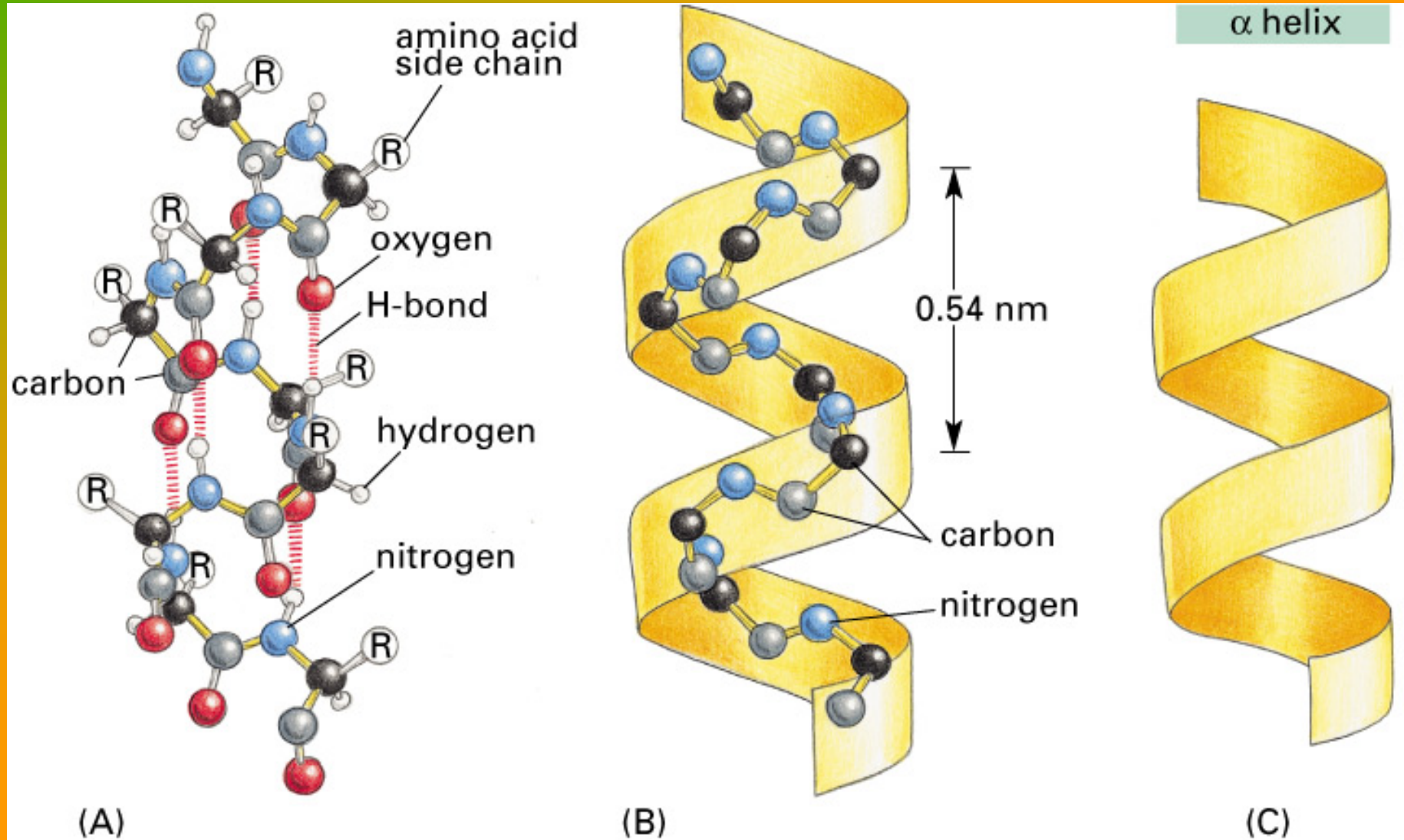
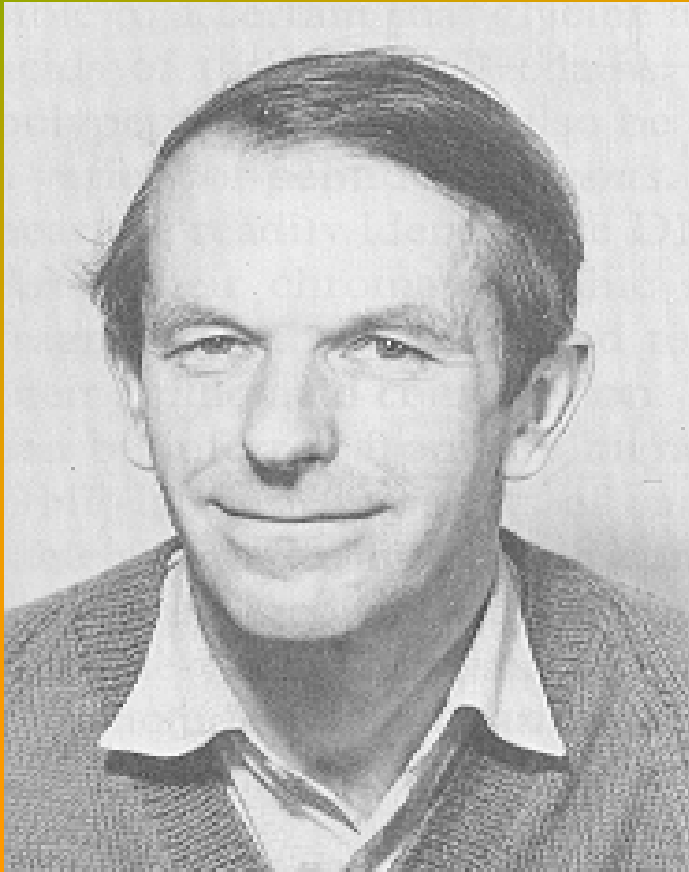
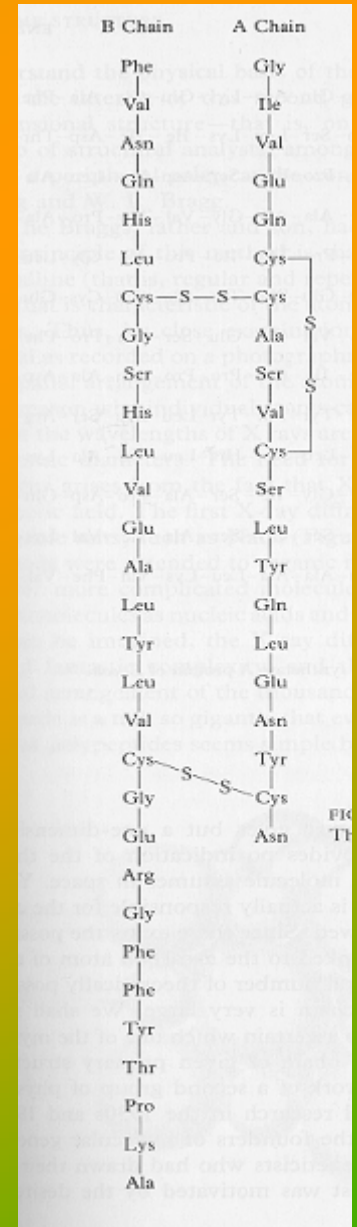


Figure 3-9 part 1 of 2. Molecular Biology of the Cell, 4th Edition.

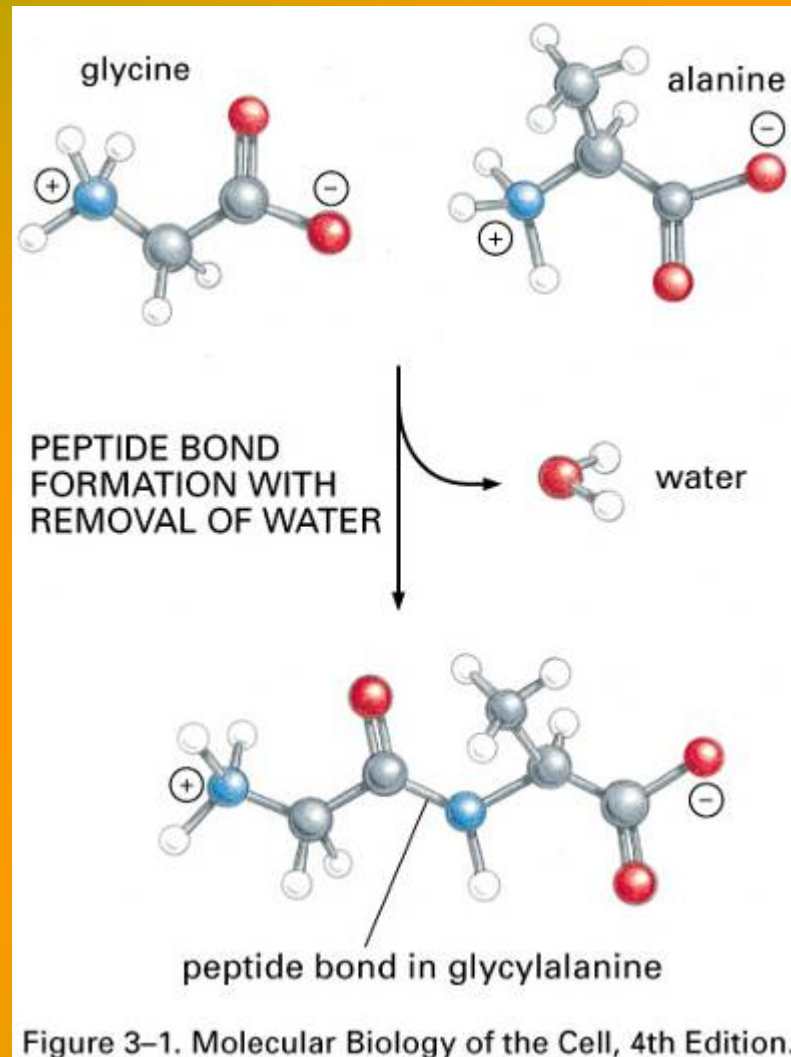
# 1955: Frederick Sanger



Reconstruction of the exact amino acid sequence of the whole insulin molecule



# Proteins are monomeric units bound together



# Getting particular shape and function at pH 7.4

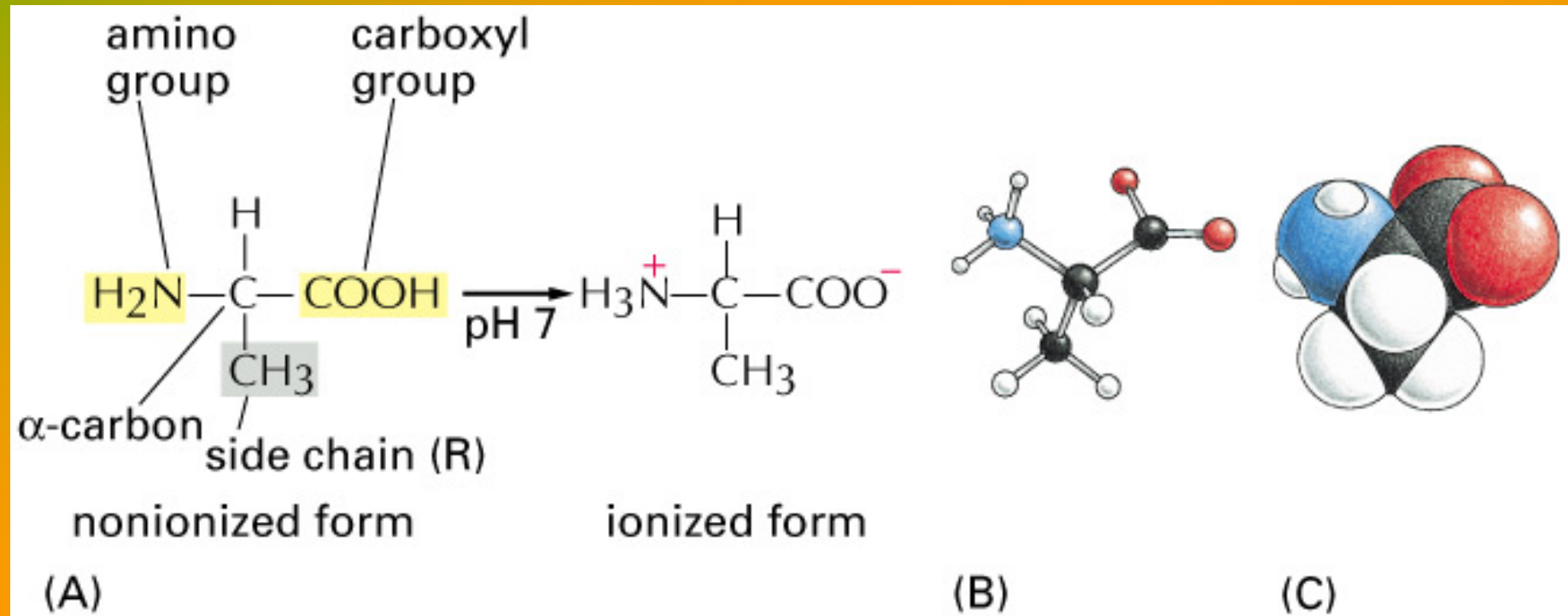
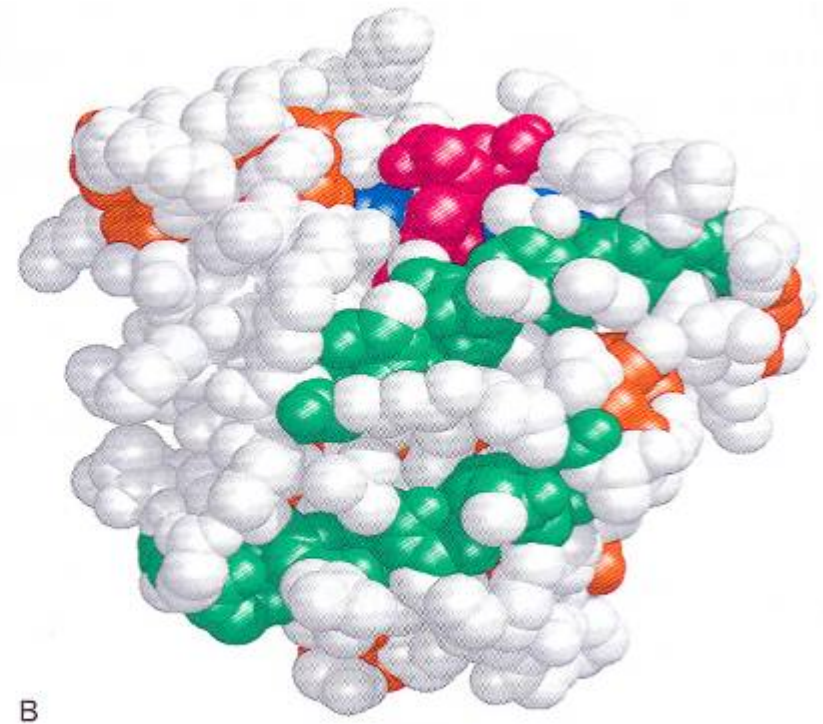
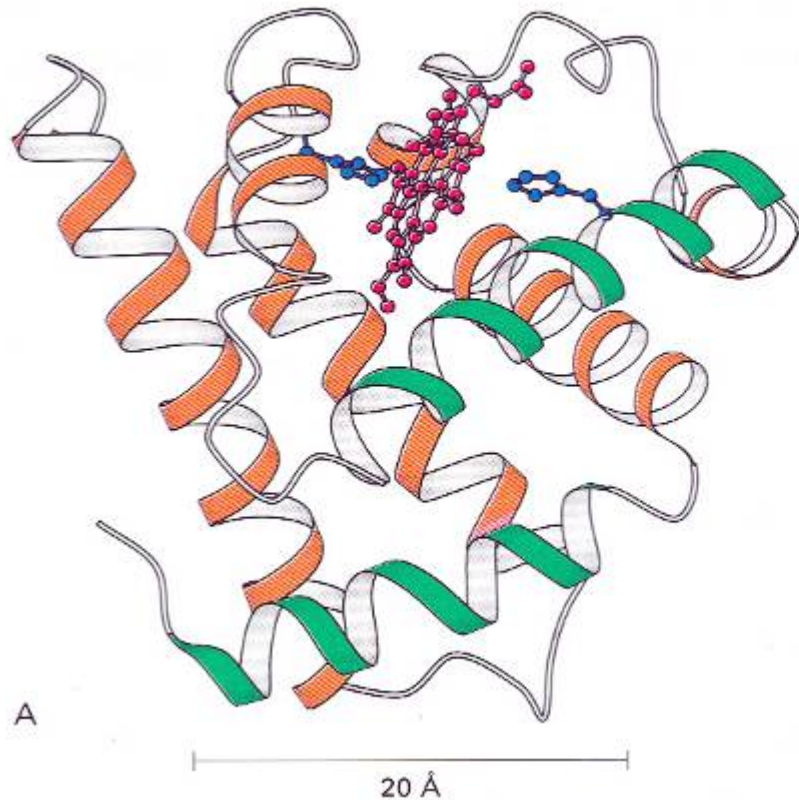


Figure 2-23. Molecular Biology of the Cell, 4th Edition.

The tridimensional structure of the protein is the direct consequence of its aminoacid sequence



And proteins shape our body



DNA is at every time accessible to DNA-binding proteins (regulation of gene expression)

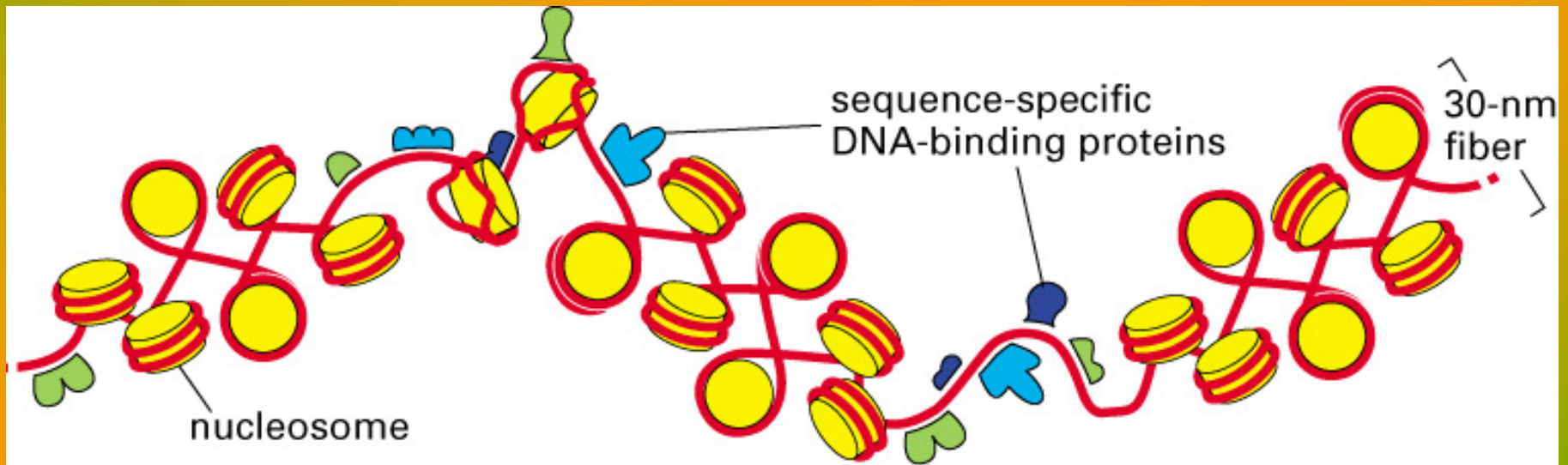


Figure 4-30. Molecular Biology of the Cell, 4th Edition.



# Proteins and DNA interact continuously

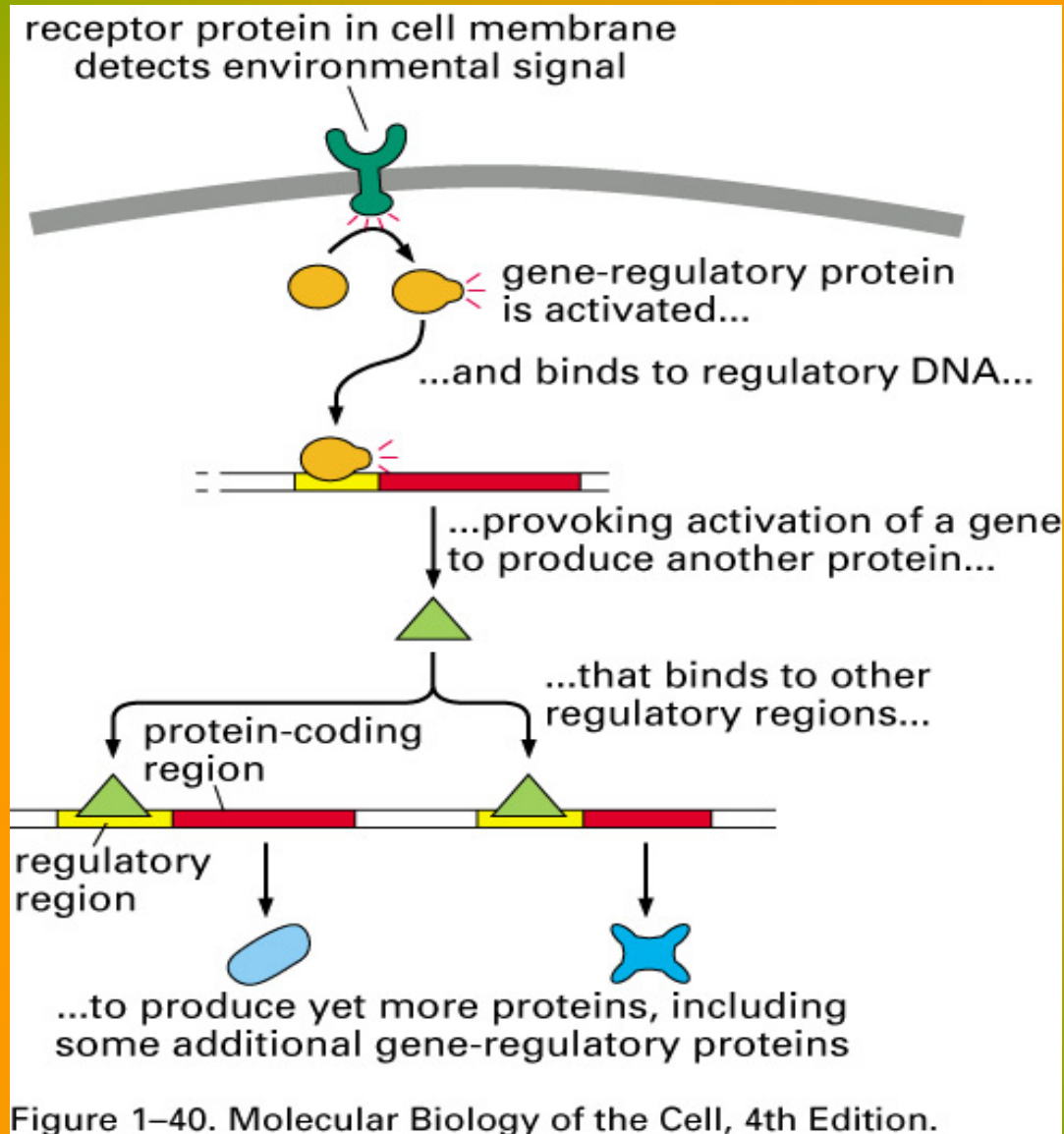


Figure 1-40. Molecular Biology of the Cell, 4th Edition.

# DNA variation is at the origin of evolution



Figure 1-50. Molecular Biology of the Cell, 4th Edition.

# Genes can mutate

C  
A  
A  
A

G A T C G A T C

SEC

CTT  
AAA  
GAT  
TAT  
AAA  
CTA

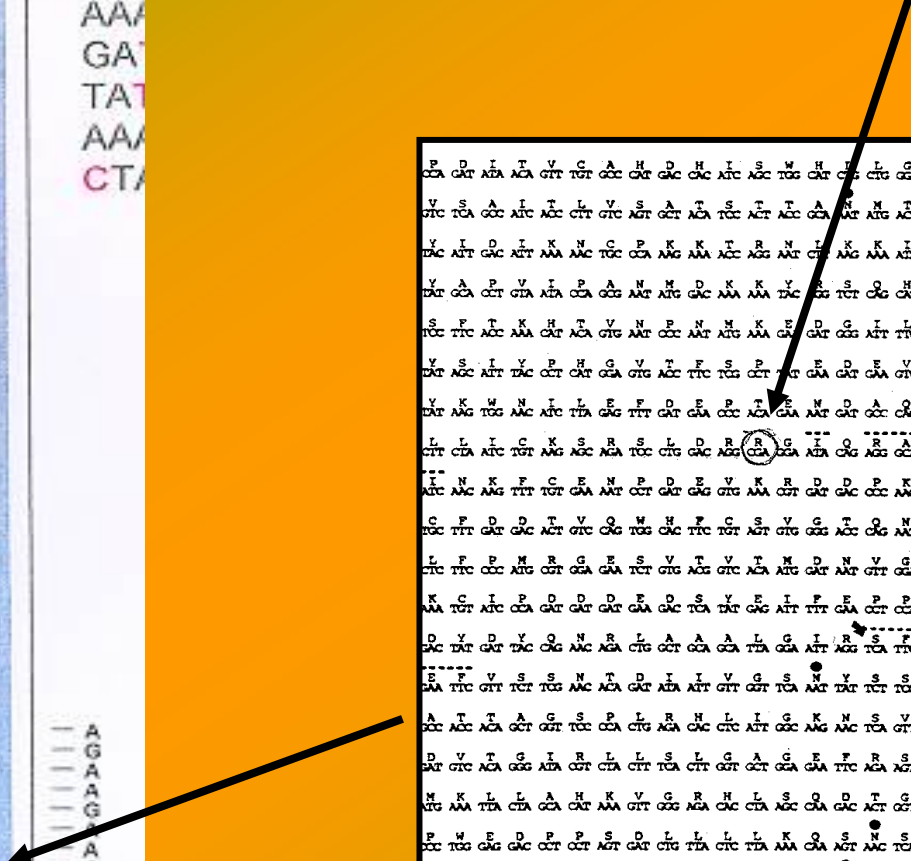
A  
G  
A  
G  
A  
A  
C  
T  
T  
T  
T  
C

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

SEQUENZA 1 SEQUENZA 2

```

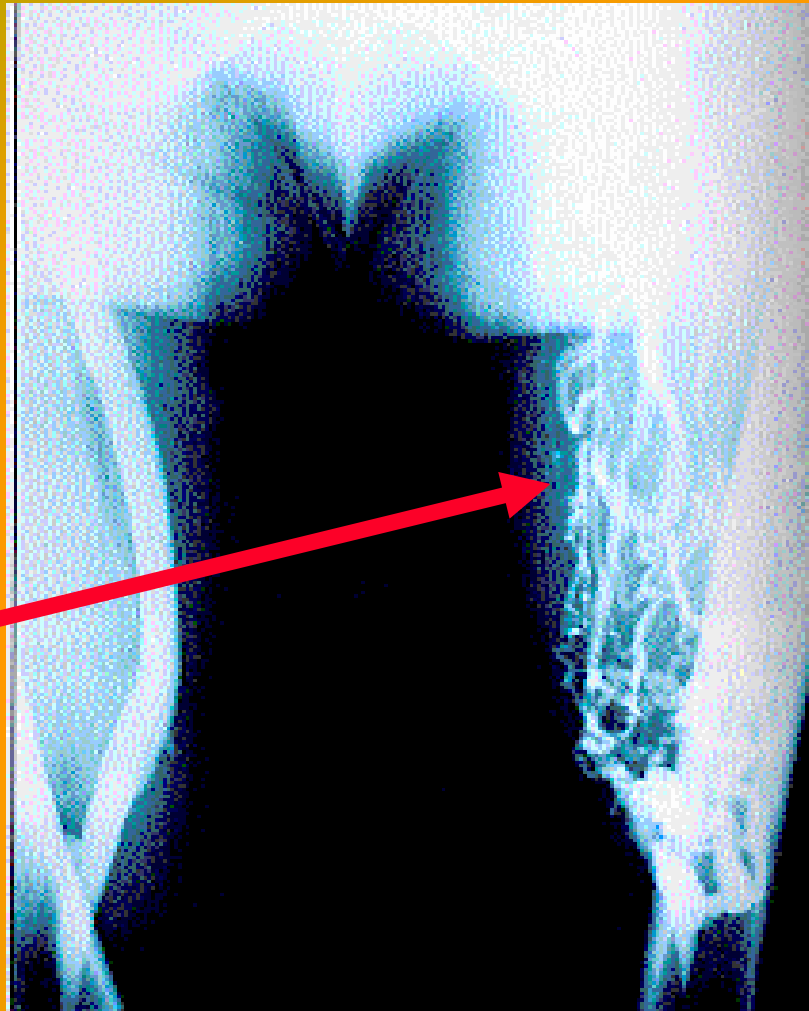
CFA GRT I AEA T GFT VC GGC ACC HGR DHC I AIC SACC WGG HGT CFV LIG GGA ATG SAC SGG P GGA E LTA FTC TTC I CRT FFC NAC GGC GCG V GTC LG GAG GCG
VTC TCA A ITC ACC LTT VTC SAGT ACA SACC TACT ACC AAA MTA MTA TCT VTG GGC P GGA GCG GGA KAG WGG ITC AEA TCT TCT CTC ACC PCA AAA HAT TTG OA OCT
VAC AIT D ITC KAA NAC TC P CA NG AAA TACC KAG SACT CHA KAA AEA TACT RCT ECG OGG RR GCG HCH M KAG RGG W EAA Y FFC I A GCA E GAA
YIAT GCA OCT VEA AEA OGA OGG AAT M ATG GAC AAA AAA TAC GGG TCT CCG CAT TTG GAT MAT P TCA NAC OCA IAT GGA AAA CAT YAT AAG AAA GTF AIG TAC A CA CAG
TCC TTC ACC AAA HAT AEA VTG NAC PAA M ATG KAA EAD GGG AIT TTG GAT P CTI AT I AEA OGG VTC RGA DC TFC KAA ITC VGT TTC KAA N ATG
YIAT SAC IAT YAC P CAT HGT GGA VTG TFC S TGG P TCT EAA D EAA VTC AAC TCT TCT TAC TCA GCG RGG AAC AAC TAT M ATG ITC RGA OCA GTT OCA P OGG
YIAT KAG W NAC AIT TFA GCG TTT D GRT ACA GCA P OCA AEA M GAT A OGC OCT CAG TCC TCT TCA TGA OCA YAC YAC SAC D V GTC D I M RGA DC A GGC TCT GCG
LCT LCT I CT KAG SAC RGA S TCC L GAC RAG R CCA GCA AEA CTG AGG OCA OCA CAC ATC GAG OCG OCT ATG GTC TTT GAT GAG NAC AAA SAC TGG TAC
---
N KAG F TTT TGT GGA AAT CTT D GAG VTG KAA R D GTC DGC OCC KAG TTT VIT EEA S NAC ITC MTC SAC TTT ATC MAT GG VY VGT PCT GFC AIT T
CTC FCT D DACT VTG CAG W GAC HCT TCT CTS S VTC GGG TCT OGA NAT GAA IAT LIG ACC ITC H C PTC ACT GGG HNC TCA TTC IAT YAT GGA KAG R HAT GAG
LCT TTC P OCA HGT GEA GAA TCT V ATG VTC ACA MTC GAT NAT VGT GCA ACT TGG ATG L TCA ACT TOC ATG NAT S S OCA R S KAA KAG L R L K
KAA C ITC P OCA GAT GAT GAA GAC TCA TAT GAG AIT TTT EEA P OCT P OCA GAA S TCT ACA VTG MTC GCT ACA CCG AAA ATG H CAT GAT OCT TPA GAA OCT E GAA GAT GAA
D YIAT D GAT YAC CAG AAC RGA CTG ACA OCA TFA GEA AIT AGG TCA TTC GAA NAC TCA TTG AAC CCG GAA GAA GAG TTC NAT LTT ACT OCT CTA OCT LCT
EAA TTC GTT TCT SGG NTC ACA DACT IA I GTT GTS NAT Y TCT S TCC PCA NAT MAT TAT TAT OCT ACA NAT S TCT SCA GCA GAG OUS TOC SAC PCA YAT TCT EGA DC OCT IST I GAG D GAT
A ACC TCA OCT ACA GGT TCC PCA CTG RGA H CAC LCT IAT GGC KAG NAC TCA GTT CTC NAT TCT TCT S TCC TCA OCA GAG OUS TOC SAC PCA YAT TCT EGA DC OCT IST I GAG D GAT
D V T G I R L L T L T G A G E F R S O E H A K R K G P K V E R A D O A A K H C
PAT GTC ACA GGG ADA OCT LCT CTA CTI GGT OCT GGA GAA TTC RGA AGT CAA GAA OCT GAT ANG OCT ANG GCA OCC ANG GTA GAA AGA GAT OCA GCA GCA AAG CAC
M K L L CTA GCA CAT GAA VTT GGG RGA CAC CTA SAC OCA DAC TCT OCT S TCC GCA ATG RGG OCC TGG GAG D GAC LCT OCT SAC OCA DAC TCT OCT S TCC OCT
PCC W E D P P S D DACT IA I GTT GTS NAT Y TCT S TCC PCA NAT MAT TAT TAT OCT ACA NAT S TCT SCA GCA GAG OUS TOC SAC PCA YAT TCT EGA DC OCT IST I GAG D GAT
D GAA GCA GCA OCT GTT NAC NAC HCT G L T C SAC P OCA OCA MAT TCT TCT S TCC TCA OCA GAG OUS TOC SAC PCA YAT TCT EGA DC OCT IST I GAG D GAT
R V R H K S L O V R A O D G G A K S R L K K A A S O F L I K T R K K K K E K H C T A
GA GTT RGA CAT AAA TCT CTA OCA GTA ACA CAG GAT GGA GCA KAG AGT RGA CTC AAG AAA SAC CAG CTT CTC AIT AAG ACA OCA AAA AAG AAA AAG GAG AAC CAC TACA
LTA TCT CCG AGG ACC TTT H OCT CTA RGA AGT EAA GCC TAC NAC NGA TTT TCA GAA RGA RCT AAC CAT TCG TTG VTG CTT CAT AAA S TCC NAT EGA TCA TCT CTT
NAT O T TTG P S M GAT DTT GGC W I A S C LTT OCT D CAG H N O N TCC SCA NAT D GAC ACT GGT Q GCA AGC CTT OCT OCA GGT CTT YAT
  
```



# Mutation is different from SNP (Single Nucleotide Polymorphism)

- **Mutazione:** rare change in DNA sequence with deleterious effects
- **SNP**
  - more common genetic variation (>1% of the population)
  - actually more than 4 million SNPs are registered in DNA banks
  - they determine susceptibility to a particular pathology
  - they are responsible for a differentiated metabolism of drugs
  - they have no clinical significance (present in non-coding regions of DNA or do not induce any difference at the protein level)

Mutation is a change in DNA coding sequence leading to protein dysfunction



Polymorphism is a more common change in DNA coding sequence leading to a change in protein function

Example the cytochrome 2D6 enzyme responsible for the metabolism of many drugs

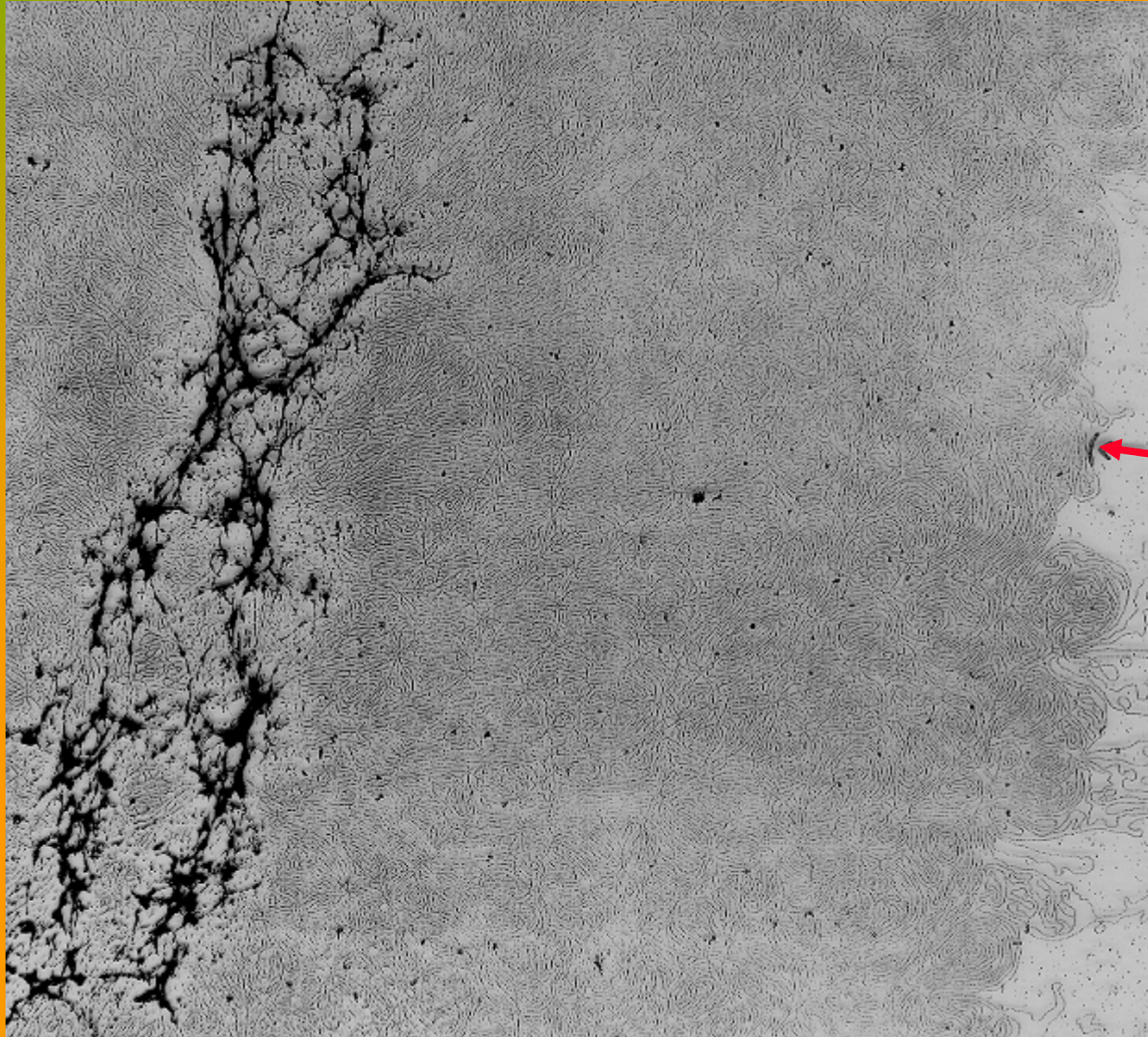
CYP2D6\*1 (WT with normal activity)

...CATCTCCCACCCCCAGGACGCCCTTTTCGC...

CYP2D6\*4 (with reduced activity)

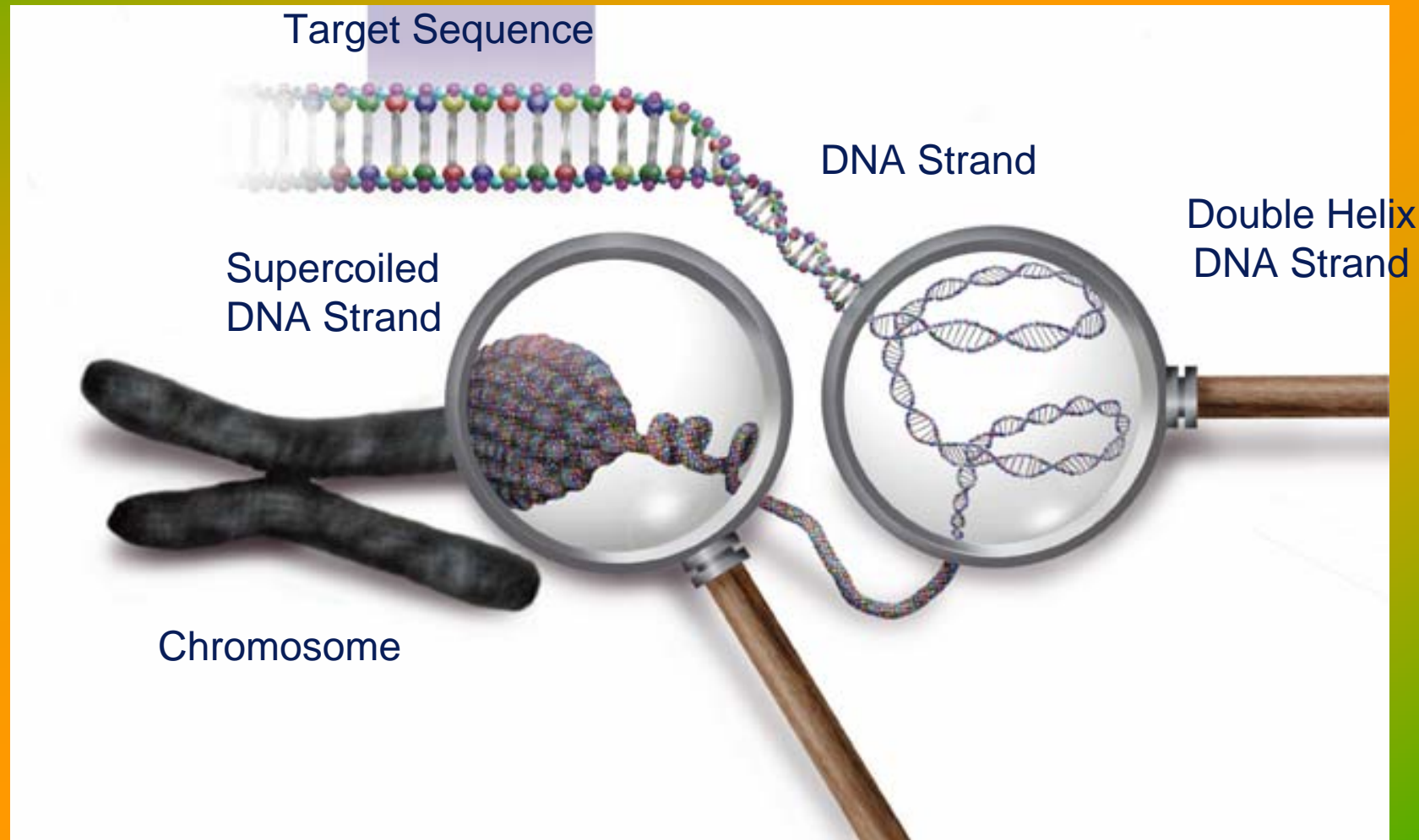
...CATCTCCCACCCCCGGGACGCCCTTTTCGC...

The game is...find the gene!



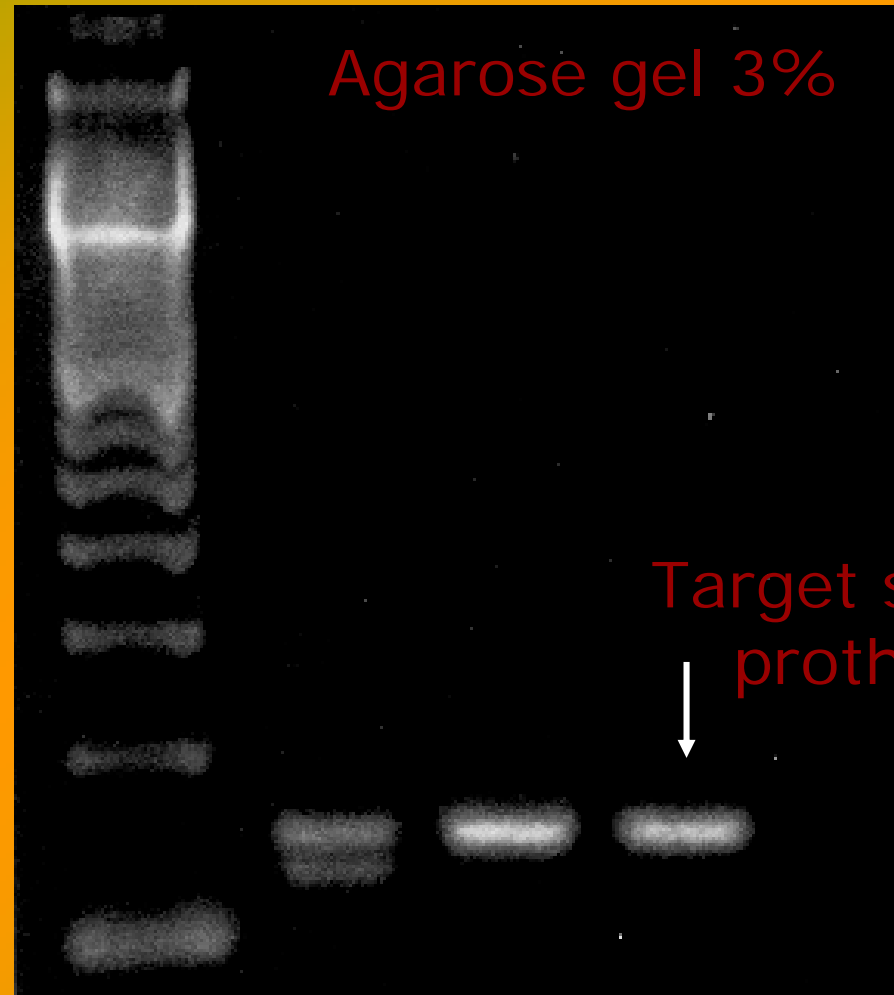
Here's one!

# PCR Amplifies Targeted DNA Sequences

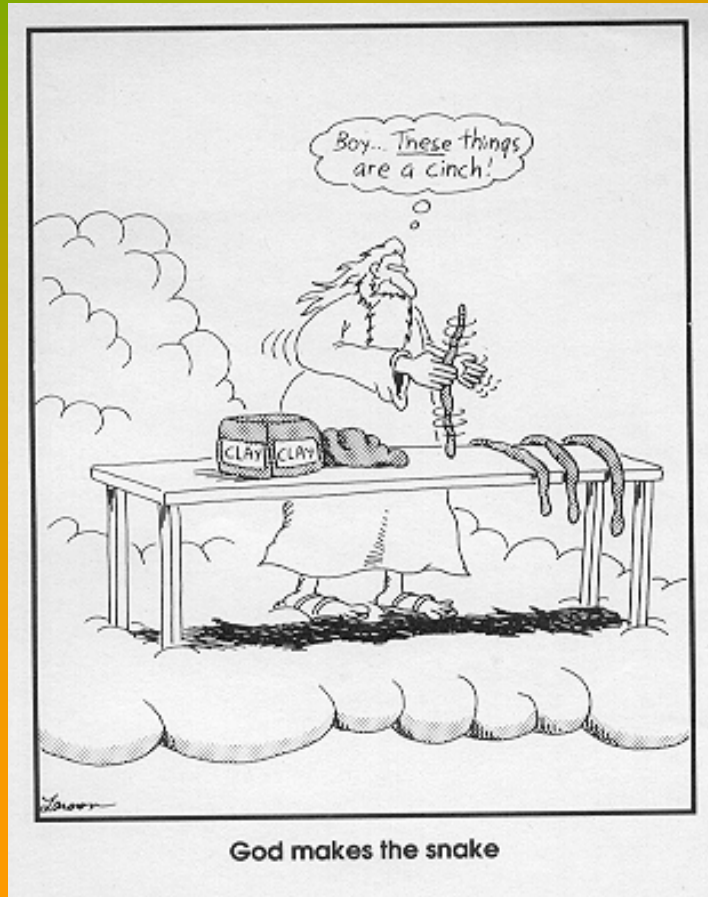




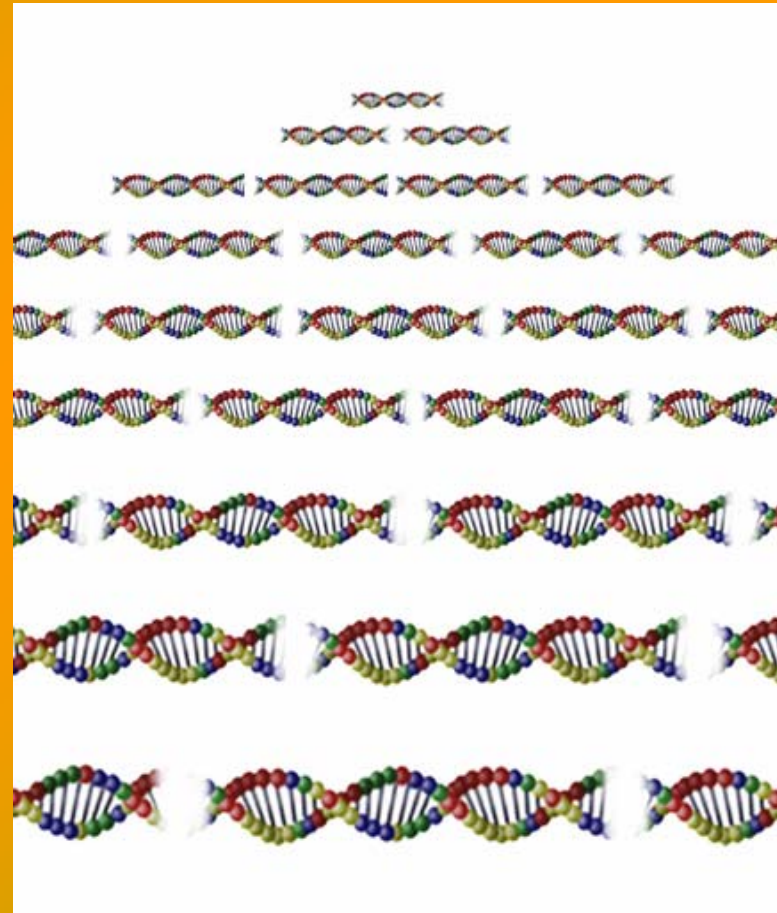
In billions of copies to render them visible by eye



# But how does PCR work?

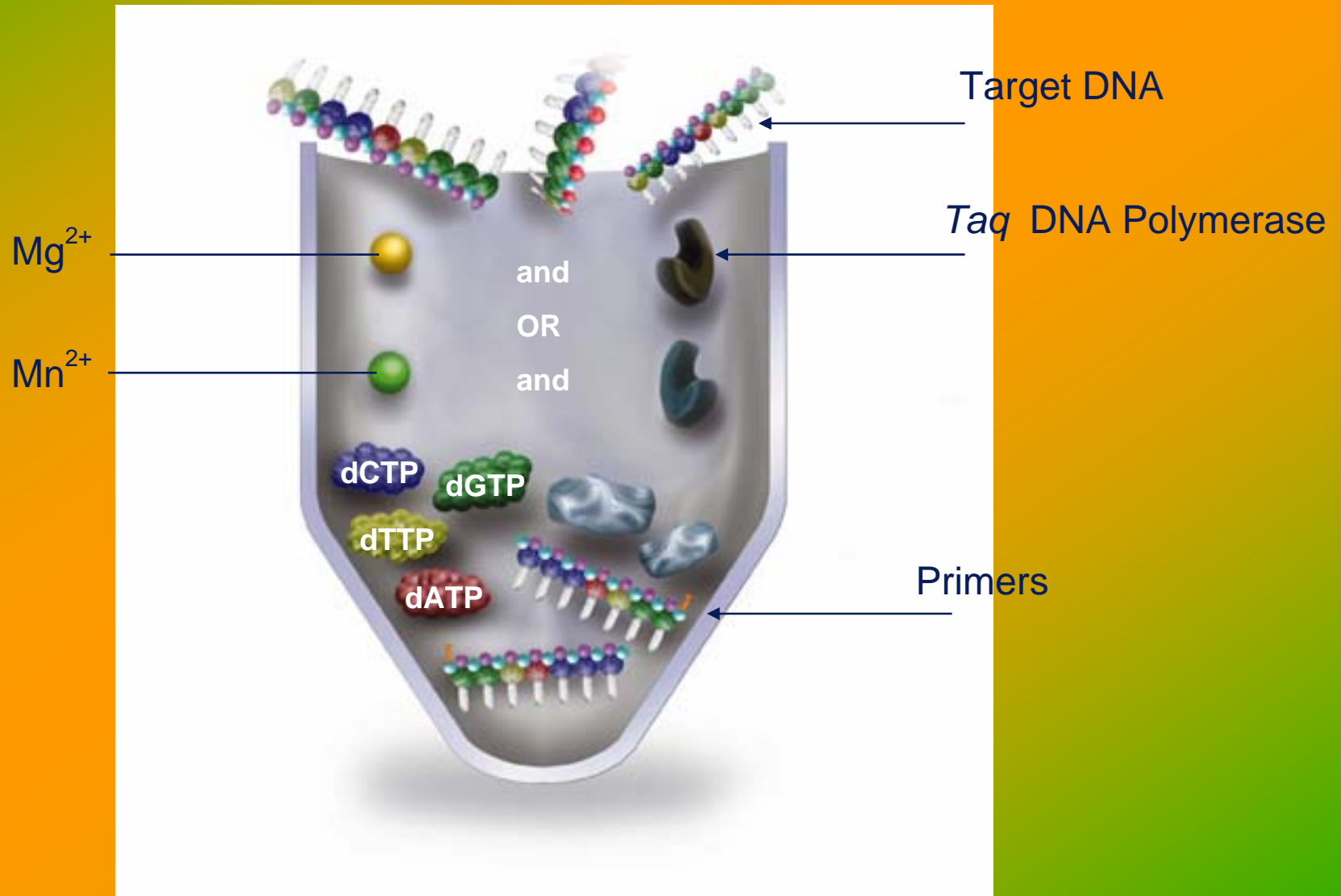


1 copy of target sequence

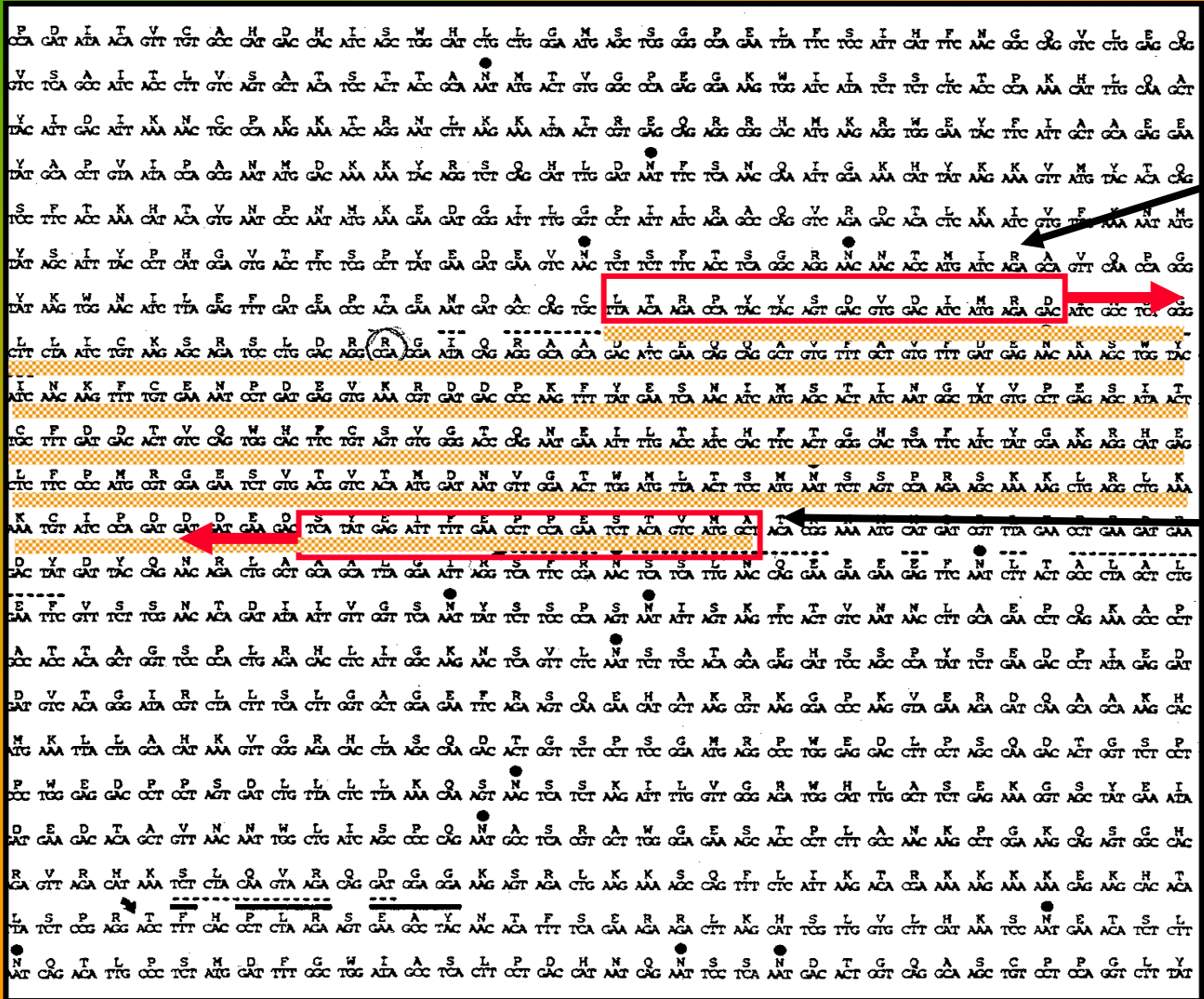


Billions of copies of target sequence

# Cooking PCR: Master Mix Components



# PRIMERS identify the sequence that will be amplified



A forward primer

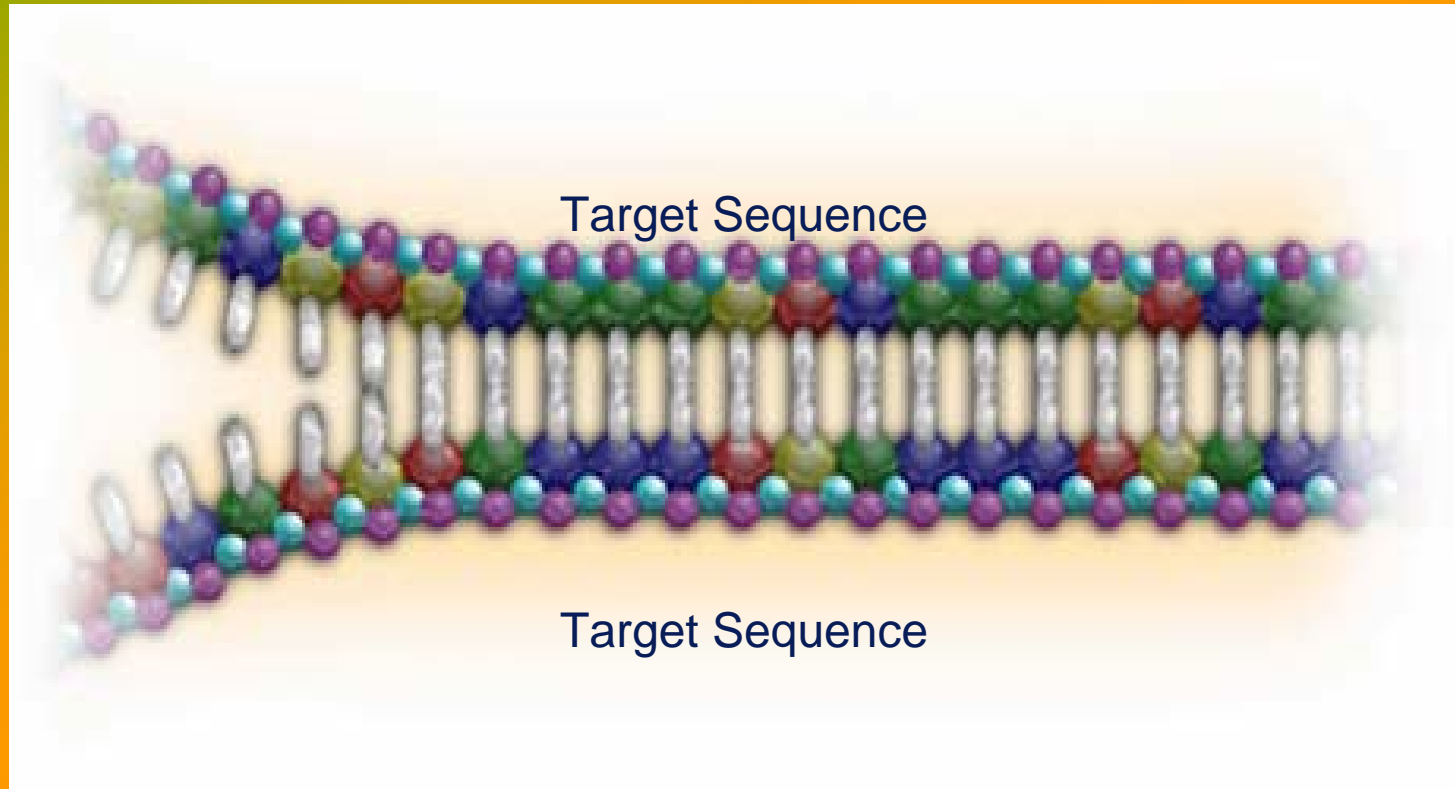
And a reverse primer

# Taq Polymerase adds nucleotides to the growing amplicon

	AmpliTaq	Stoffel	Vent	Pfu
Molecular weight	94kD	61kD	?	92kD
Processivity	50-60 nt	5-10 nt	30-40 nt	?
Extension rate	75nt/sec	>50nt/sec	>80nt/sec	60nt/sec
Thermostability				
half-life time at - 97.5°C	10min	20min	130min	>3h
- 95.0°C	40min	90min	360min	>2h
- 92.5°C	130min	?	?	?
5'-3'-exonuclease activity (pmol/sec/pmol enzyme)	0.3pmol	<0.00001pmol	?	?
3'-5'-exonuclease activity	no	no	yes	yes
Mg <sup>2+</sup> optimum	1.5mM	3.0mM	2.0mM	1.5-2.0mM
KCl optimum	50mM	10mM	10mM	10mM
Reverse transcriptase activity	(yes)	?	?	?

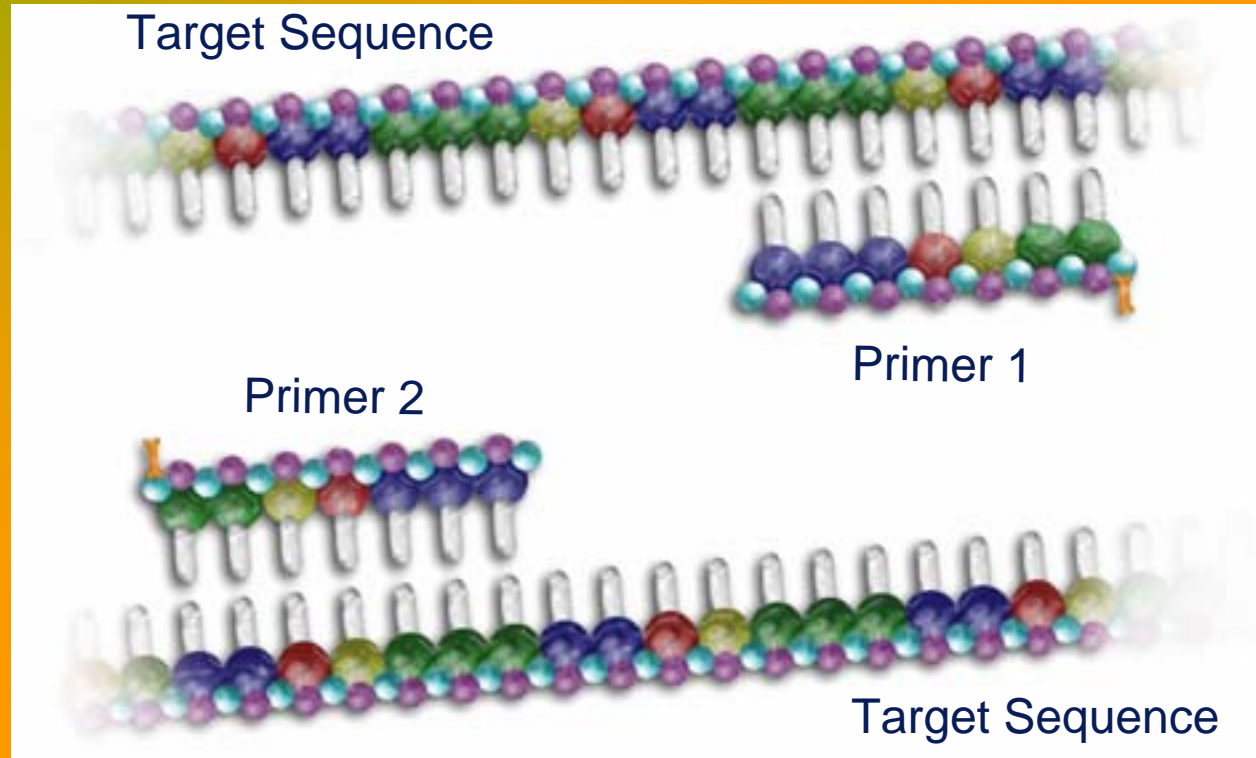
Table 21.2: Some characteristics of different polymerases

# PCR Cycle - Step 1 - Denaturation by Heat



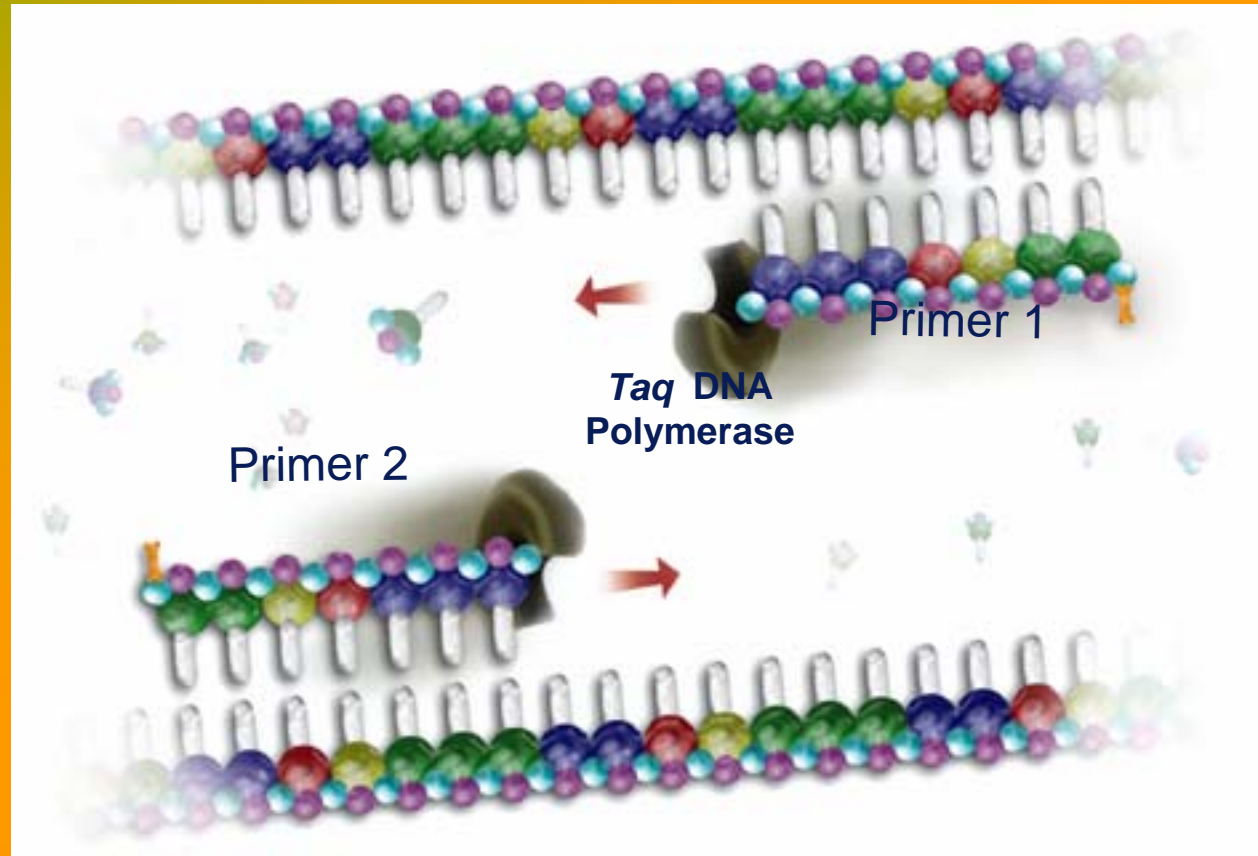
Hydrogen bond denaturation (94°C)

# PCR Cycle - Step 2 - Primer pair anneals to ends of target sequence



Annealing of primers (58-68°C)

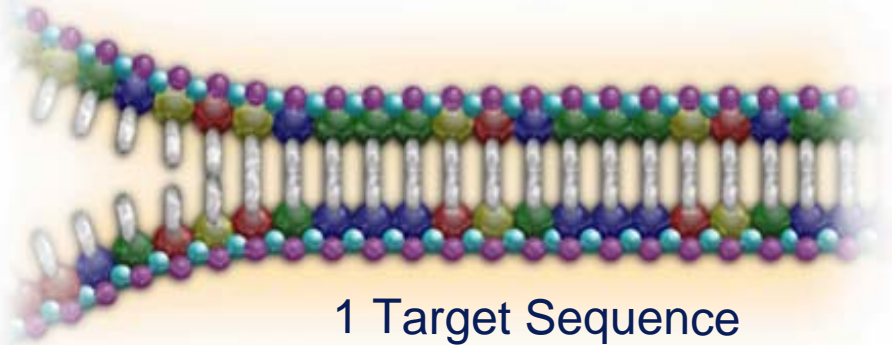
# PCR Cycle - Step 3 - *Taq* DNA Polymerase Catalyses Primer Extension as Complementary Nucleotides are Incorporated



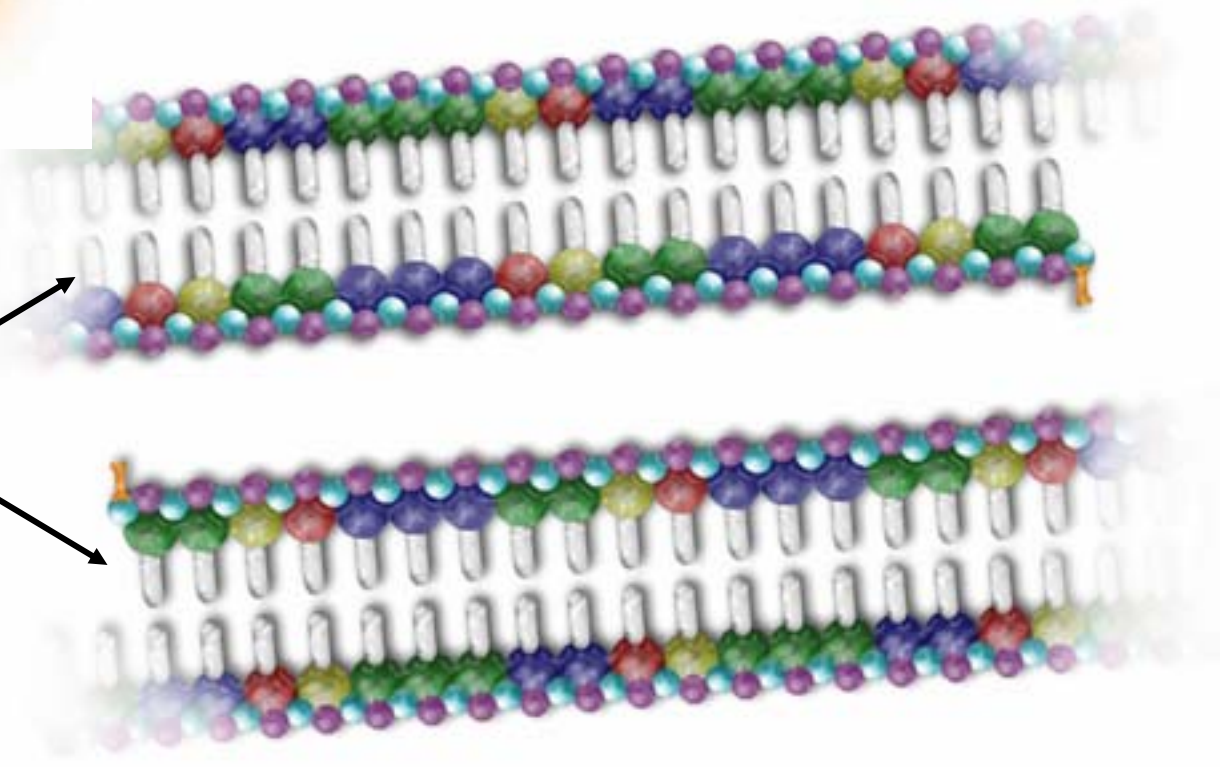
Extension of the target sequence copy (72°C)



# End of the 1st PCR cycle results in two copies of the target sequence



2 identical Target Sequences



# PCR: the first cycle

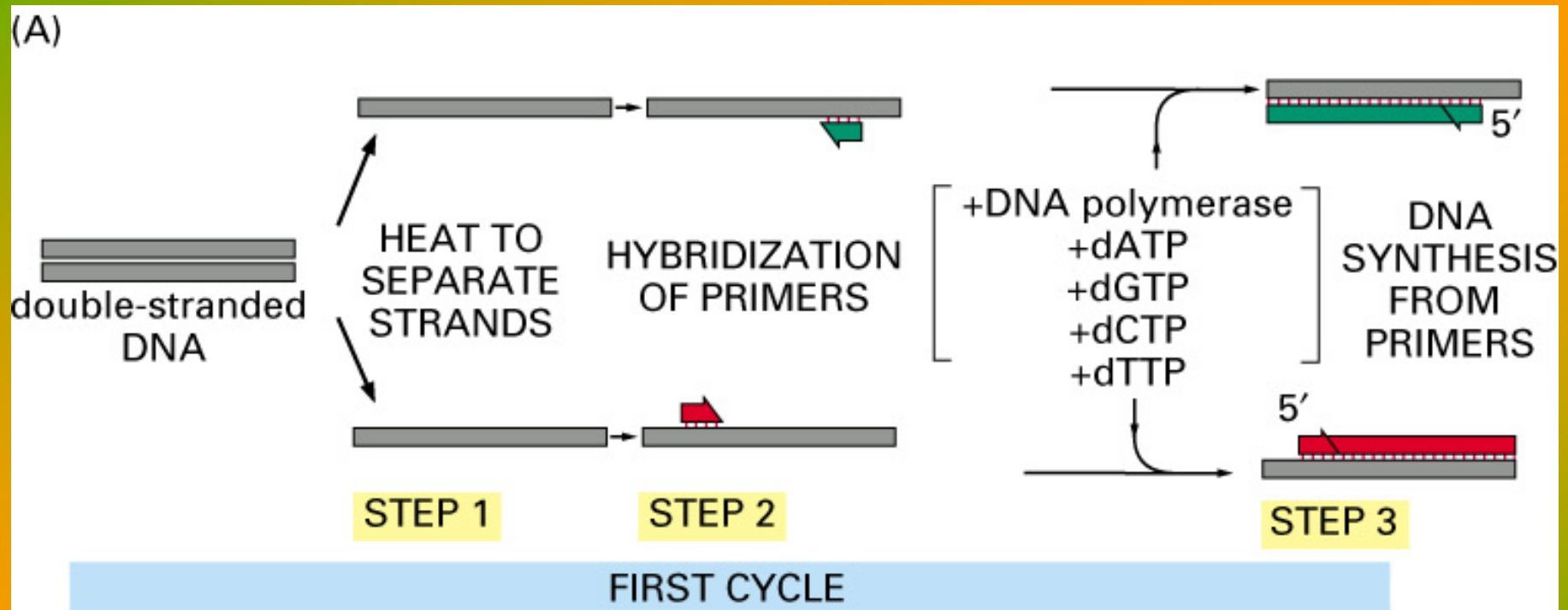


Figure 8-39 part 1 of 3. Molecular Biology of the Cell, 4th Edition.

# PCR: the second cycle

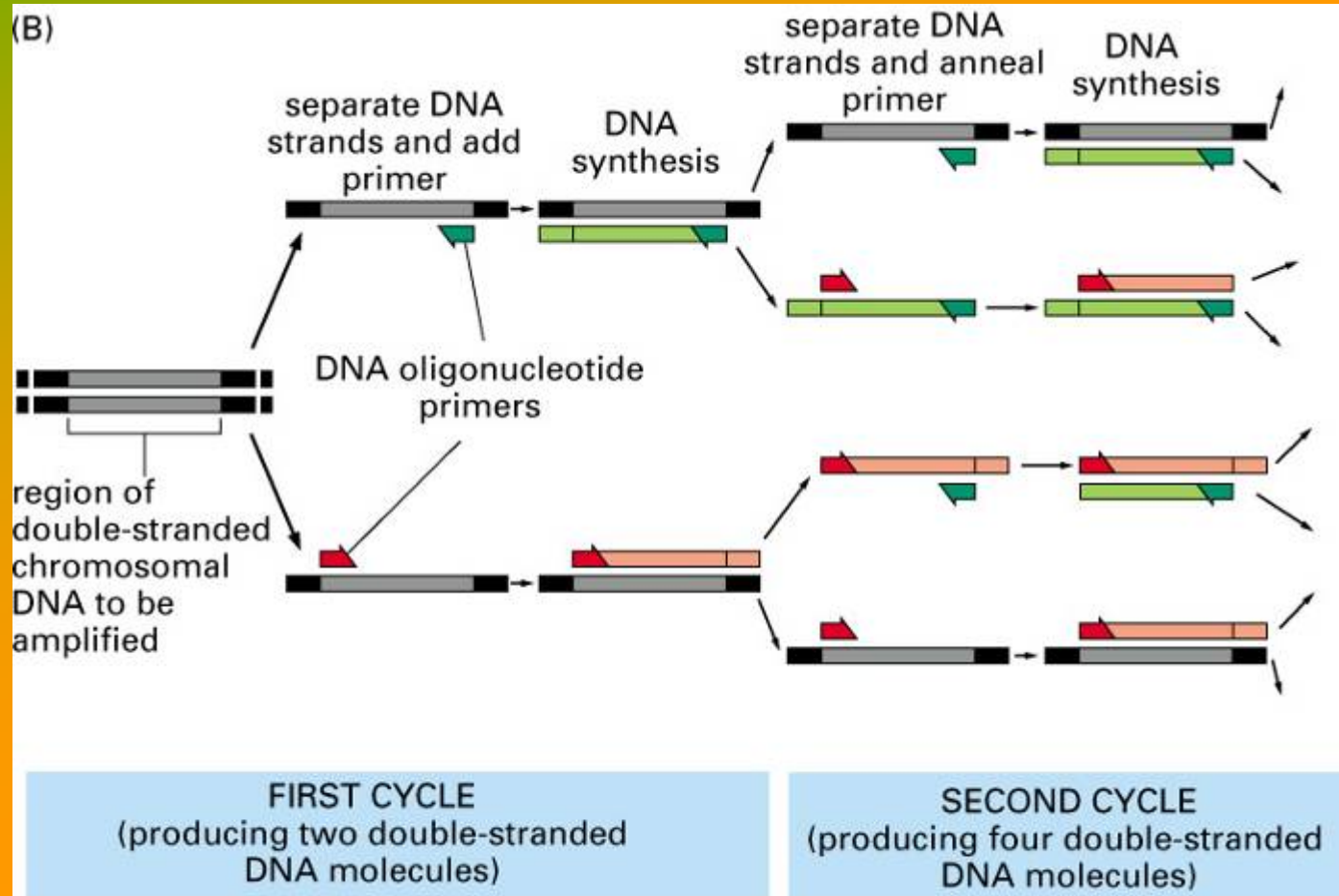


Figure 8-39 part 2 of 3. Molecular Biology of the Cell, 4th Edition.

# PCR: the third cycle

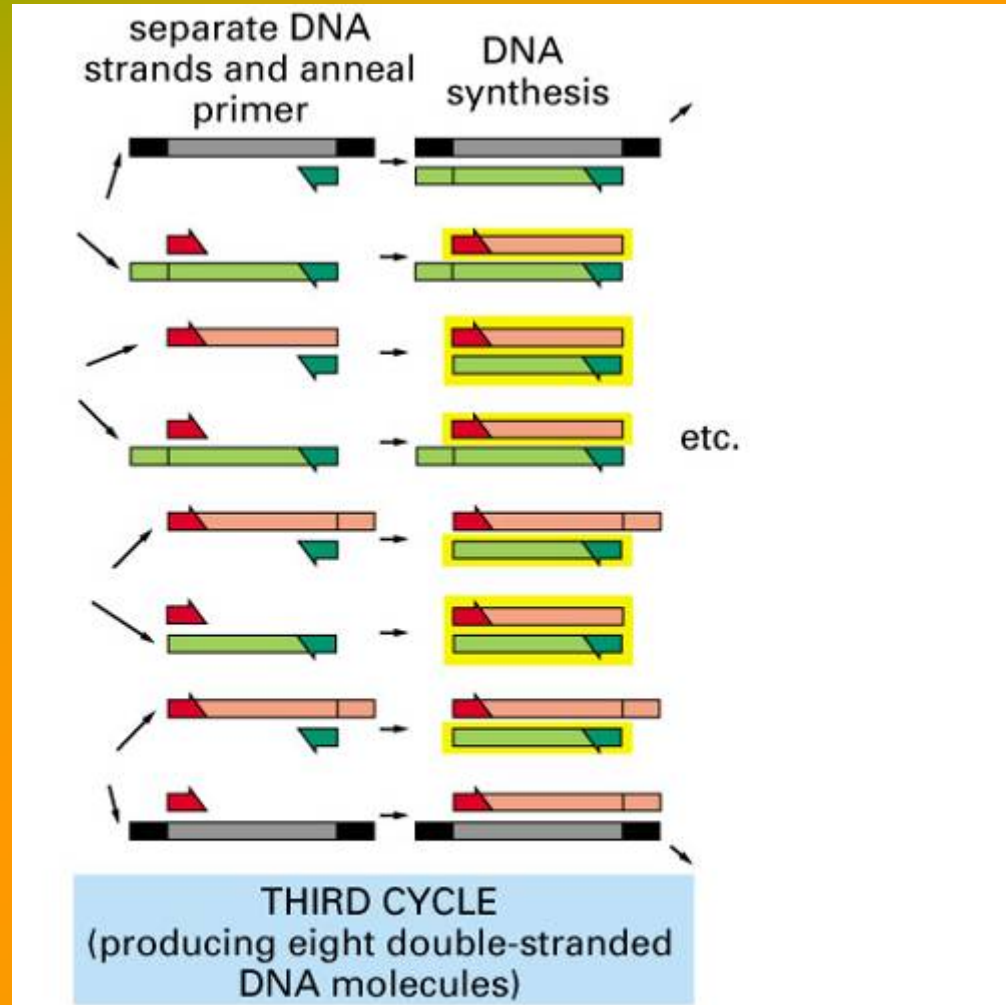
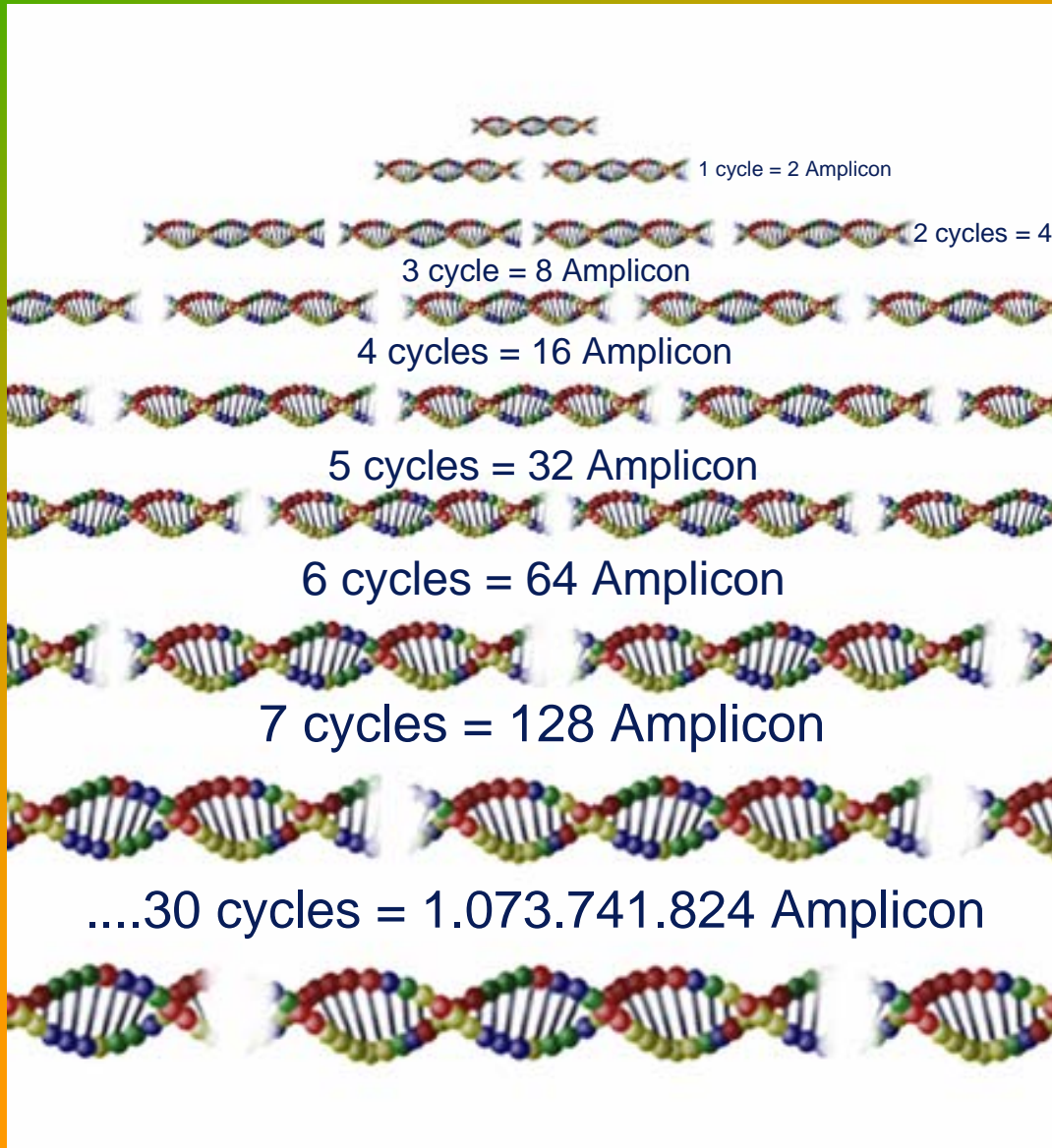


Figure 8-39 part 3 of 3. Molecular Biology of the Cell, 4th Edition.

# Target Amplification

$$2^n$$



No. Of cycles (n)	No. Of amplicon copies of the target
1	2
2	4
3	8
4	16
5	32
6	64
20	1,048,576
30	1,073,741,824

# Using PCR in diagnostics. An example: Hemochromatosis

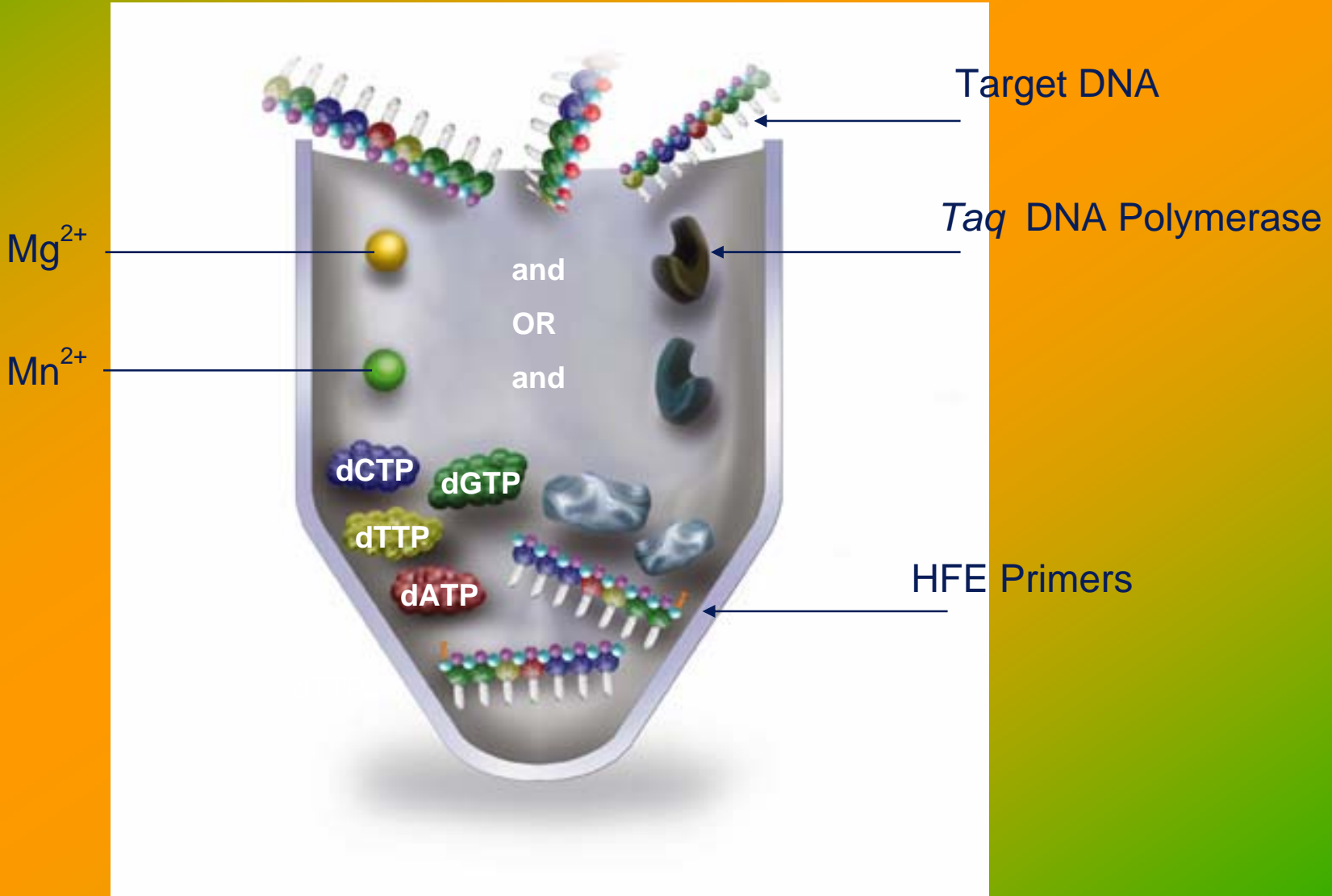
- Autosomal recessive disease of iron overload due to intestinal hyperabsorption
- Common genetic disorder: approximately 1 in 200/300 humans
- The mutation C282Y that causes iron overload is relatively frequent (1 to 5 % of the general population)
- Storage of iron in the liver and other tissues cause a number of symptoms
- The symptoms appear frequently above age of 40

# Design of primers in the HFE gene outside the C282Y mutation (expected amplicon length = 307 bp)

1 ttttctgaaa agggtatttc cttcctccaa cctatagaag gaagtgaaag ttccagtctt **Forward primer**  
61 cctggcaagg gtaaacagat cccctctcct cctccttctt ctttcctgtc aagtgcctcc  
121 tttggtgaag gtgacacatc atgtgacctc ttcagtgacc actctacggt gtcgggcctt **Site of mutation**  
181 gaactactac ccccagaaca tcaccatgaa gtggctgaag gataagcagc caatggatgc  
241 caaggagttc gaacctaaag acgtattgcc caatggggat gggacctacc agggctggat **Reverse primer**  
301 aaccttggct gtaccccctg ggaagagca gagatatacg tgcacaggtgg agcaccagg  
361 cctggatcag ccctcattg tgatctgggg tatgtgactg atgagagcca ggagctgaga  
421 aaatctattg ggggttgaga ggagtgcctg aggaggtaat tatggcagtg agatgaggat  
481 ctgctctttg ttaggggggtg ggccgagggt ggcaatcaaa ggctttaact tgctttttct  
541 gtttttagagc cctcaccgtc tggcacccta gtcattggag tcatcagtg aattgctgtt  
601 tttgtcgtca tcttgttcat tgaattttg ttcataatat taaggaagag gcagggttca  
661 agtgagtagg aacaaggggg aagtctctta gtacctctgc cccagggcac agtggggaaga  
721 ggggcagagg gga

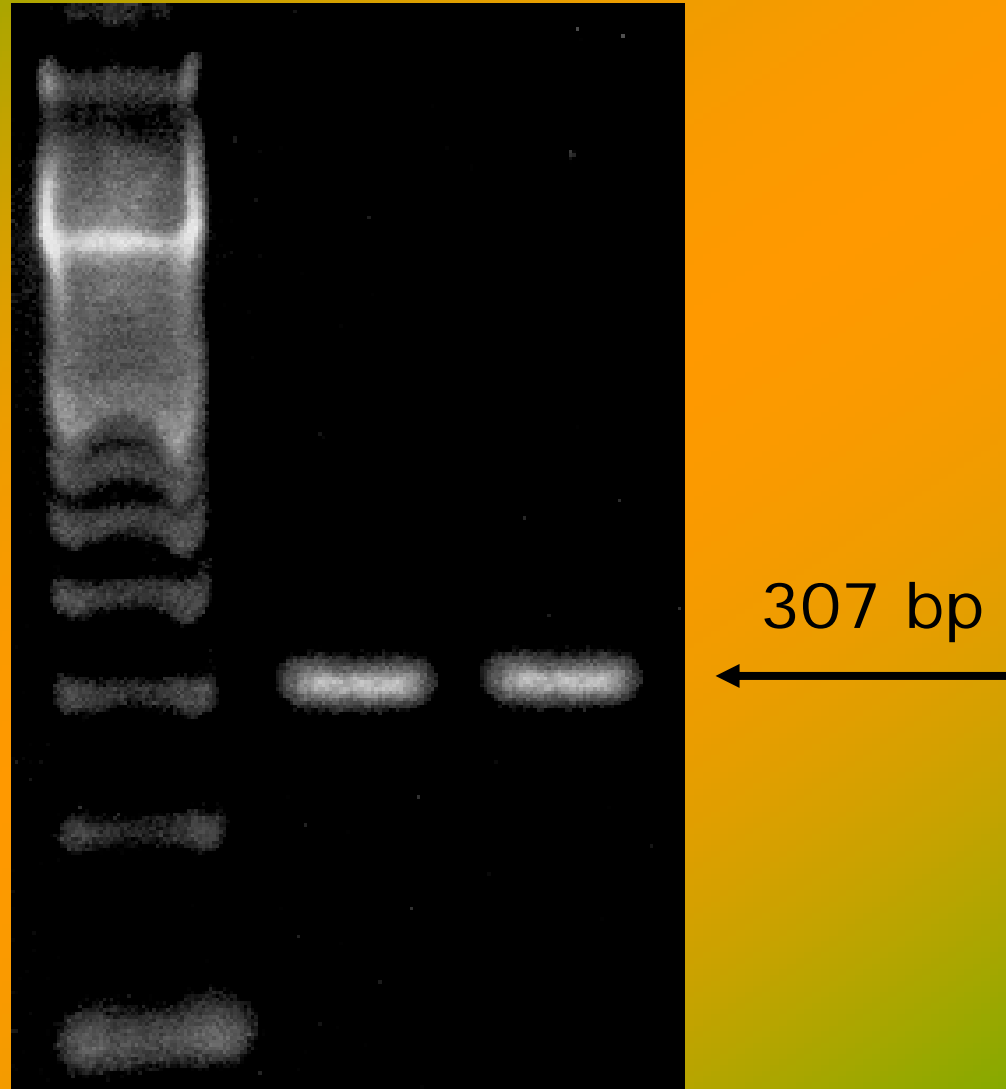
TGC = Cys  $\xrightarrow{\text{C282Y}}$  TAC = Tyr

# PCR for the amplification of the HFE gene

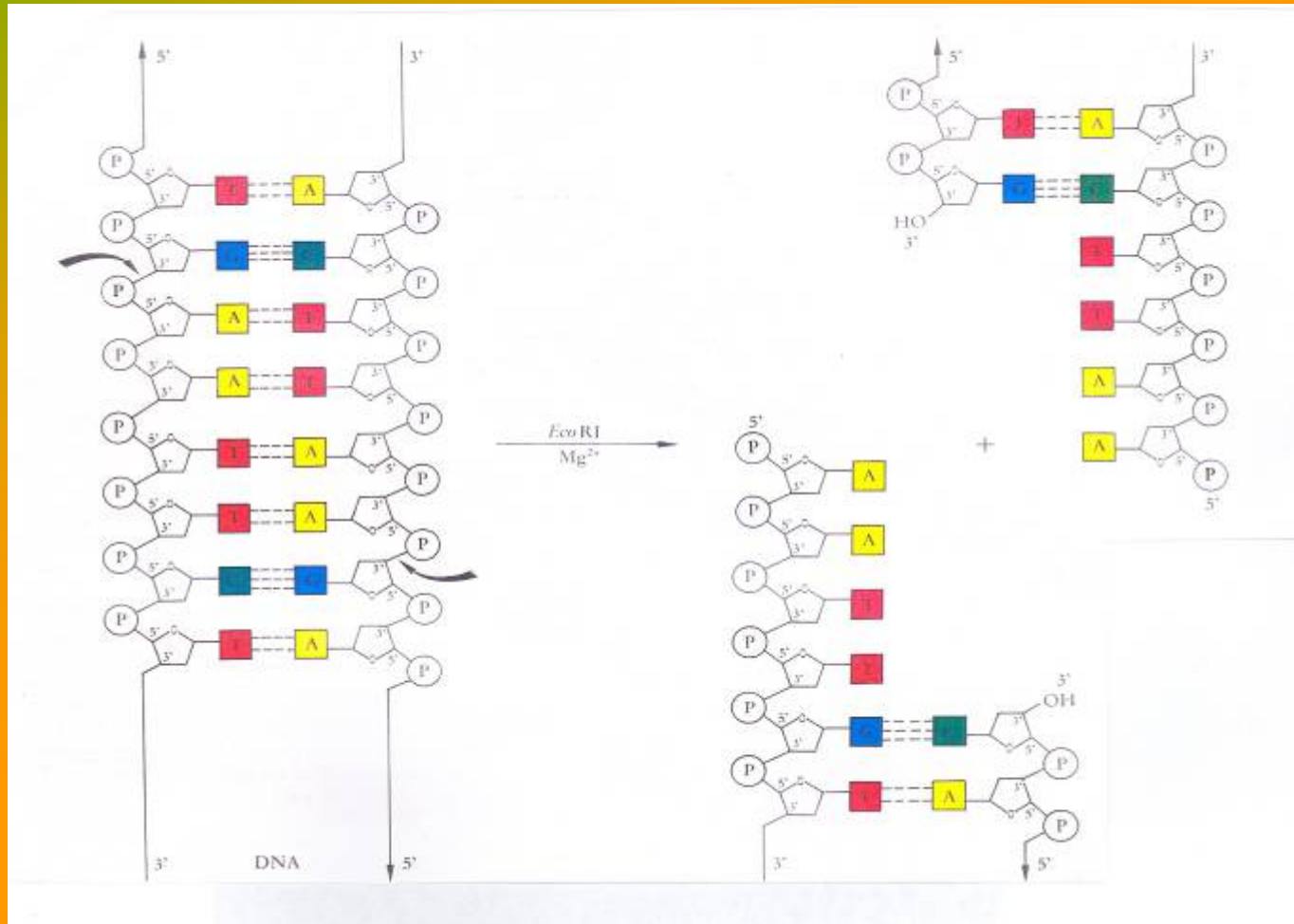




And expected PCR product



# Find a restriction enzyme that cuts the mutated base (C282Y)



# Restriction sites in the HFE gene amplicon

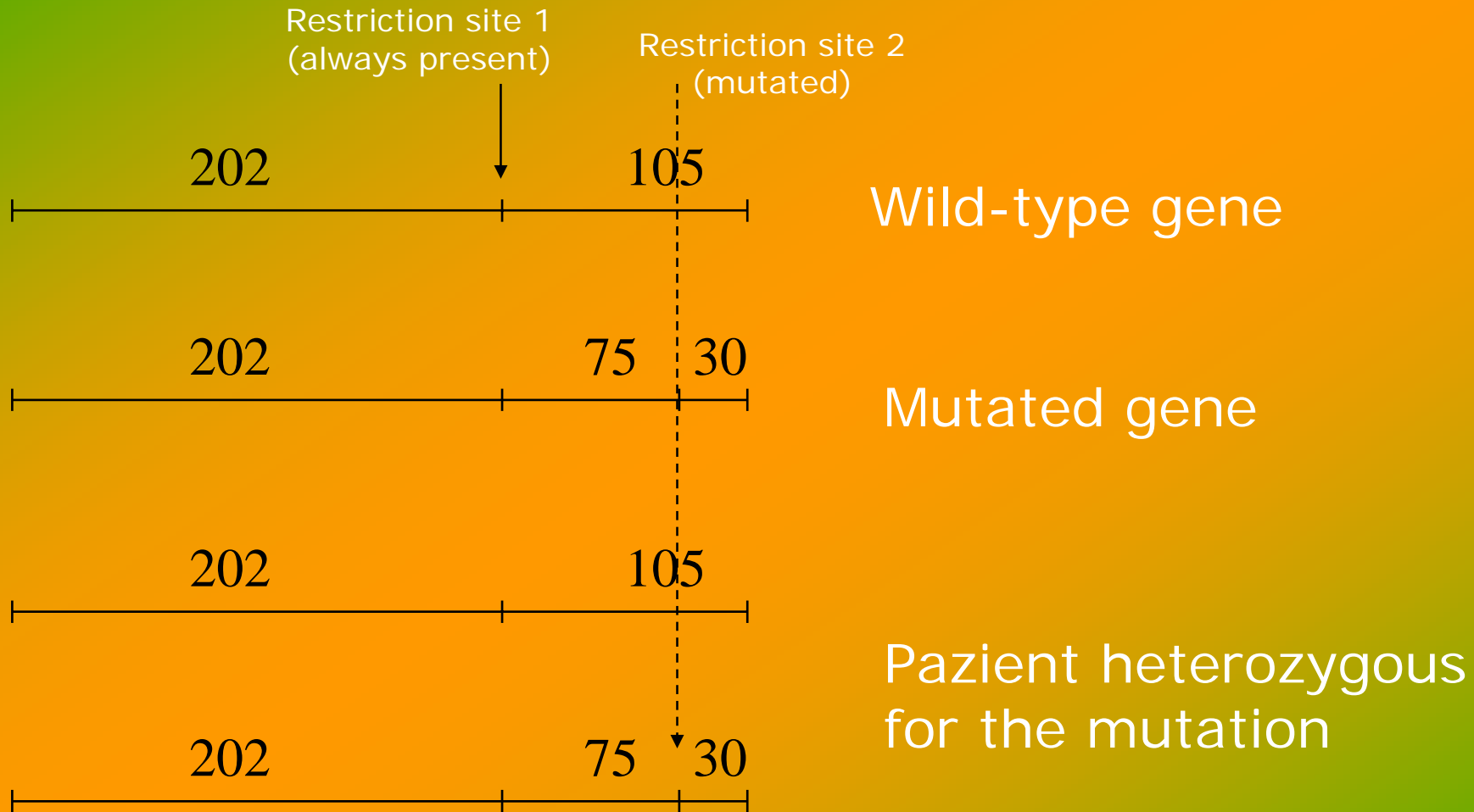
1 ttttctgaaa agggatatttc cttcctccaa cctatagaag gaagtgaaag ttccagtett  
61 cctggcaagg gtaaacagat ccctctcct catccttctt ctttctgttc aagtgcctcc  
121 tttggtgaag gtgacacatc atgtgacctc ttcagtgaac actctacggt gtcgggcctt  
181 gaactactac ccccagaaca tcaccatgaa gtggctgaag gataagcagc caatggatgc  
241 caaggagttc gaacctaaag acgtattgcc caatggggat gggacctacc agggctggat  
301 aaccttggct **gtac**cccctg gggaagagca gagatatacg **tgc**caggtgg agcaccagg  
361 cctggatcag ccctcattg tgatctgggg tatgtgactg atgagagcca ggagctgaga  
421 aaatctattg gggggtgaga ggagtgcctg aggaggtaat tatggcagtg agatgaggat  
481 ctgctctttg ttaggggggtg ggccgagggg ggcaatcaaa ggctttaact tgctttttct  
541 gtttttagagc cctcaccgtc tggcacccta gtcattggag tcatcagtggt aattgctggt  
601 tttgtcgtca tcttgttcat tggaaatttg ttcataatat taaggaagag gcagggttca  
661 agtgagtagg aacaaggggg aagtctctta gtacctctgc cccagggcac agtggggaaga  
721 ggggcagagg gga

Restriction site 1  
(always present)

Restriction site 2  
(present only if  
mutated)

TGC = Cys  $\xrightarrow{\text{C282Y}}$  TAC = Tyr

# Enzymatic restriction (Rsa I) of the PCR product



# Results

