

SYLICA 2013
Bowater lectures

**Contemporary DNA
Sequencing Technologies**

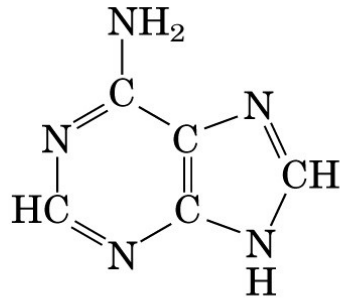
Bowater Lectures in Brno, Feb. 2013

4 lectures on linked topics will be delivered during the coming week:

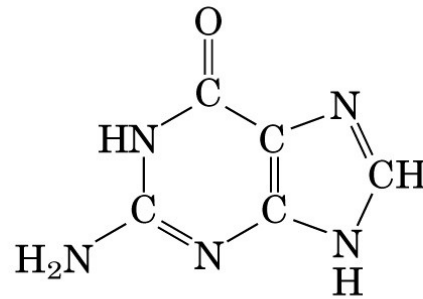
- *Contemporary DNA Sequencing Technologies – 26/2/2013 @ 10:00*
- *Using ‘Omic Technologies to Investigate Gene Function – 26/2/2013 @ 14:00*
- *Biophysical Methods to Study Molecular Interactions – 27/2/2013 @ 10:00*
- *Synthetic Biology & Nanotechnology: Tomorrow’s Molecular Biology? – 28/2/2013 @ 10:00*

DNA STRUCTURE

Bases

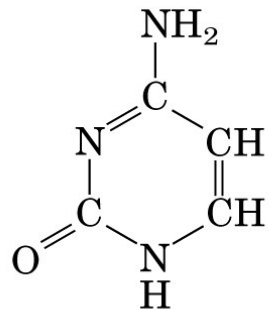


Adenine

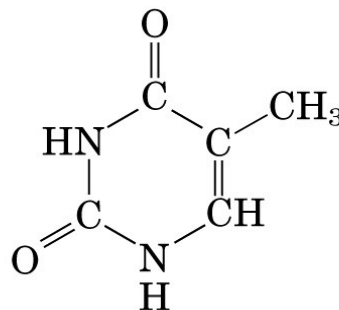


Guanine

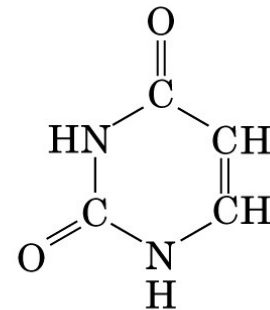
Purines



Cytosine



Thymine
(DNA)



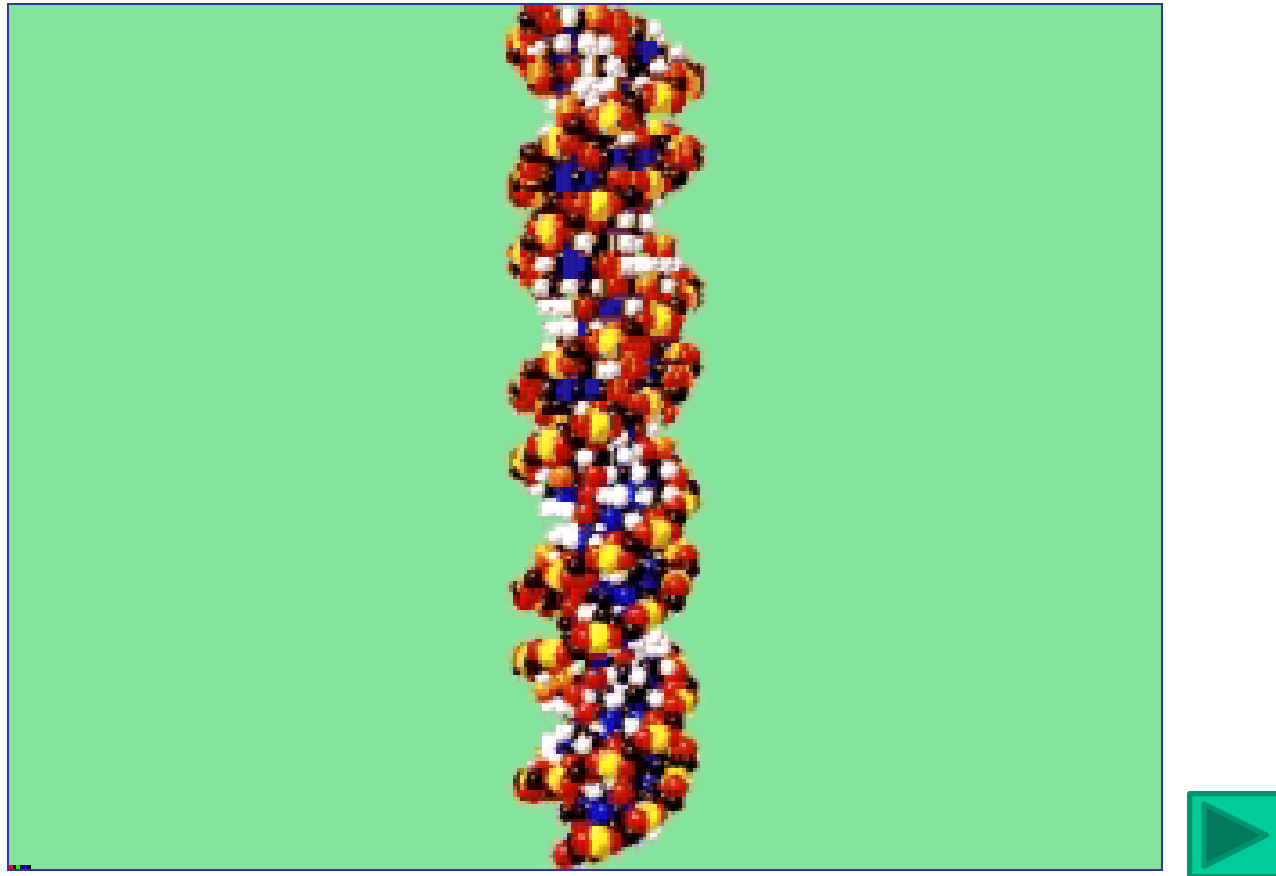
Uracil
(RNA)

Pyrimidines

Nelson & Cox, "Lehninger, Principles of Biochemistry", 5th edn, 2008, p. 272

DNA

DNA is a double-stranded, helical molecule



Alternative DNA Helices

- The DNA helix may take different forms:

A- and B-helices are

right-handed.

Z-DNA

hande.

The typical (average) structure of a DNA molecule is in the B-form.

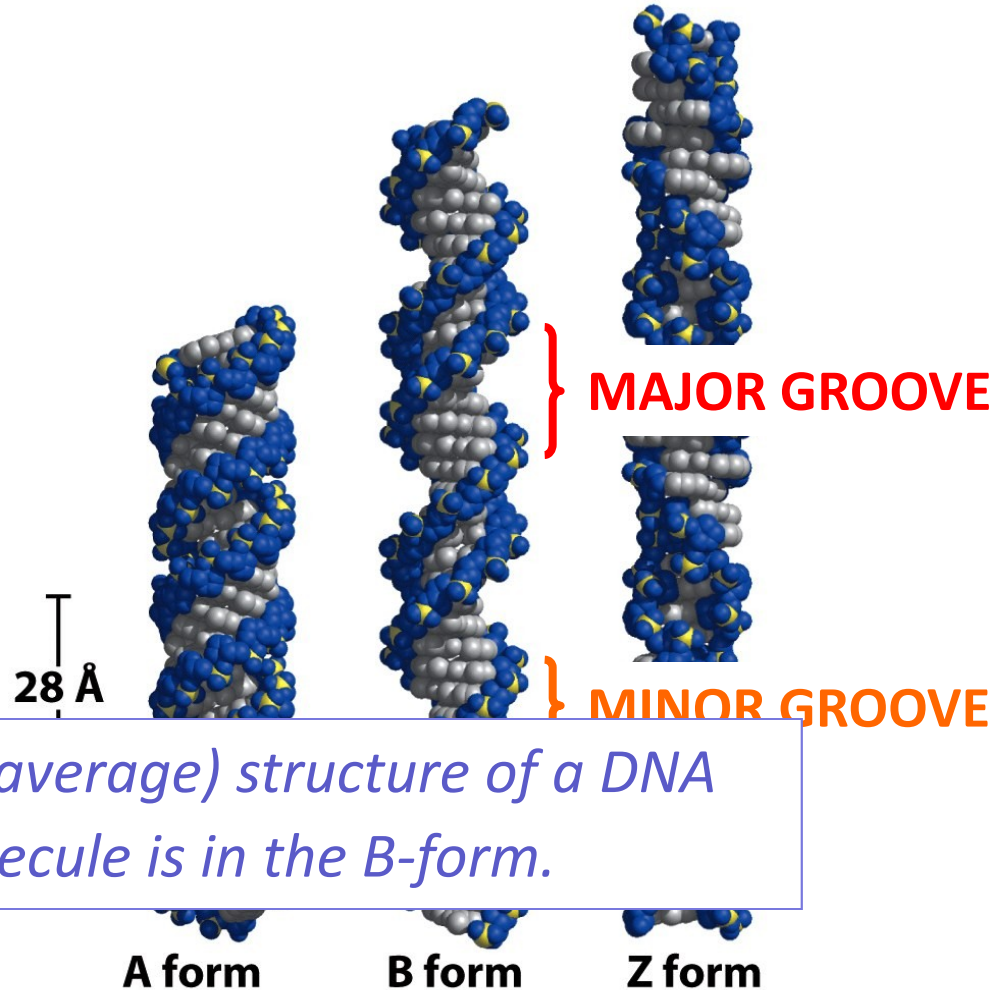
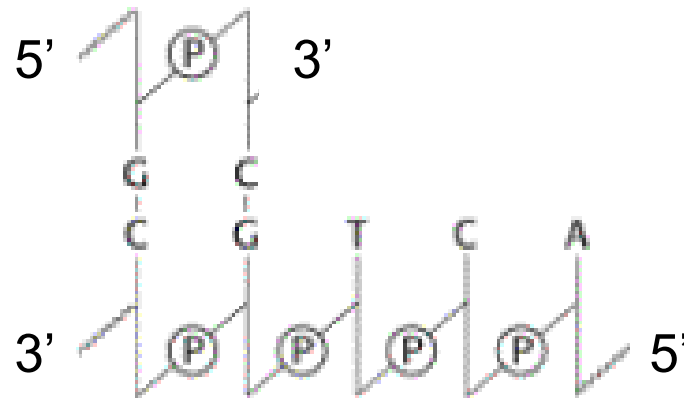
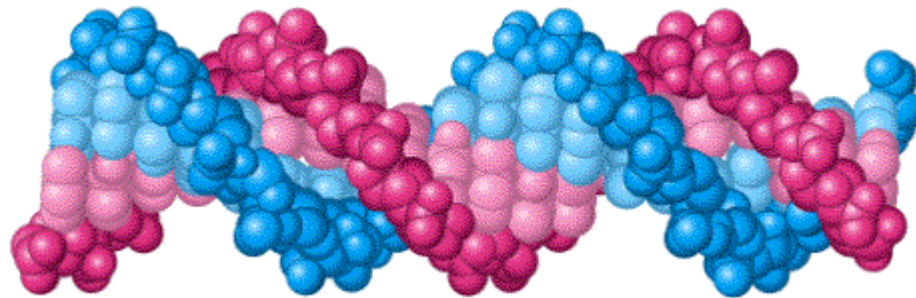


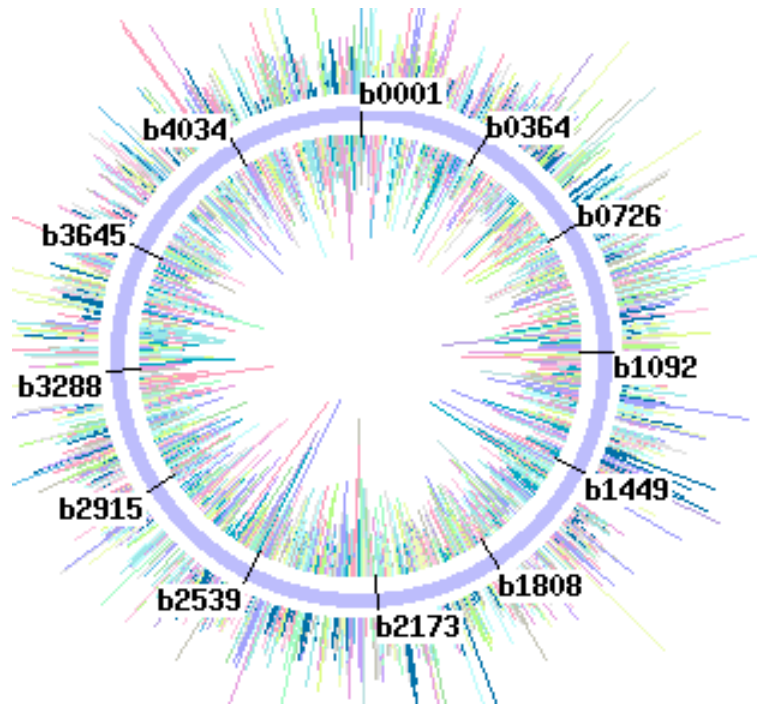
Figure 8-17 part 1
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W.H. Freeman and Company

dsDNA has an Anti-Parallel Structure

The two strands of dsDNA have an anti-parallel polarity

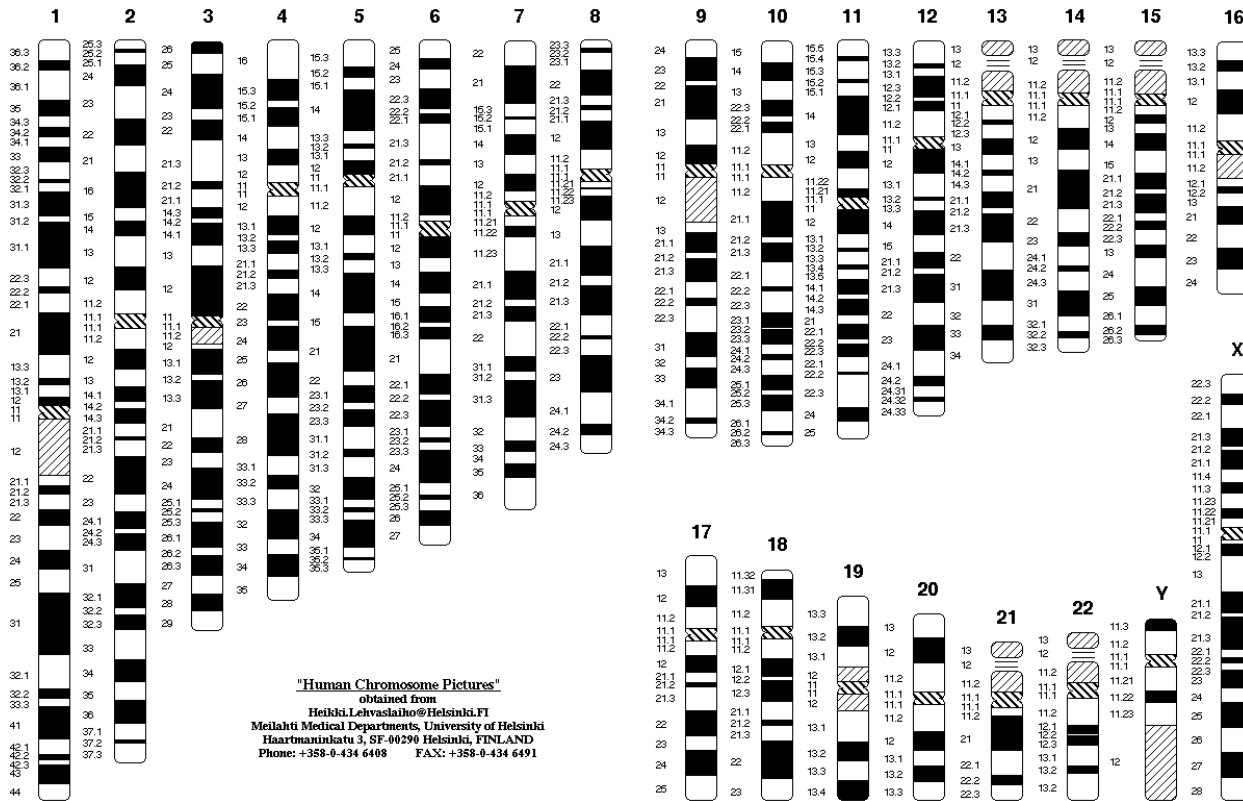


Genomes



Genome of *E. coli* codes for 4,500 genes in 4.6 Mbp

Genomes



Human genome codes for ~30,000 genes in 3,000 Mbp

Cellular DNA

- “Coding” DNA is name given to genes that are transcribed & translated to make protein
- Eukaryotic genomes contain large amounts of *non-coding* DNA
- The length of DNA inside cells is extremely large relative to cell size

The Length of DNA in Cells is Very Large!

- length of DNA inside cells is extremely large relative to cell size

E. coli – lysed to show chromosomal DNA

For related discussion, see also: Nelson & Cox, “Lehninger, Principles of Biochemistry”, 5th edn, 2008, p. 950



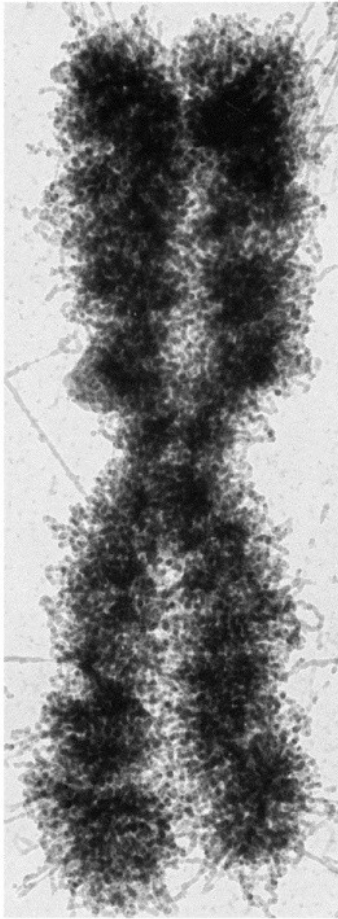
Coding & Non-coding DNA

| Organism | No. of genes | Total size of DNA (Mbp) | % of genome as coding DNA* |
|--------------------------------|--------------|-------------------------|----------------------------|
| <i>E. coli</i> | 4,500 | 4.6 | 98 |
| Yeast (<i>S. cerevisiae</i>) | 5,885 | 12 | 49 |
| Human | 30,000 (?) | 3,000 | 1 |

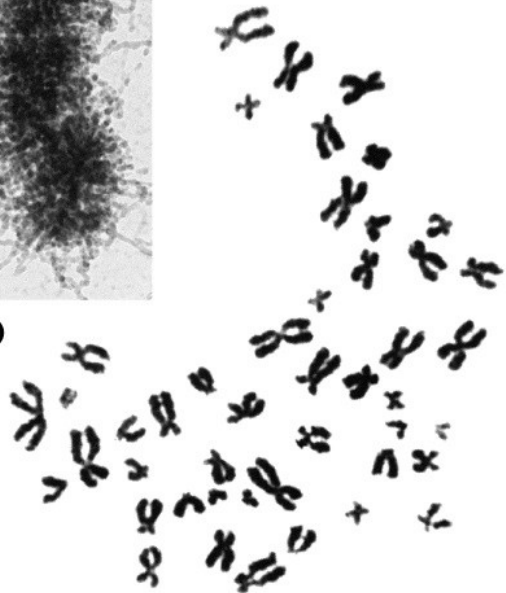
* Assuming average gene size ~ 1,000 bp

Chromosomal DNA

- *chromosomes* are complexes of protein & DNA



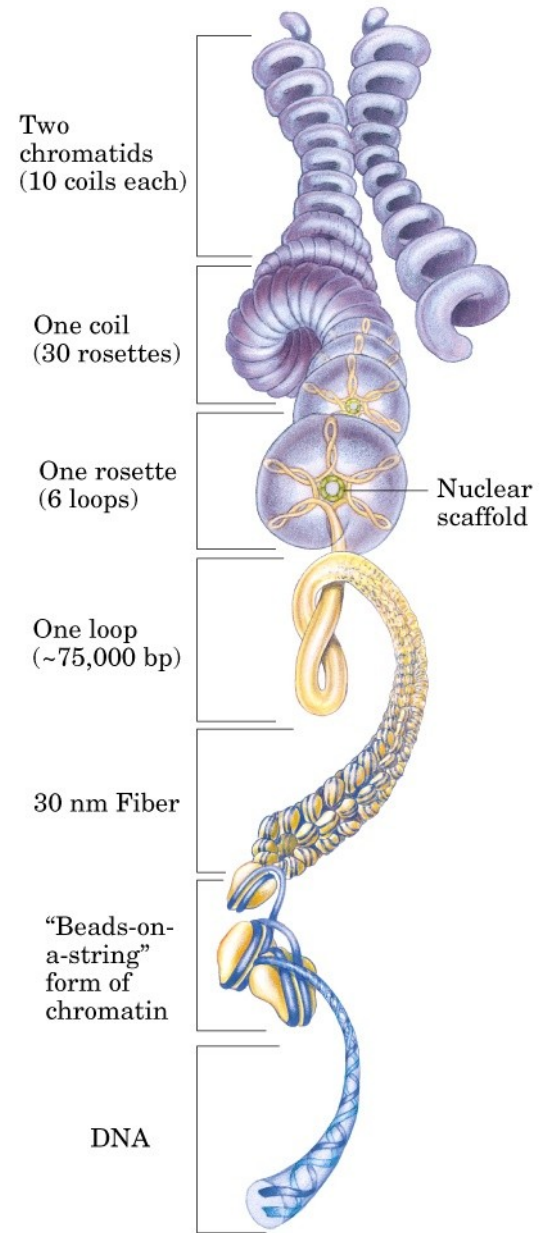
(a)



(b)

Nelson & Cox, "Lehninger, Principles of Biochemistry",
5th edn, 2008, p. 951

DNA Compaction within Cells



Nelson & Cox, "Lehninger, Principles of Biochemistry",
5th edn, 2008, p. 968

DNA Structure: Overview

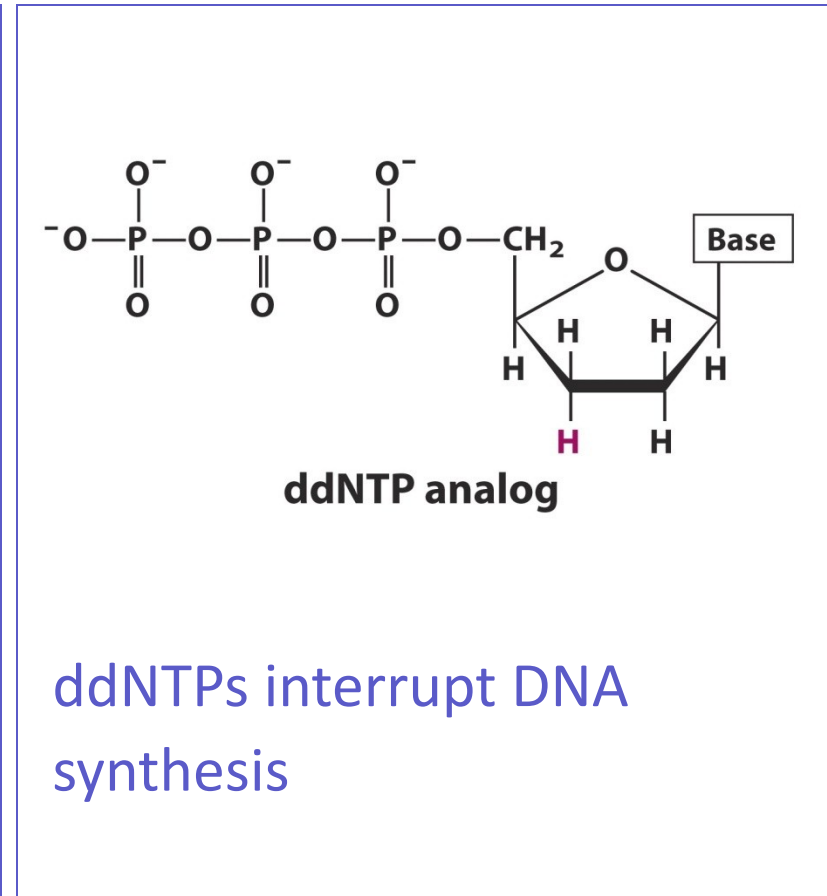
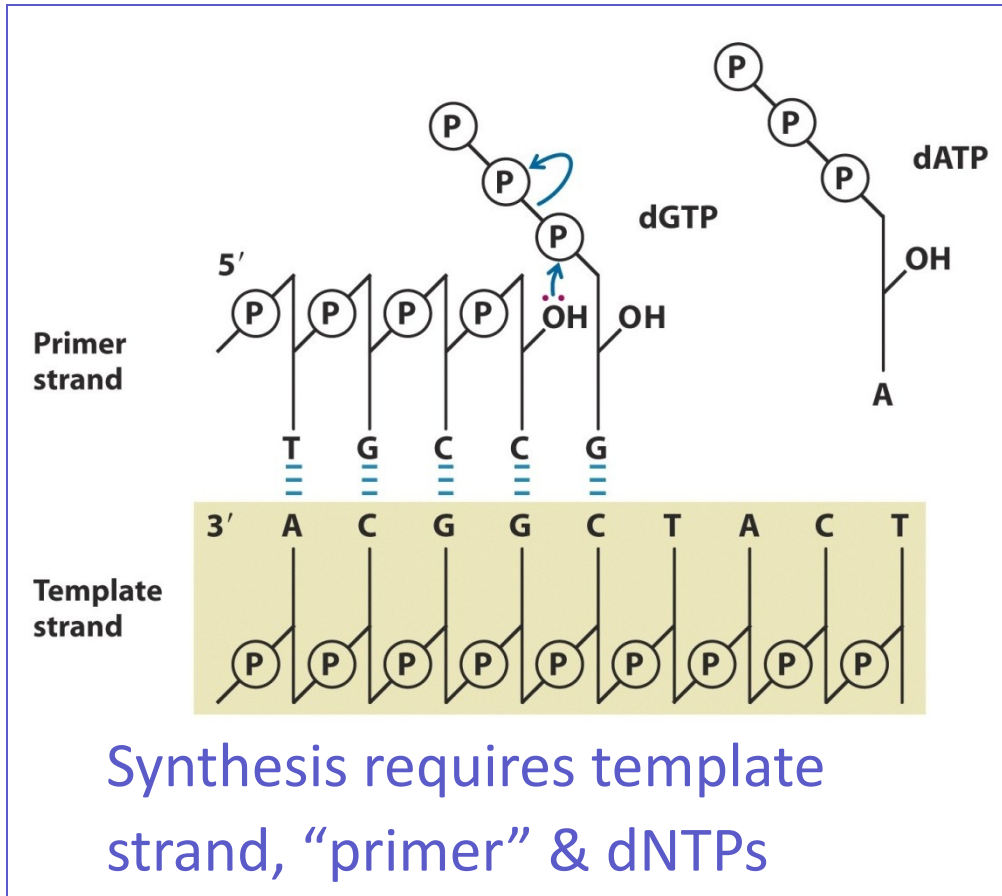
- Inside cells, the structure of DNA is *dynamic*: usually in B-form helix, but can exist in different structures/conformations
- DNA molecules can be linear or circular
- Most genomes have significant amounts of non-coding DNA
- DNA molecules in cells are *very long* – therefore the helix has further levels of well-organised structure that allow it to be contained and used in cells

Genomics & Technology

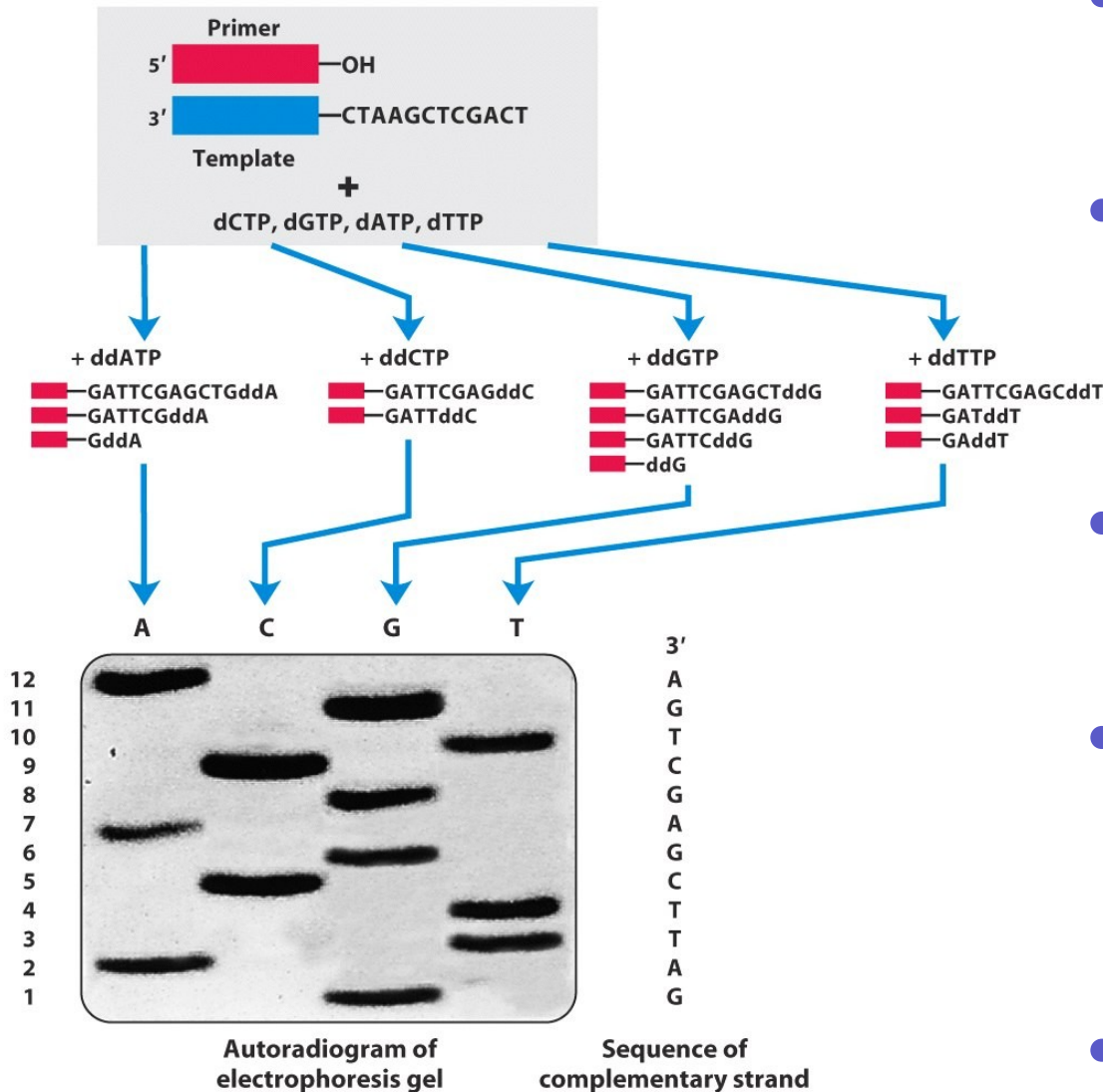
- Molecular biology: major scientific discipline for past ~50 years
- Genomics: became important science during 1990's
- Transcriptomics/Proteomics: developed during past 5 years
- Bioinformatics: has developed as major branch of science - enables efficient analysis of data from "omics" experiments

A Primer About DNA Sequencing

- Major advance in DNA sequencing occurred with use of DNA polymerases



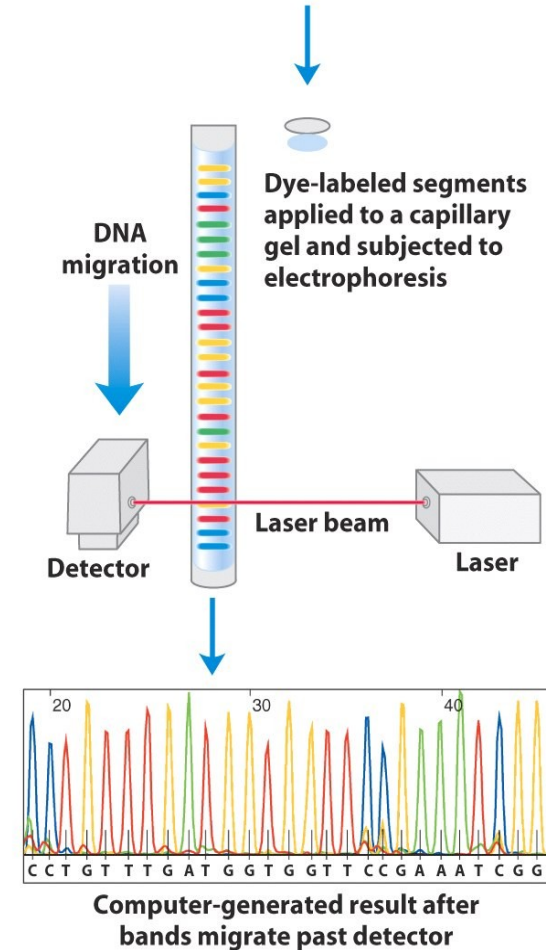
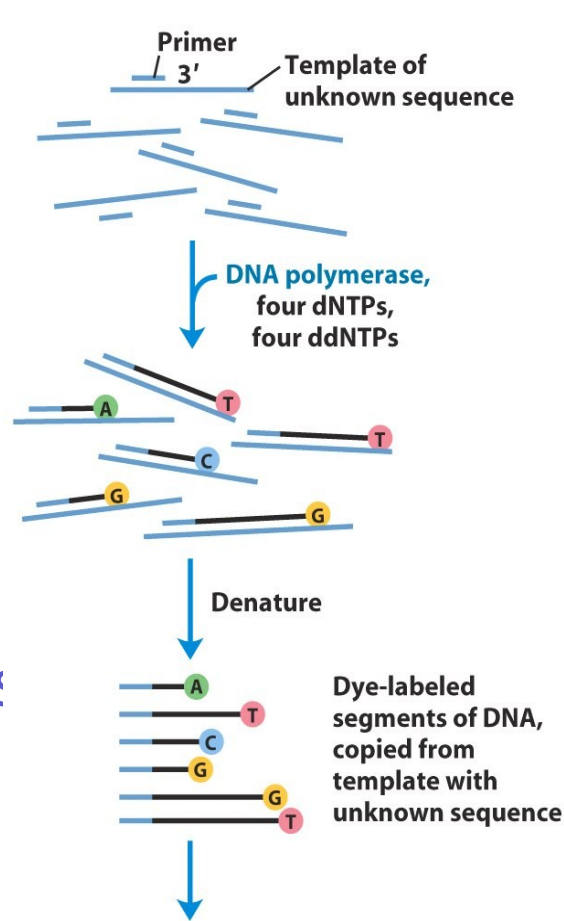
A Primer About DNA Sequencing



- DNA to be sequenced acts as template
- Oligonucleotide allows sequencing to start at any point
- Small amounts of “labelled” ddNTP
- Identification of specific bases after electrophoresis
- *<500 bases per day*

Improved DNA Sequencing

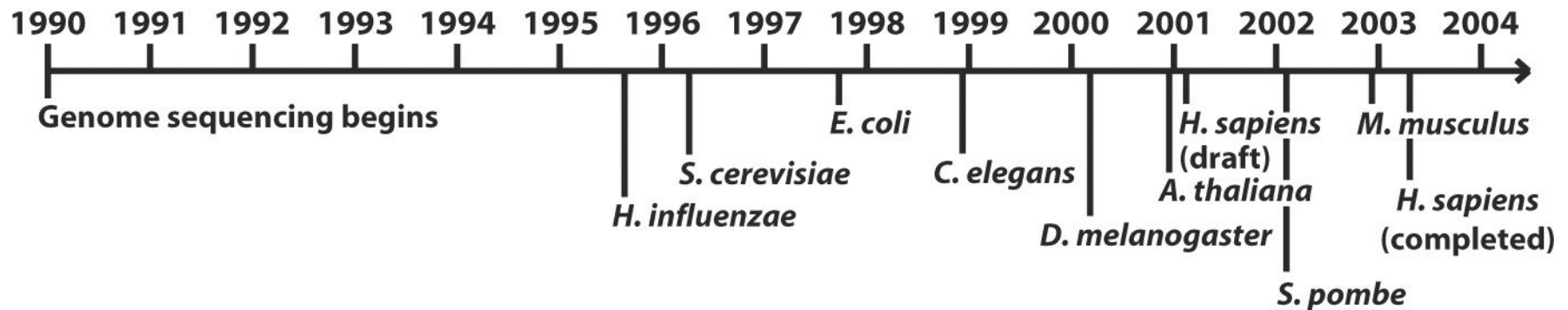
- “Labelled” DNAs separated by capillary electrophoresis
- DNA sequence read as series of colours
- Computer deciphers sequence
- *>2,000 bases per day*



Nelson & Cox, “Lehninger, Principles of Biochemistry”, 4th edn, 2004, p. 298

Genomic Sequencing

- Sequencing centres have hundreds of machines working continuously
- Each can generate equivalent of human genome sequence each month



Nelson & Cox, "Lehninger, Principles of Biochemistry", 4th edn, 2004, p. 324

Genomic Sequencing



Francis S. Collins



J. Craig Venter

Unnumbered 9 p322

Lehninger Principles of Biochemistry, Fifth Edition

© 2008 W.H. Freeman and Company

The Human Genome Project

- Sequencing of the human genome allows for:
 - Identification and categorization of different haplotypes
 - Understanding the differences between humans and chimpanzees
 - Based on phylogenetic trees and comparison of differences
 - Especially in regulatory sequences, which may be more important to evolution than protein changes
 - Identification of genes involved in disease
 - Track the path of human migration

Human genome contains many different sequence types

Human genome: DNA sequence types

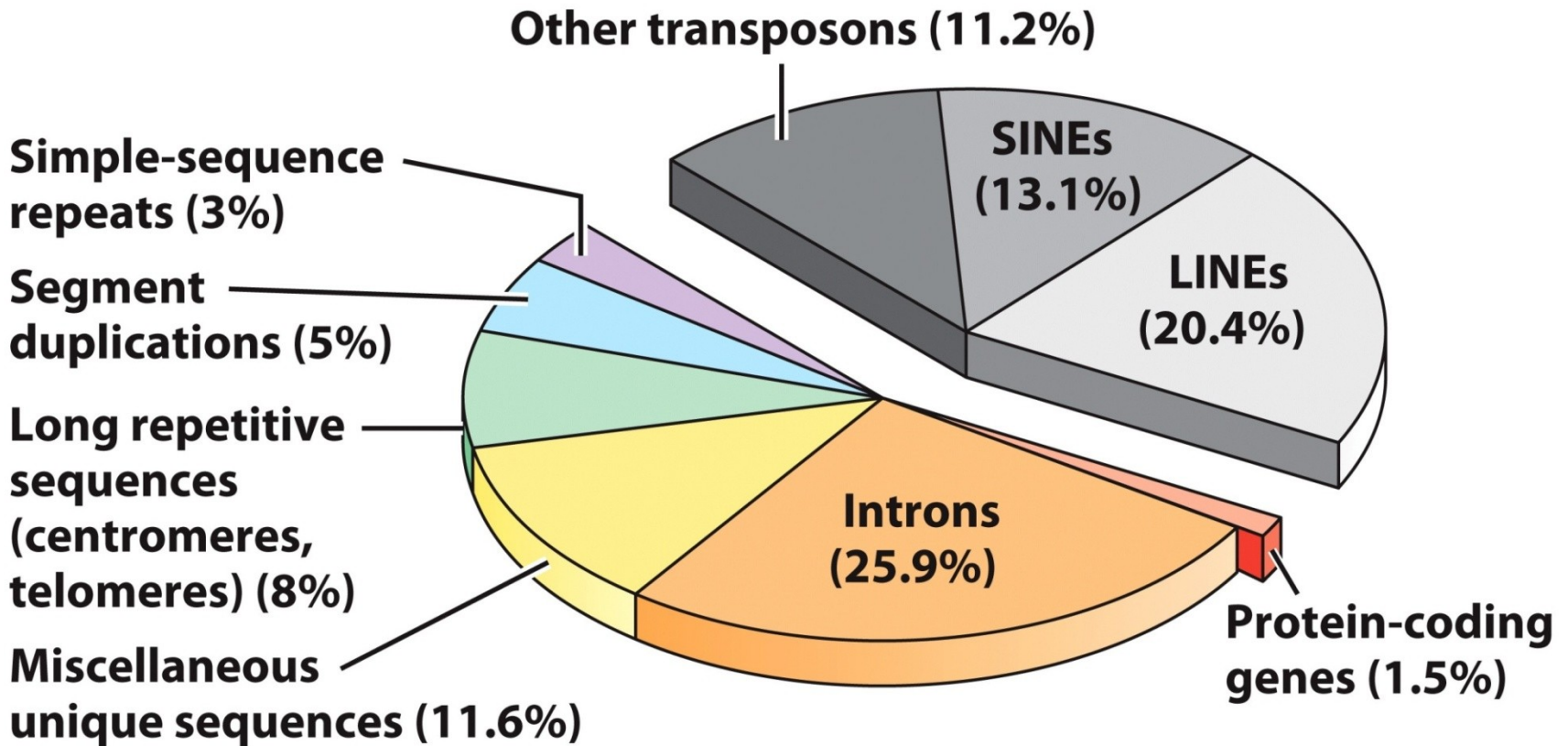


Figure 9-29a

Lehninger Principles of Biochemistry, Sixth Edition

© 2013 W. H. Freeman and Company

Human genome contains many different protein types

Human genome: Protein-coding genes

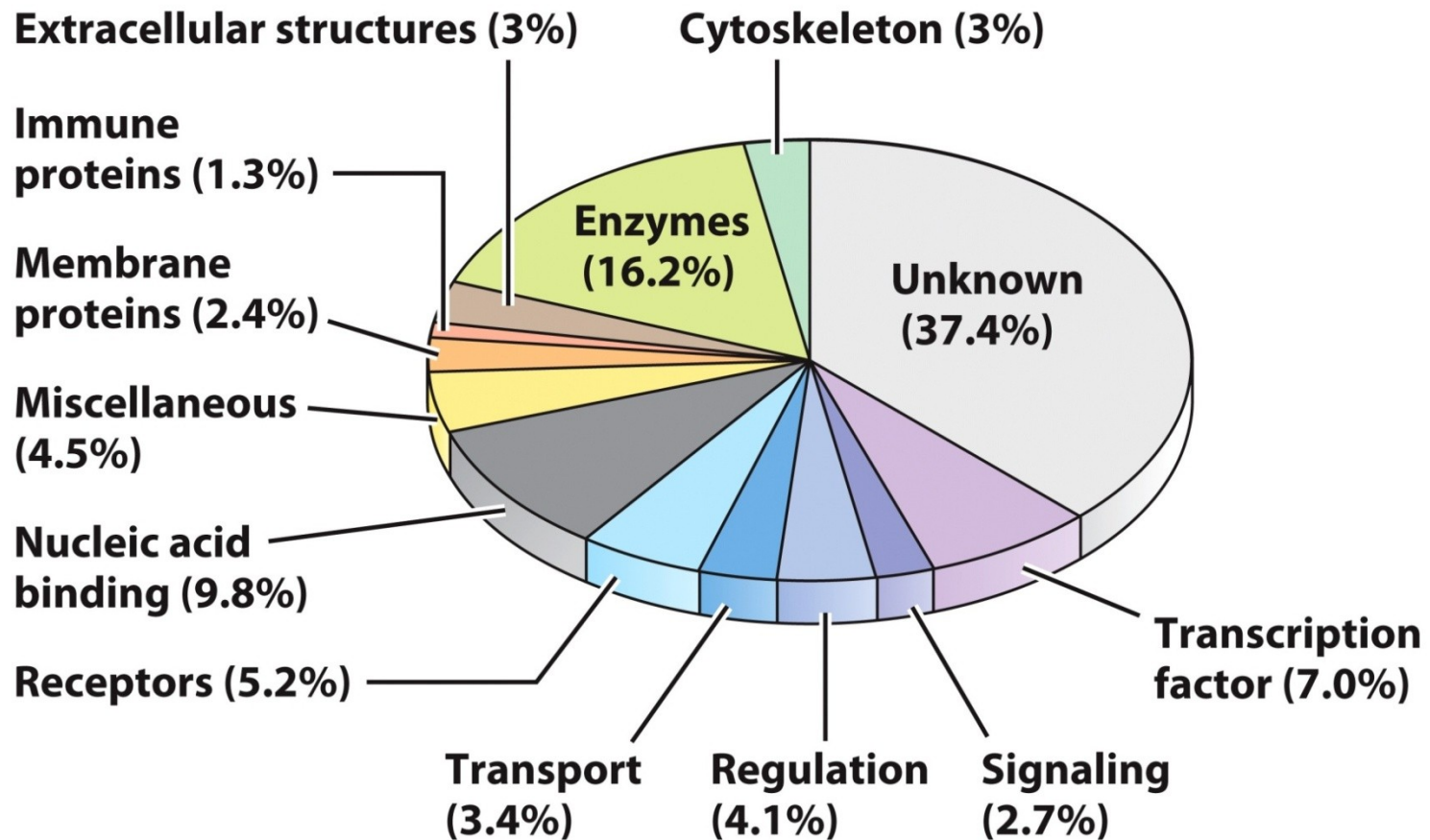


Figure 9-29b
Lehninger Principles of Biochemistry, Sixth Edition
© 2013 W. H. Freeman and Company

New Generation of DNA Sequence Analysis

- Full genome is immobilized on a chip in fragments a few hundred bases long
 - All sequenced at once, allowing for faster detection
- Pyrosequencing
 - DNA synthesized from the template a single nucleotide at a time, each generating a pulse of light
 - Can read 400–500 nucleotides in the sequence
- Reversible terminator sequencing
 - Fluorescently labeled terminal nucleotide is added to the sequence and detected
 - Terminal nucleotide is removed, sequence extended, and next nucleotide is detected

Pyrosequencing

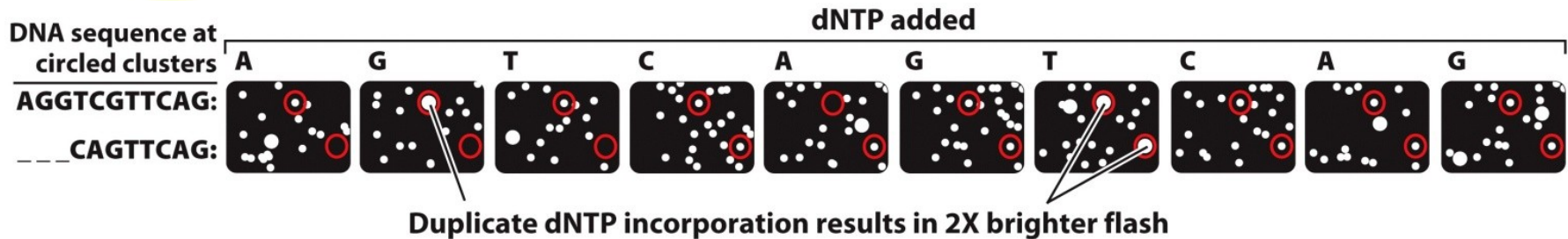
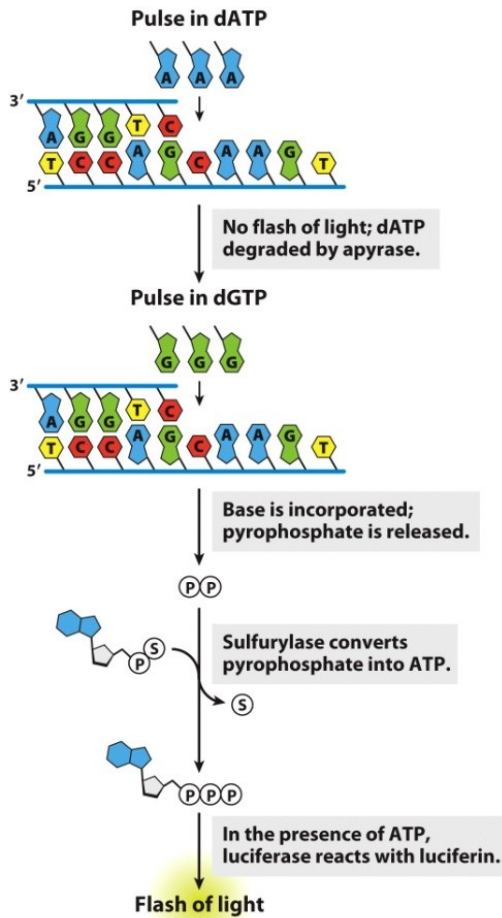
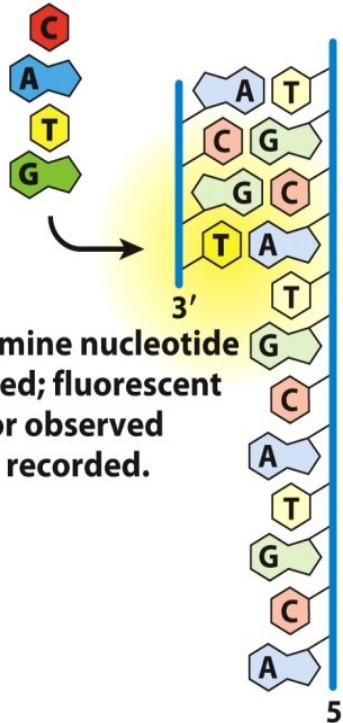


Figure 9-25b
 Lehninger Principles of Biochemistry, Sixth Edition
 © 2013 W. H. Freeman and Company

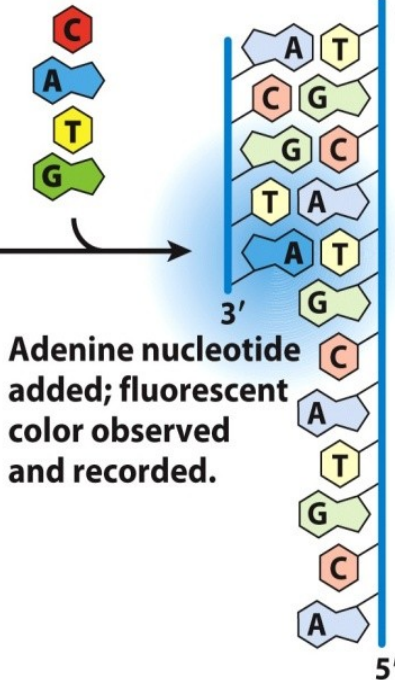
Reversible Terminator Sequencing

(a)

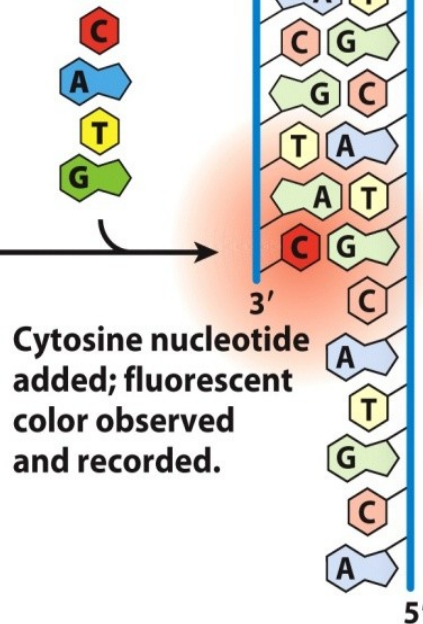
Add blocked, fluorescently labeled nucleotides.



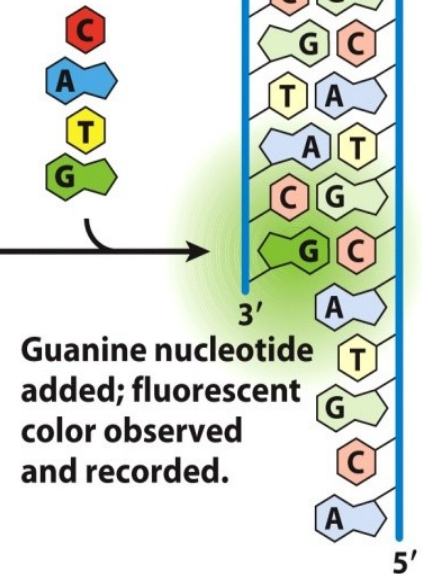
Remove labels and blocking groups; wash; add blocked, labeled nucleotides.



Remove labels and blocking groups; wash; add blocked, labeled nucleotides.



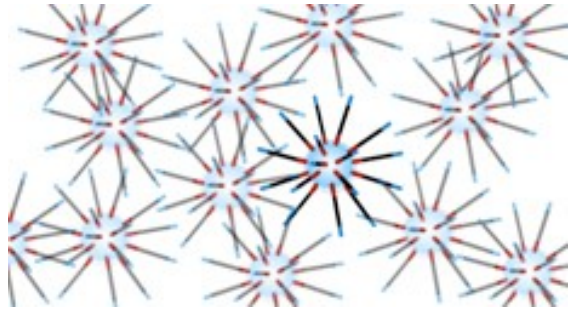
Remove labels and blocking groups; wash; add blocked, labeled nucleotides.



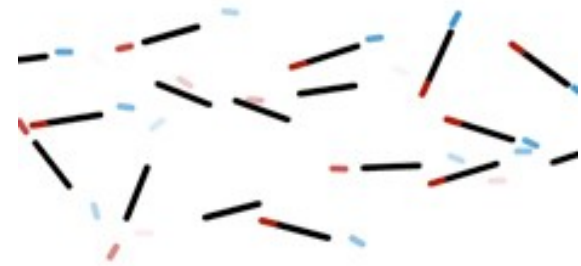
High-throughput Sequencing Technologies

- Recently, several new technologies have increased the throughput and reduced the cost for genome sequencing
- Examples are:

454 Sequencing



Illumina method



- Animations illustrating these methods available at:
- www.wellcome.ac.uk/Education-resources/Teaching-and-education/Animations/DNA/index.htm

DNA Sequencing: Overview

- Production of pure DNA polymerases made it feasible to consider sequencing of genomes
- Sequencing of large genomes became possible with:
 - Increased sensitivity of nucleic acid detection
 - Automated robotic technology
 - Improved computer power
- Further advances are increasing speed, reducing size and cost – suggesting it will soon be possible to sequence individual human genomes for \$1,000

Polymerase Chain Reaction (PCR)

- Used to amplify DNA in the test tube
 - Can amplify regions of interest (genes) within DNA
 - Can amplify complete circular plasmids
- Mix together
 - Target DNA
 - Primers (oligonucleotides complementary to target)
 - Nucleotides: dATP, dCTP, dGTP, dTTP
 - Thermostable DNA polymerase
- Place the mixture into thermocycler
 - Melt DNA at $\sim 95^{\circ}\text{C}$
 - Cool to $\sim 50\text{--}60^{\circ}\text{C}$, primers anneal to target
 - Polymerase extends primers in $5' \rightarrow 3'$ direction
 - After a round of elongation is done, repeat steps

General Steps of PCR

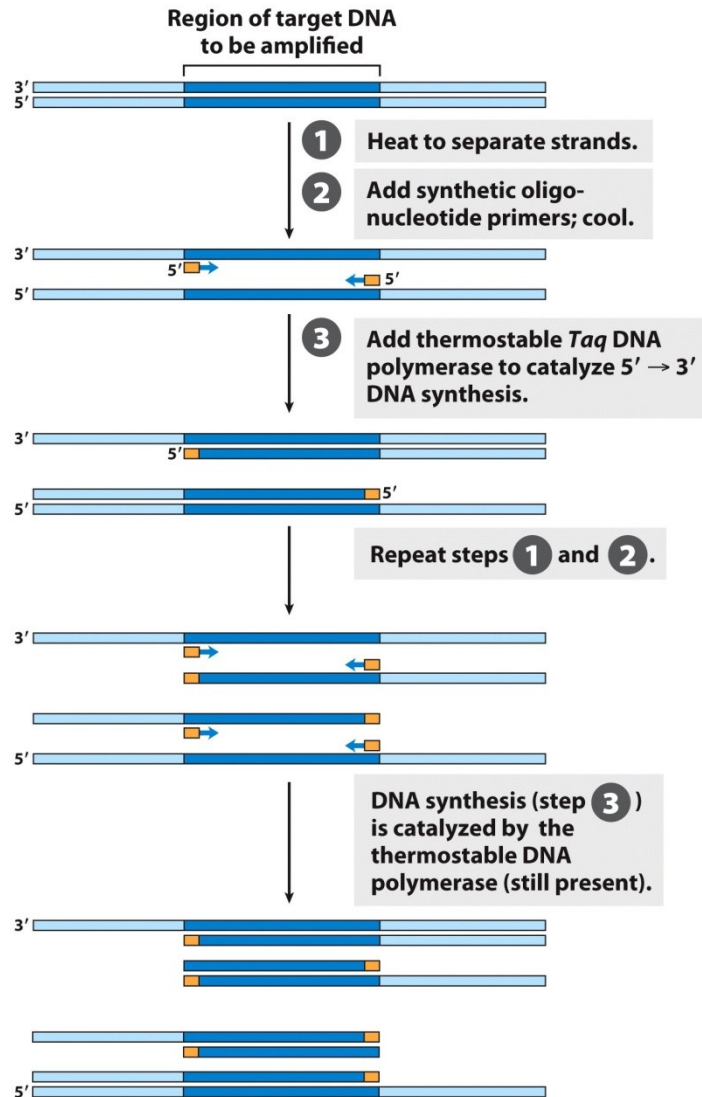


Figure 9-12a part 1

Lehninger Principles of Biochemistry, Sixth Edition
© 2013 W. H. Freeman and Company

General Steps of PCR

- Repeat steps 1–3 many times:

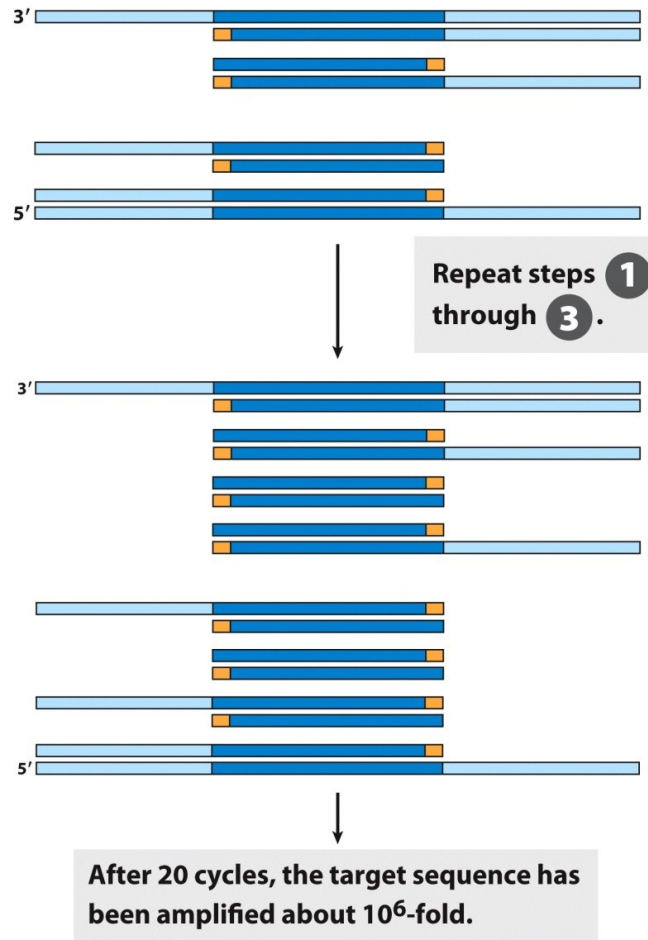
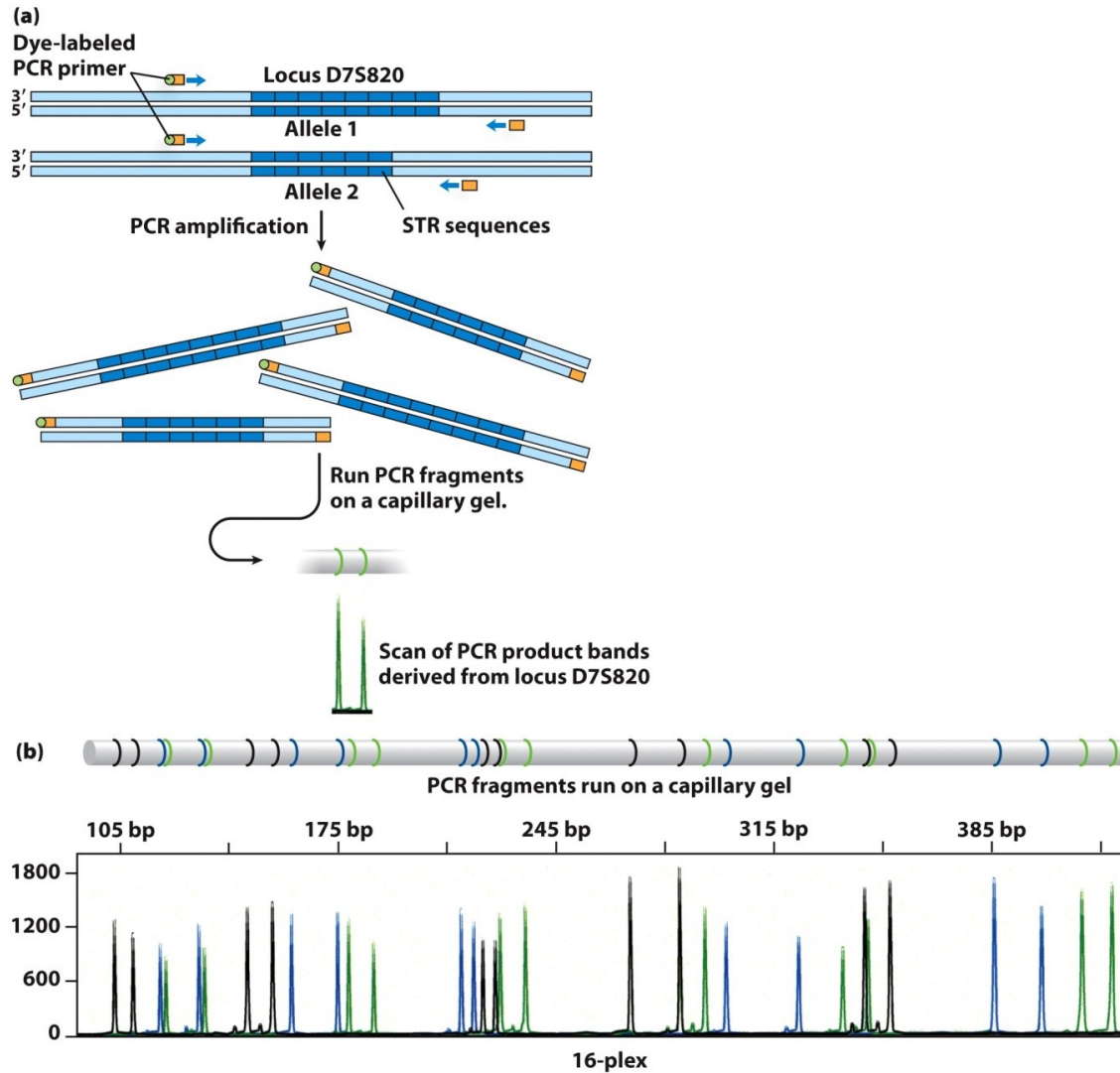


Figure 9-12a part 2
Lehninger Principles of Biochemistry, Sixth Edition
© 2013 W. H. Freeman and Company

DNA Fingerprinting

- Humans have short sequences that repeat next to each other (Short tandem repeats (STR))
- Differences in the number of repeats cause varying fragment lengths when sample subjected to PCR using a primer specific for that region
- Fragment sizes determined by using a capillary gel
- Multiple STR locations exist in the human genome
- Allows matching of “suspect” samples to known individuals
- 13 well-studied locations are used in identifications
 - Based on number of alleles at each location
misidentification is <1 in 10^{18} (with good data)

DNA Genotyping



Box 9-1 figure 1
Lehninger Principles of Biochemistry, Sixth Edition
© 2013 W. H. Freeman and Company

Adaptations to PCR

- Reverse Transcriptase PCR (RT-PCR)
 - Used to amplify RNA sequences
 - First step uses reverse transcriptase to convert RNA to DNA
- Quantitative PCR (Q-PCR)
 - Used to show quantitative differences in gene levels

qPCR

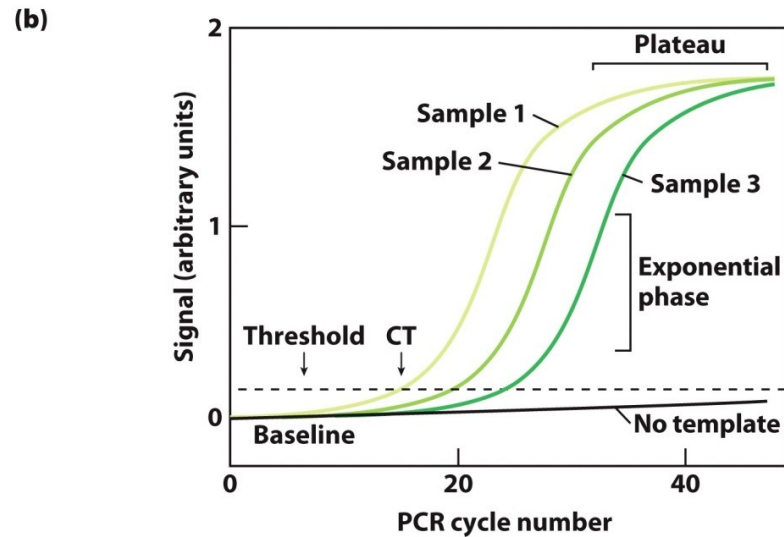
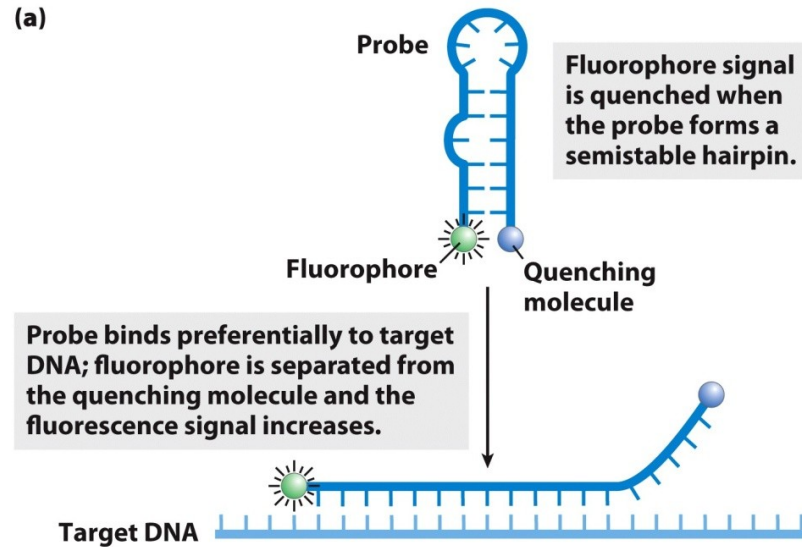


Figure 9-13
Lehninger Principles of Biochemistry, Sixth Edition
© 2013 W. H. Freeman and Company