Gel electrophoresis

Gel electrophoresis

Separates

- DNA fragments (single nucleotide resolution)
- proteins
- Protein-DNA complexes can be analyzed by gel electrophoresis

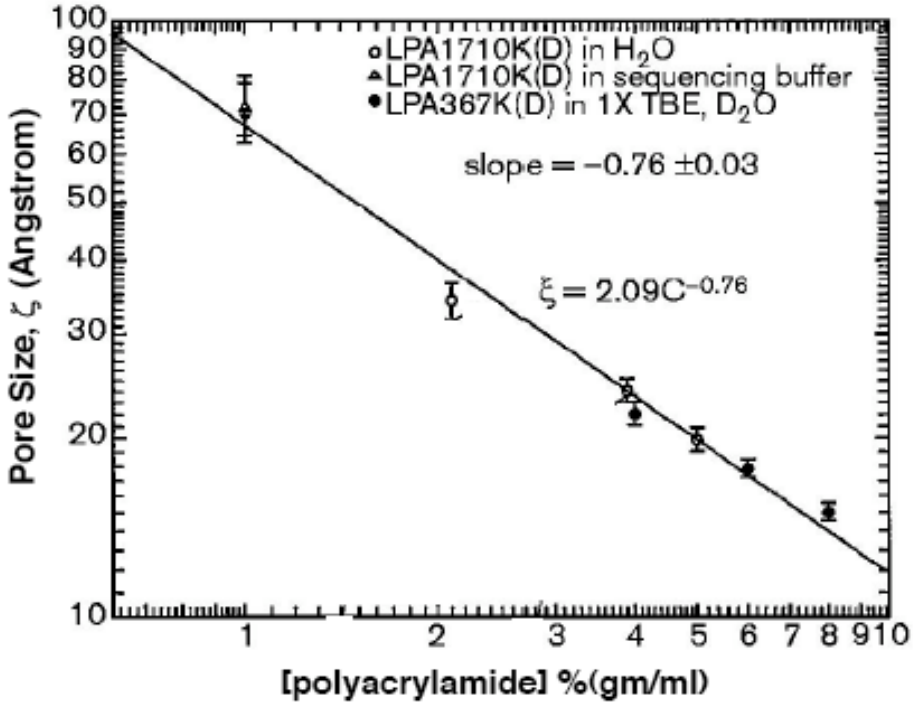
native - denatured gel-electrophoresis

agarose - polyacrylamide gel-electrophoresis

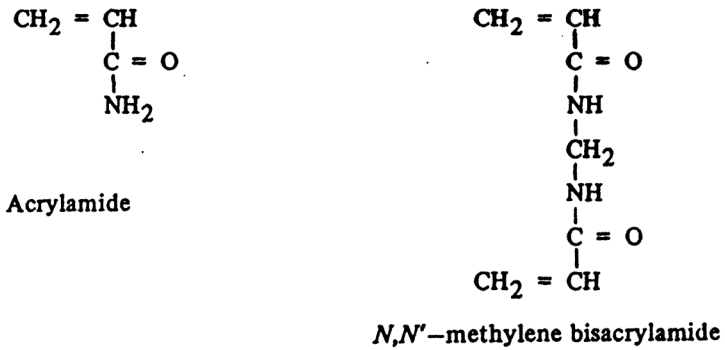
What is needed?

- Agarose - a polysaccharide made from seaweed. Agarose is dissolved in buffer and heated, then cools to a gelatinous solid with a network of crosslinked molecules
- 1% agarose has ~ 200 nm diameter pores

Polyacrylamide gels have a pore size of only a few nm

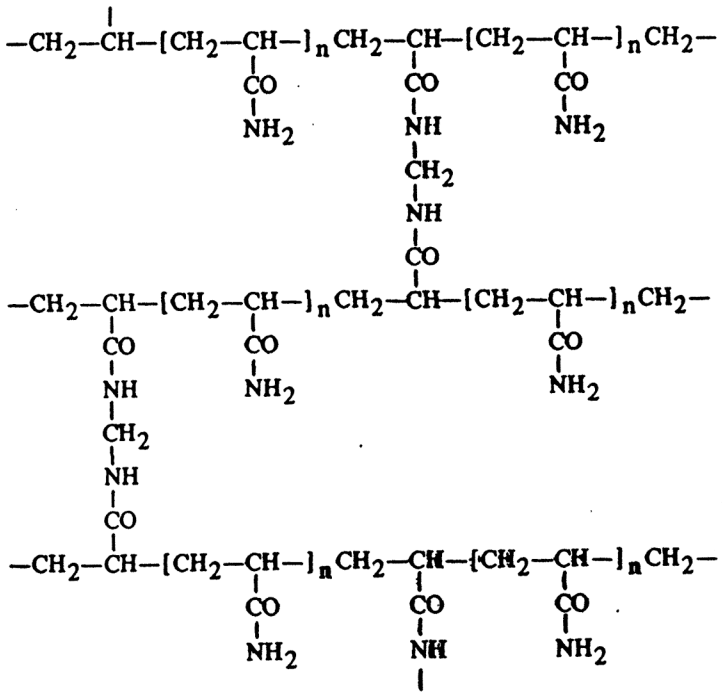


Quesada 1997 Curr. Opin. Biotech. 8:82-93



Acrylamide

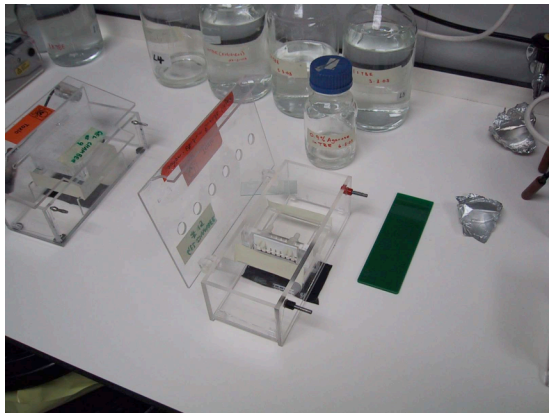
N,N'–methylene bisacrylamide



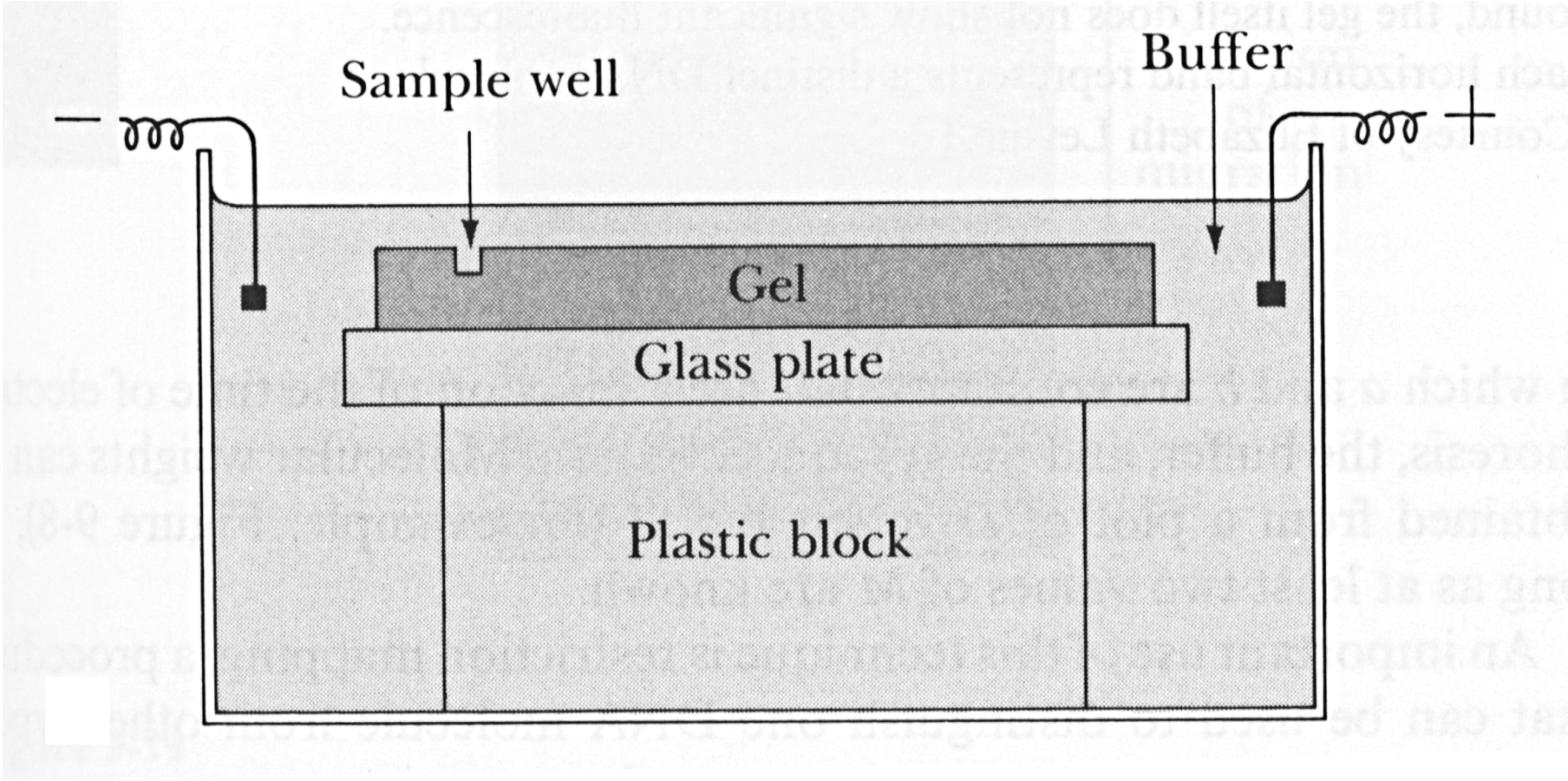
Polyacrylamide gel

- Buffer - in this case TBE
- The buffer provides ions in solution to ensure electrical conductivity.
- Not only is the agarose dissolved in buffer, but the gel slab is submerged (submarine gel) in buffer after hardening

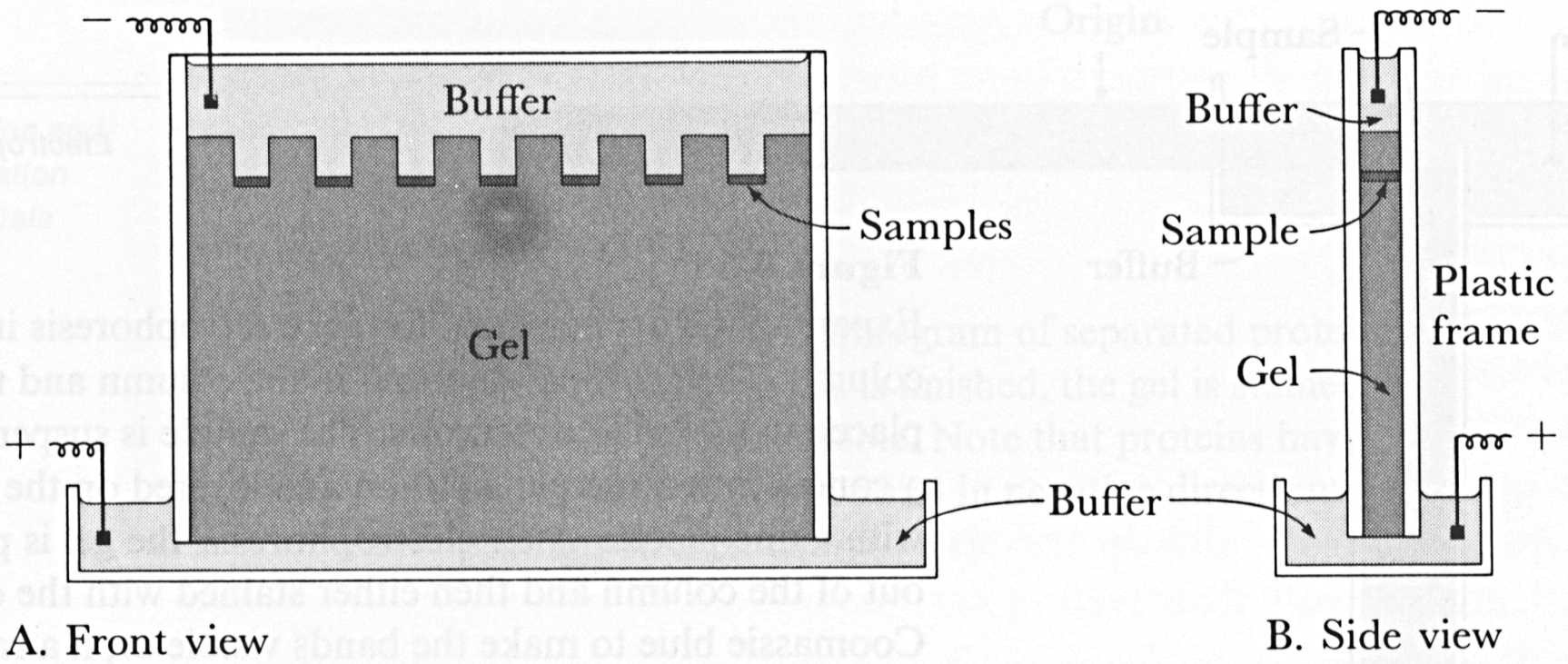
- Also needed are a power supply and a gel chamber
- Gel chambers come in a variety of models, from commercial through home-made, and a variety of sizes



Agarose gel electrophoresis apparatus



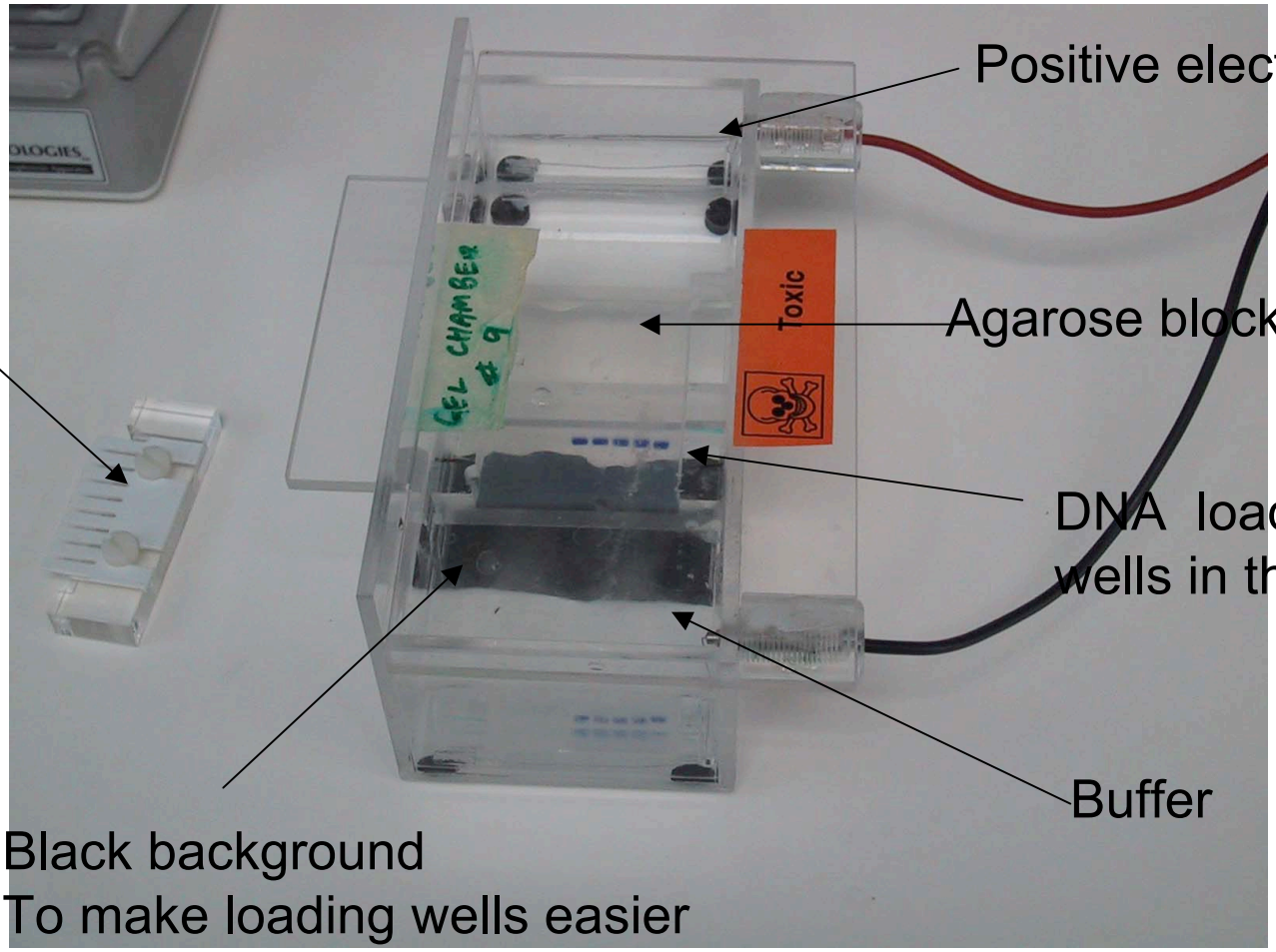
Polyacrylamide gel electrophoresis apparatus



How does it work?

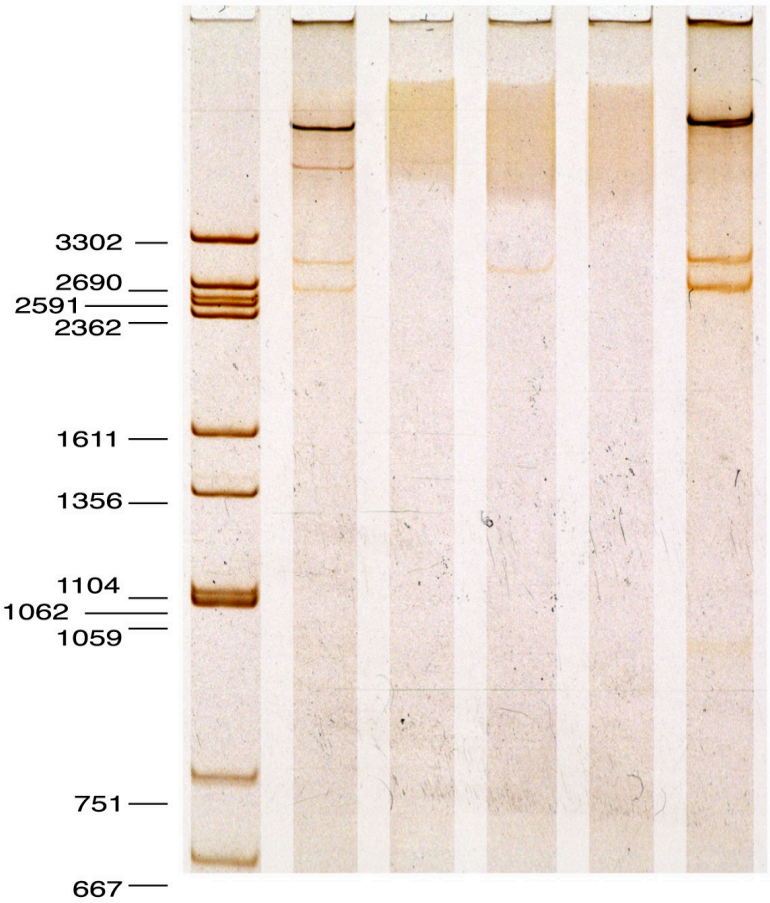
- DNA is an organic acid, and is negatively charged
- When the DNA is exposed to an electrical field, the particles migrate toward the positive electrode
- Smaller pieces of DNA can travel further in a given time than larger pieces

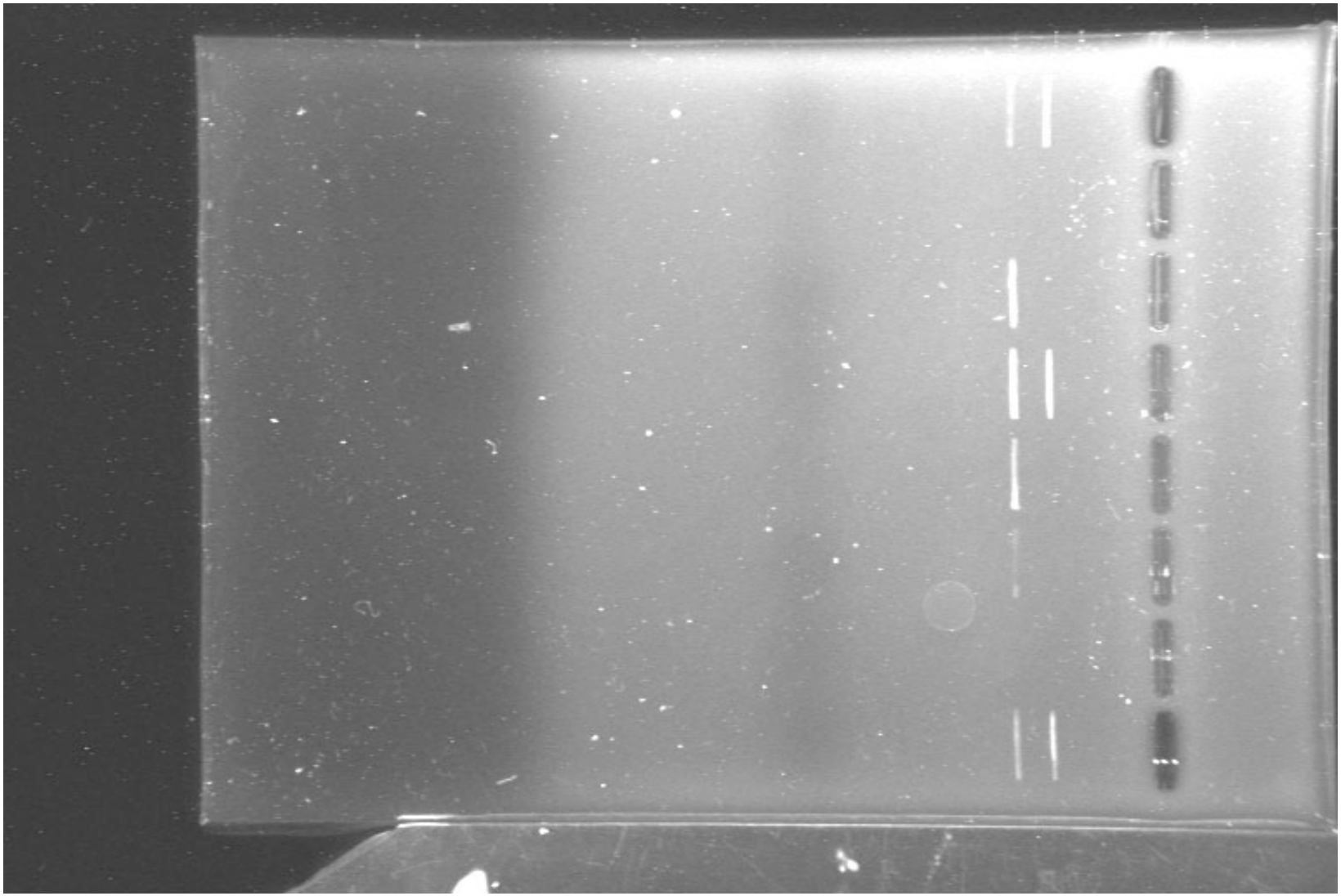
A gel being run



The results from a DNA gel

- Many samples can be run on one gel- but it is important to keep track
- Most gels have one lane as a ‘DNA ladder’ - DNA fragments of known size are used for comparison





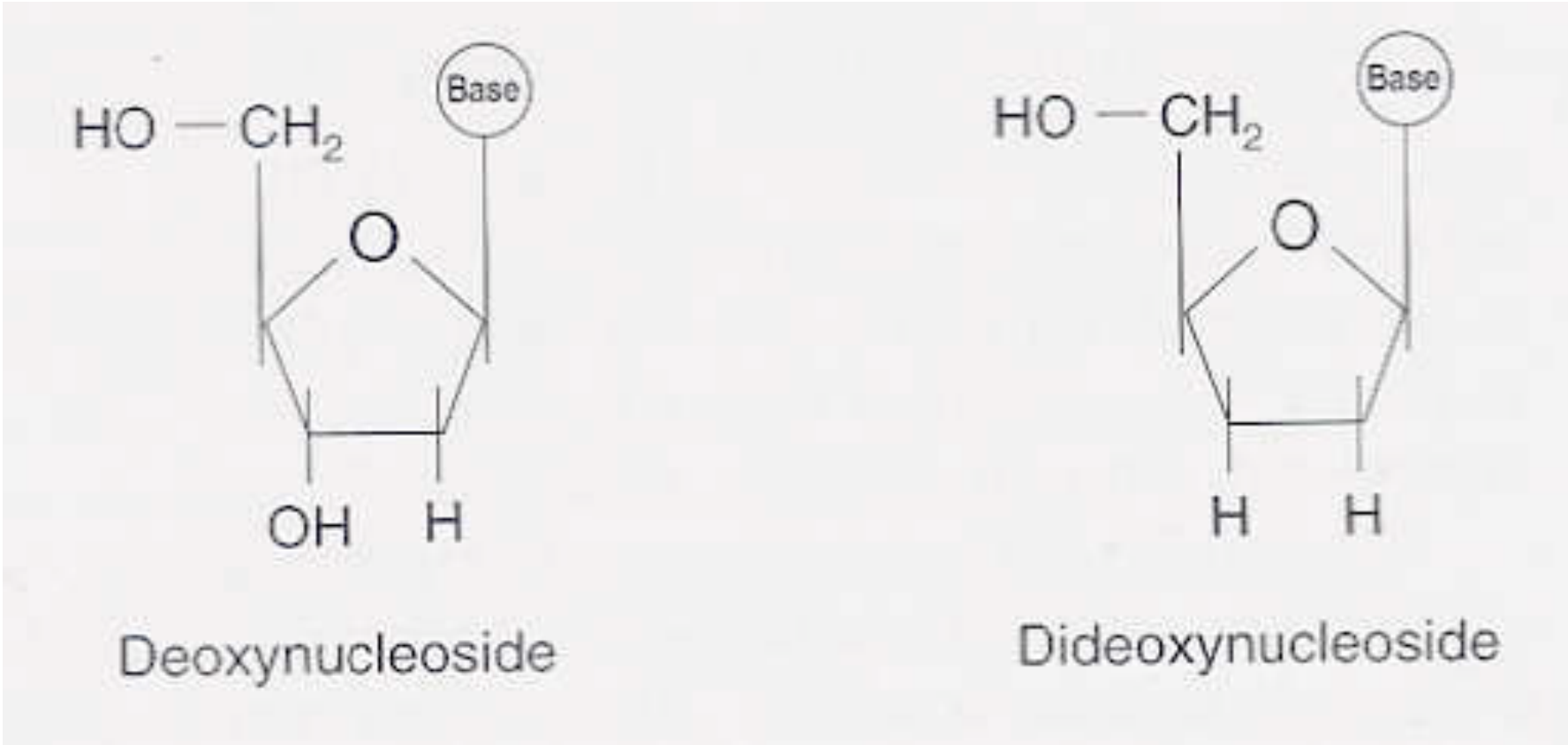
A gel as seen under UV light - some samples show 2 Bands of DNA, while others had none or one. The slower one could be the protein-DNA complex

DNA Sequencing

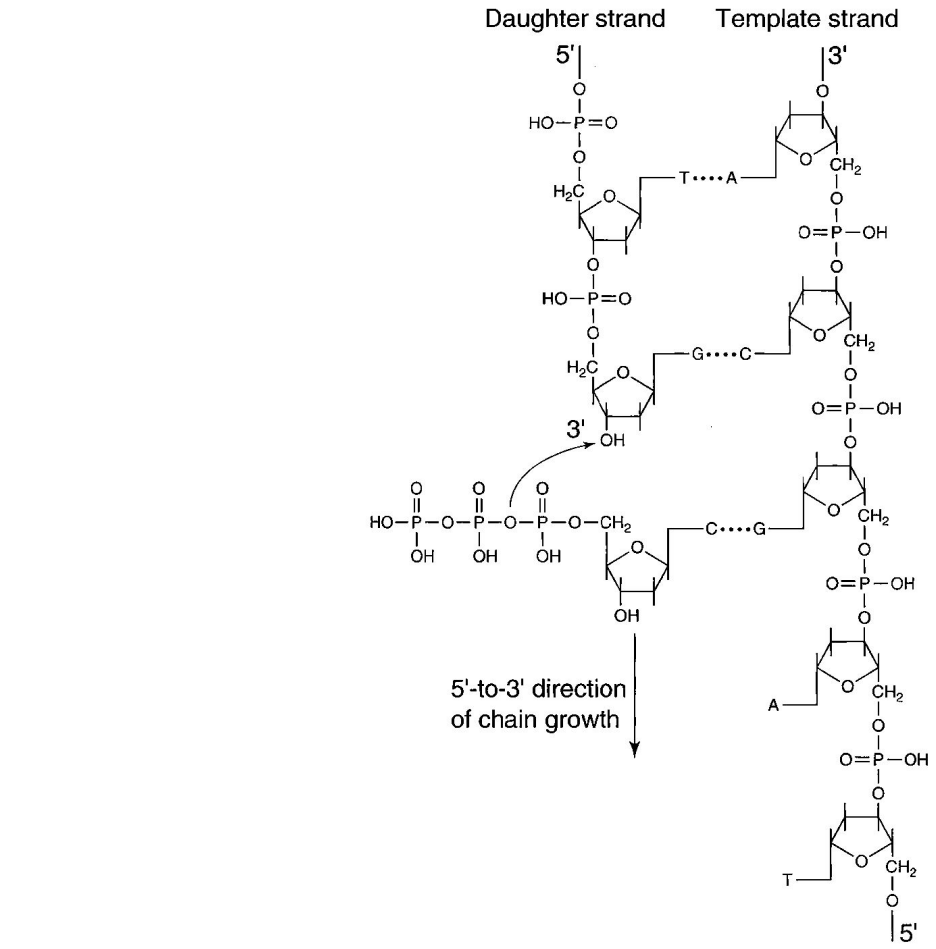
DNA Sequencing

- DNA sequencing – used to determine the actual DNA sequence of an organism. Using a computer, one can identify an organism, and predict protein sequences and functions based on the nucleic acid data.
  - The most commonly used sequencing method is the dideoxy method.
    - This method uses dideoxynucleotide triphosphates(ddNTPs) which have an H on the 3' carbon of the ribose sugar instead of the normal OH found in deoxynucleotide triphosphates (dNTPs). Dideoxynucleotides are chain terminators. In a synthesis reaction, if a dideoxynucleotide is added instead of the normal deoxynucleotide, the synthesis stops at that point because the 3'OH necessary for the addition of the next nucleotide is absent.

Deoxy versus dideoxy



DNA synthesis



DNA sequencing continued

- In the dideoxy method of sequencing, the template DNA that is to be sequenced is mixed with a primer complementary to the template DNA and the four normal dNTPs, one of which is radioactively labeled for subsequent visualization purposes.
- This mixture is then splint into four different tubes that are labeled A, C, G, and T. Each tube is then “spiked” with a different ddNTP (ddATP for tube A, ddCTP for tube C, ddGTT for tube G, or ddTTP for tube T).
- DNA polymerase is added and using the DNA template and its’ complementary primer, the synthesis of new strands of DNA complementary to the template begins.
- Occasionally a dideoxynucleotide is added instead of the normal deoxynucleotide and synthesis of that strand is terminated at that point.

DNA sequencing continued

- In the tube containing ddATP, some percentage of newly synthesized molecules will get a ddATP in each place that there is a T in the template DNA.
- The result is a set of new DNA molecules in tube A, each of which ends in an A.
- A similar type of reaction occurs in the three other tubes to result in molecules that end in C, G, and T in tubes C, G, and T respectively.
- After the synthesis reactions are complete, the products of the four different tubes are loaded onto four adjacent lane of a polyacrylamide gel and the different fragments are separated by size.
- The sequencing gel is able to resolve fragments that differ in size from each other by only one base.

DNA sequencing continued

- After electrophoresis to separate the fragments by size, the fragments are visualized to exposing the gel to photographic film (Remember that one nucleotide was radioactively labeled).
- All fragments in lane A will end in an A, fragments in lane C will all end in a C, fragments in lane G will all end in a G, and fragments in lane T will all end in a T.
- The sequence of the DNA is read from the gel by starting at the bottom and reading upward.

Dideoxynucleoside Sequencing

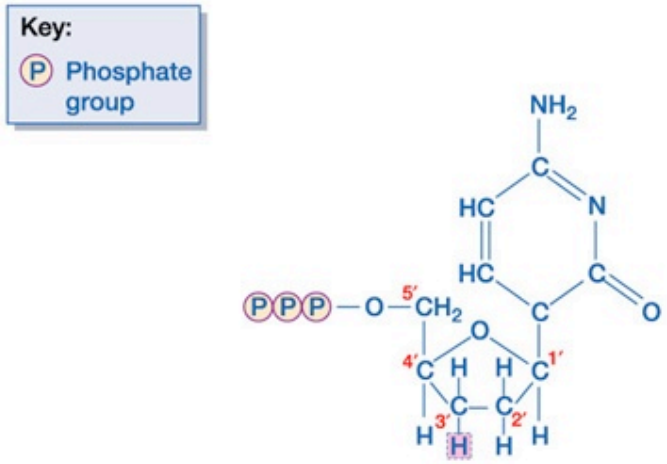


Figure 7-1 Human Molecular Genetics, 3/e. (© Garland Science 2004)

Dideoxynucleoside Sequencing

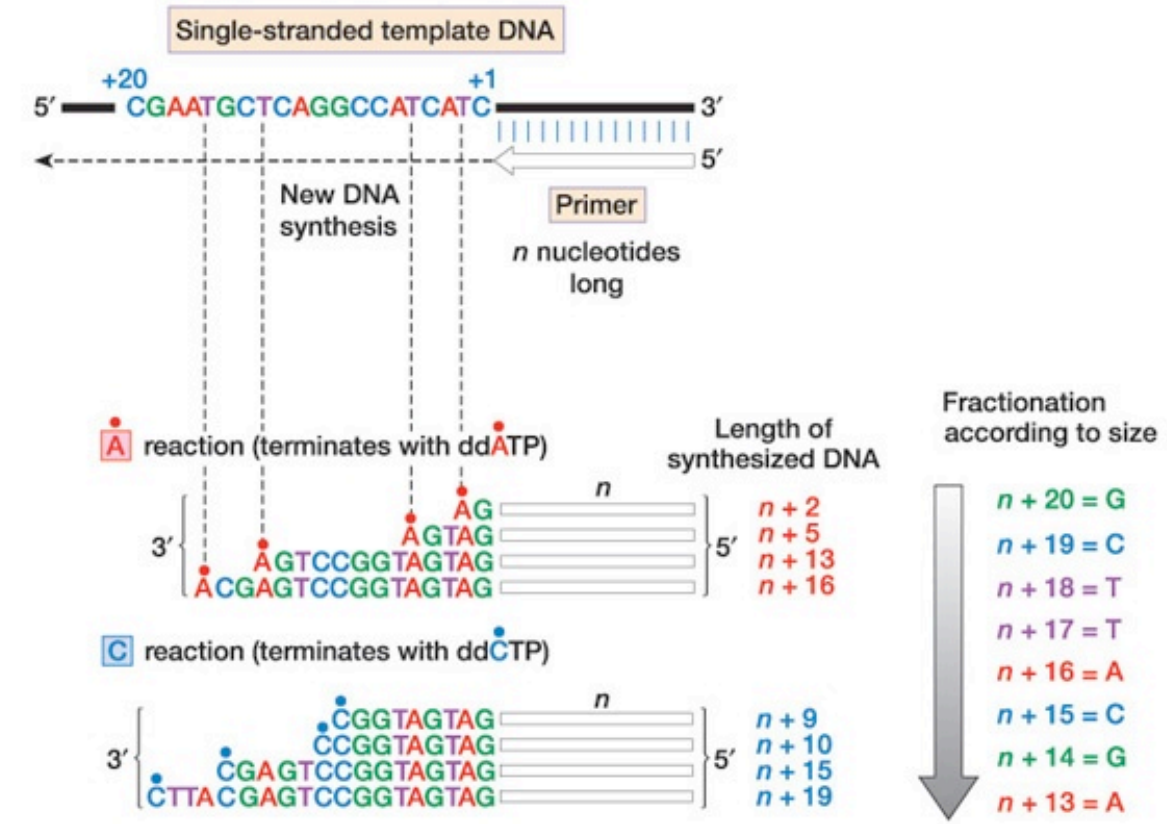
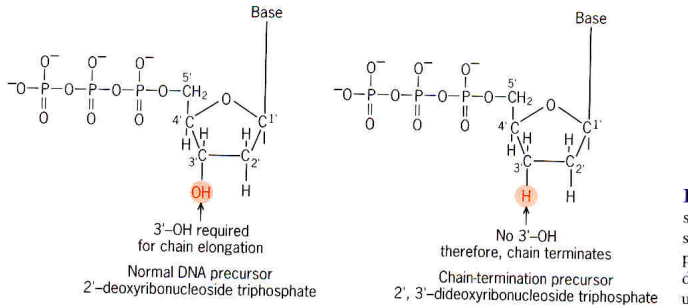
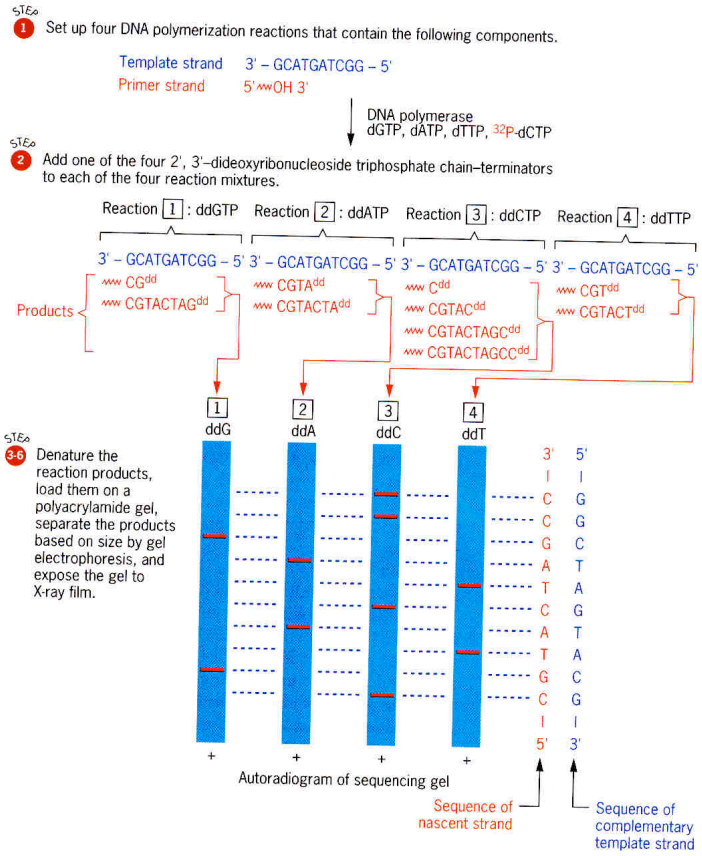


Figure 7-2 part 1 of 2 Human Molecular Genetics, 3/e. (© Garland Science 2004)

Dideoxy DNA Sequencing





DNA sequencing

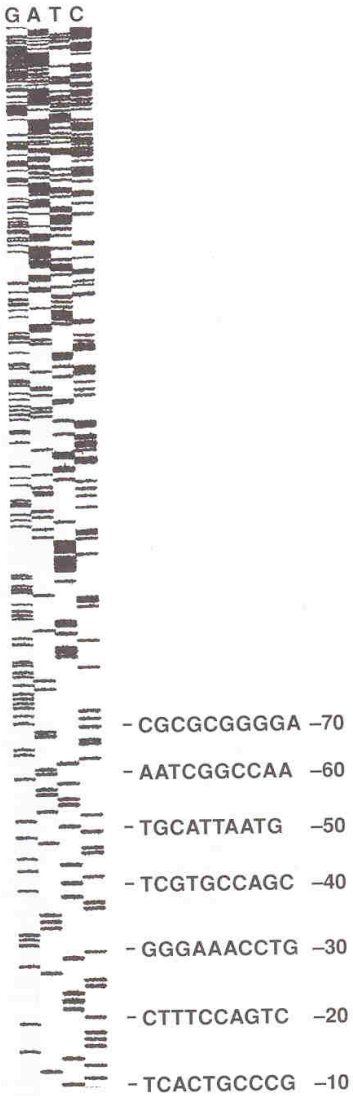


Figure 20.28 Photograph of an autoradiograph of a 2′,3′-dideoxynucleotide chain-terminator sequencing gel. The sequence defined by the lower portion of the gel is shown on the right.

DNA sequencing

- Automated DNA sequencing – in automated DNA sequencing a radioactive deoxynucleotide is not used and all four dideoxy reactions are done in a single tube.
- This is possible because each ddNTPs is labeled with a different flourescent dye.
- Therefore the dye present in each synthesized fragment corresponds to the dye attached to the dideoxynucleotide that was added to terminate the synthesis of that particular fragment.
- The contents of the single tube reaction are loaded onto a single lane of a gel and electrophoresis is done.

DNA Sequencing

- A flourimeter and computer are hooked up to the gel and they detect and record the dye attached to the fragments as they come off the gel.
- The sequence is determined by the order of the dyes coming off the gel.

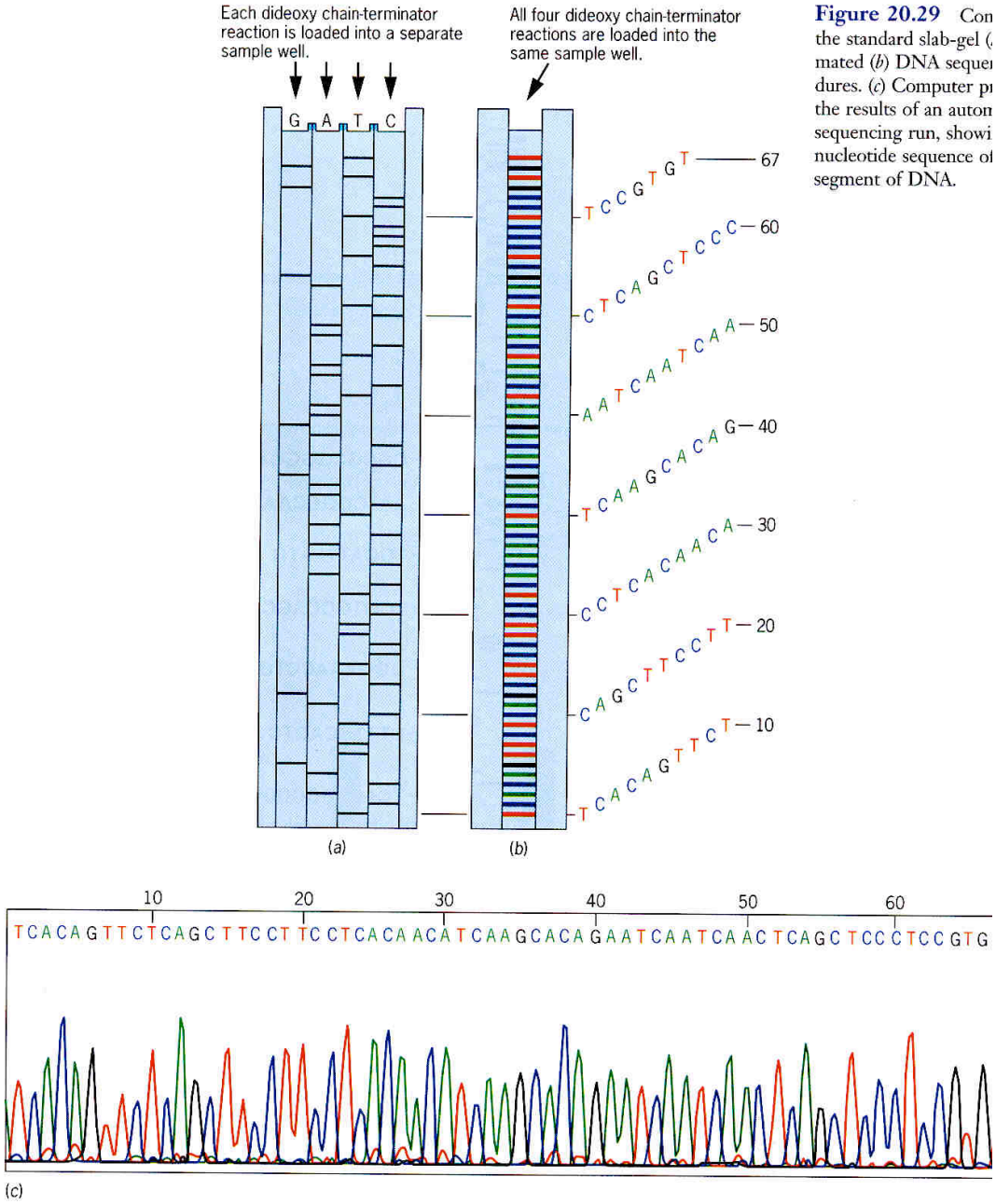


Figure 20.29 Con
the standard slab-gel (
mated (*b*) DNA sequer
dures. (*c*) Computer pr
the results of an auton
sequencing run, showi
nucleotide sequence of
segment of DNA.

Automated DNA sequencing

Dideoxynucleoside Sequencing

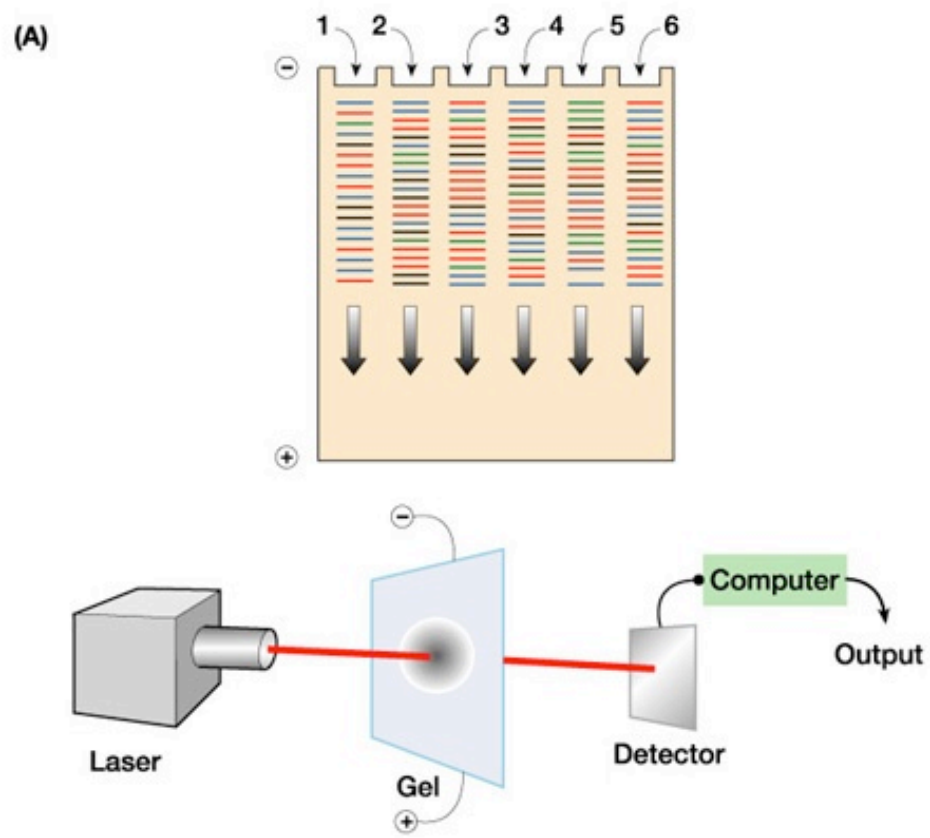


Figure 7-3 part 1 of 2 Human Molecular Genetics, 3/e. (© Garland Science 2004)

Genome Sequencing Approaches

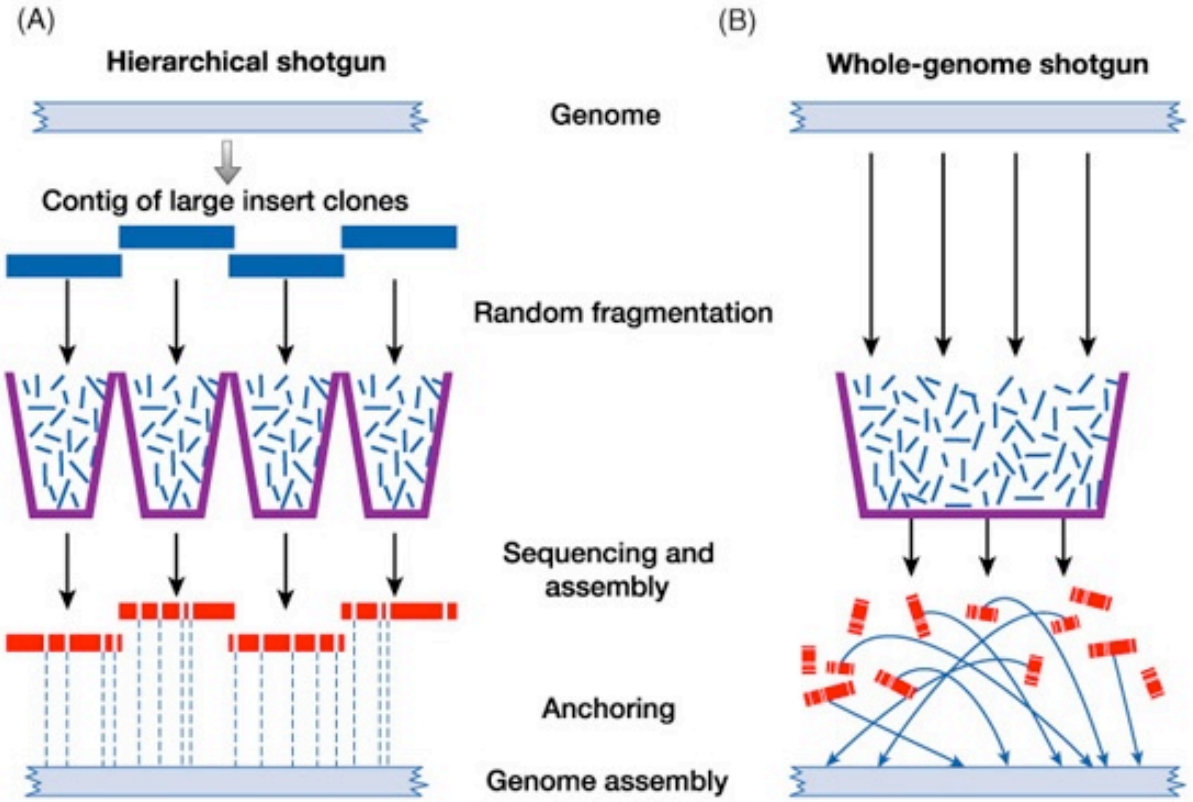
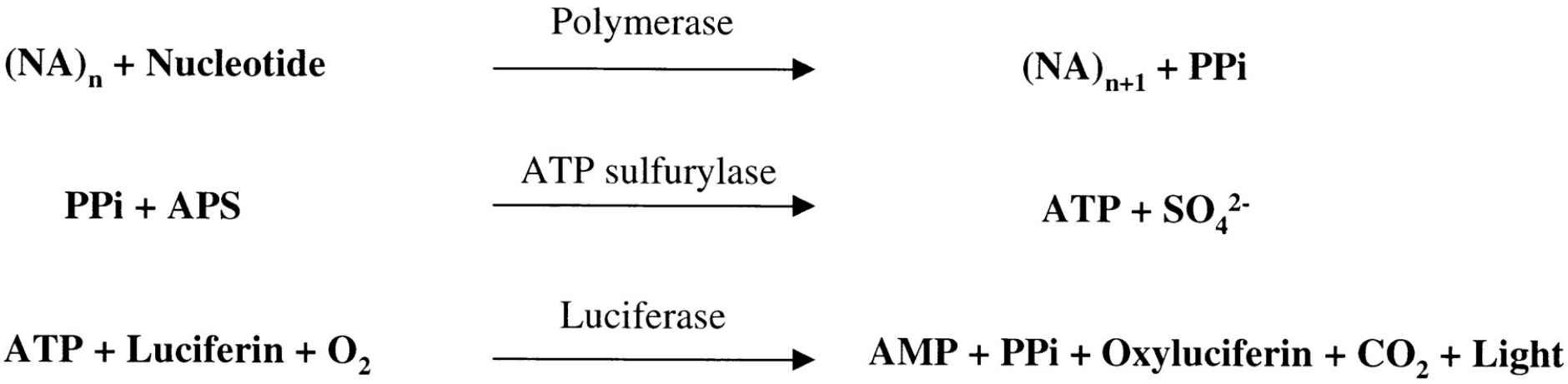


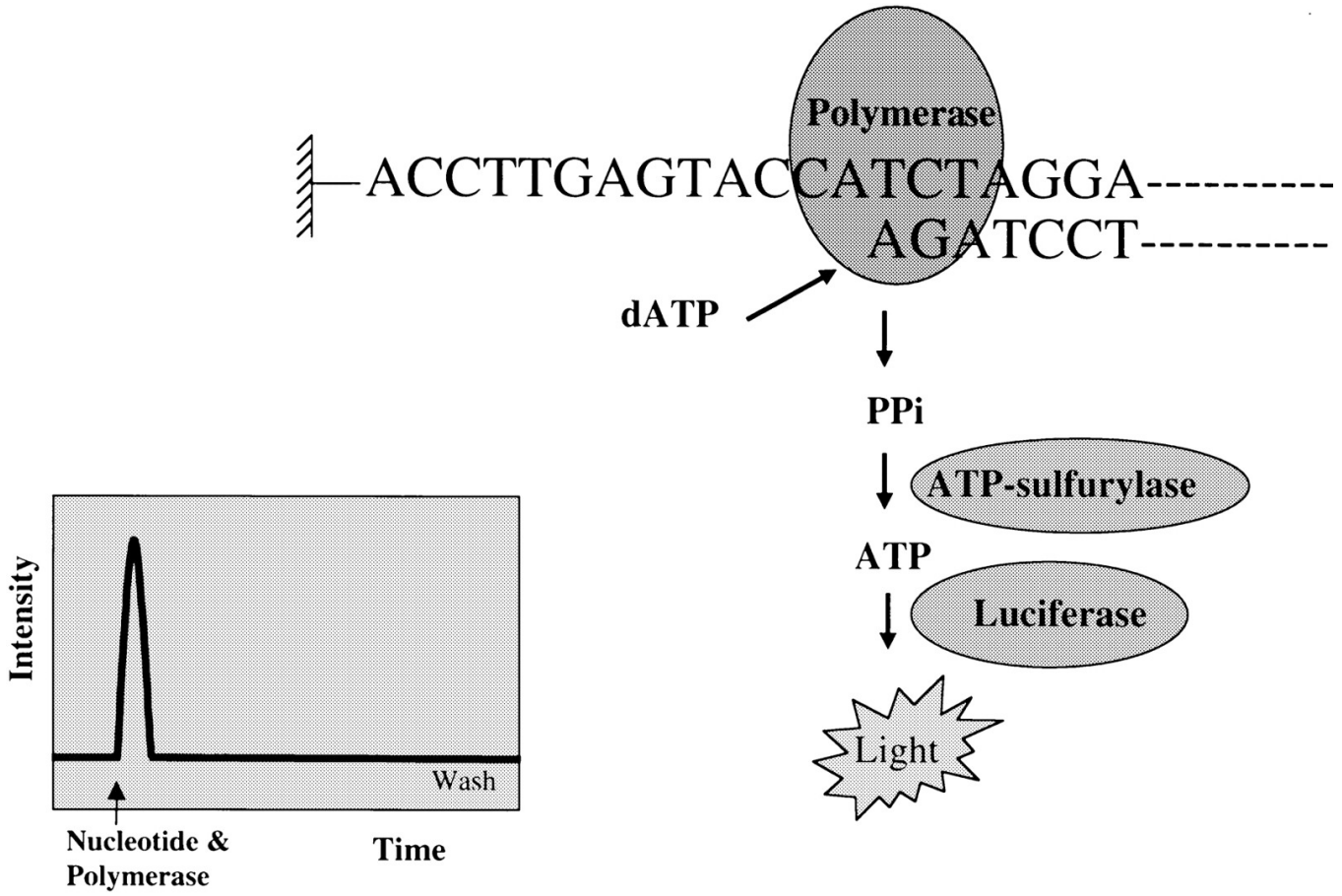
Figure 8-3 Human Molecular Genetics, 3/e. (© Garland Science 2004)

Pyrosequencing



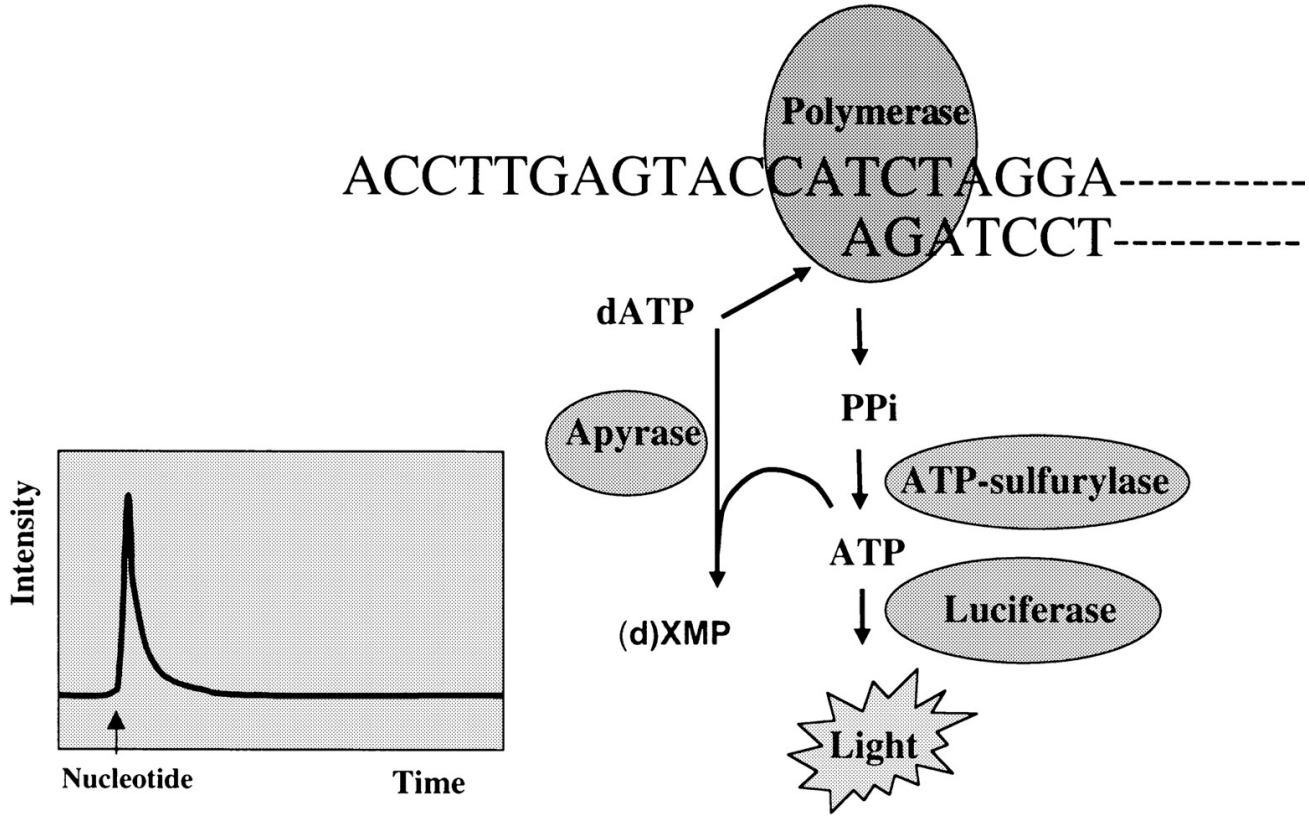
Ronaghi M. Pyrosequencing sheds light on DNA sequencing. Genome Res 2001

Pyrosequencing - Solid Phase



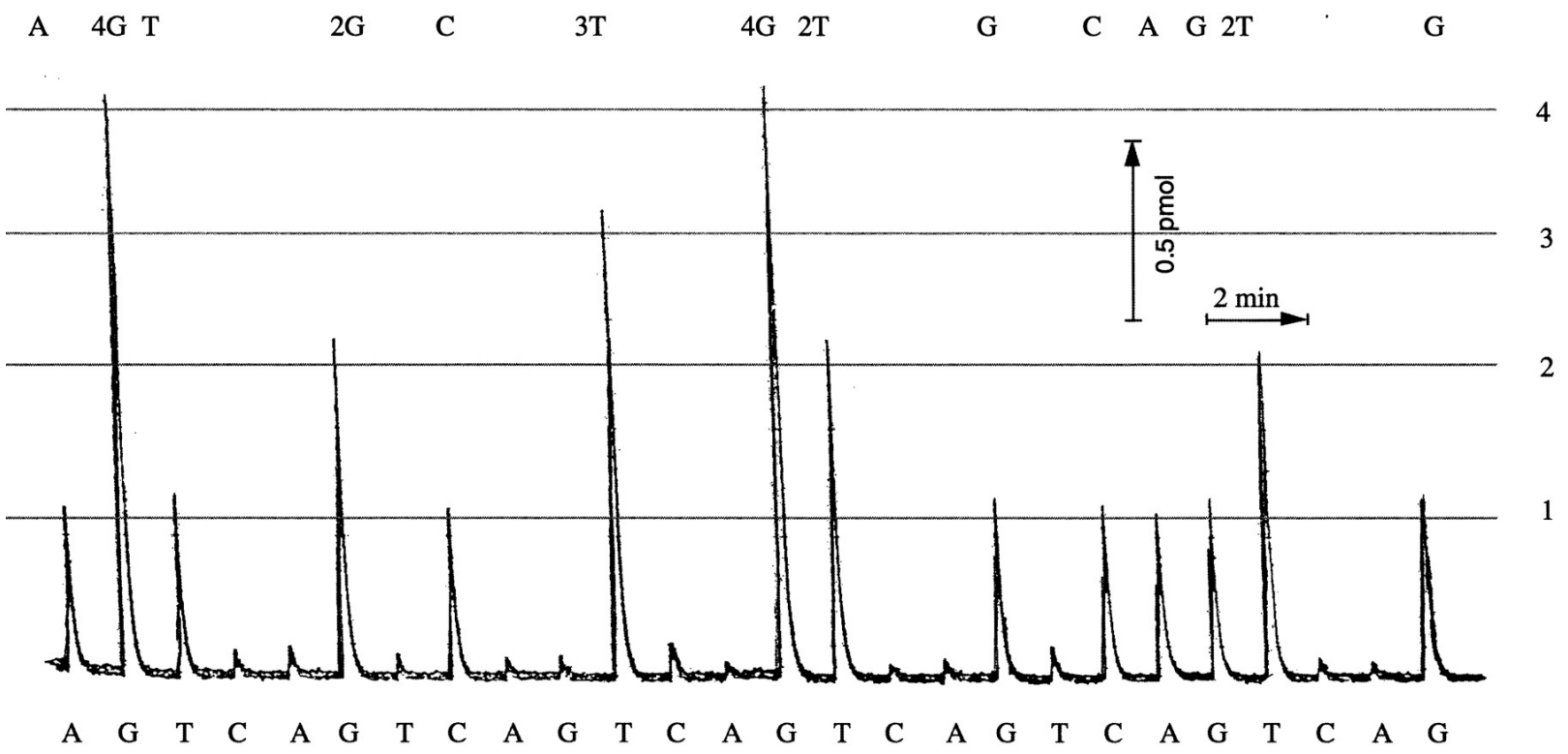
Ronaghi M. Pyrosequencing sheds light on DNA sequencing. Genome Res 2001

Pyrosequencing - Liquid Phase



Ronaghi M. Pyrosequencing sheds light on DNA sequencing. Genome Res 2001

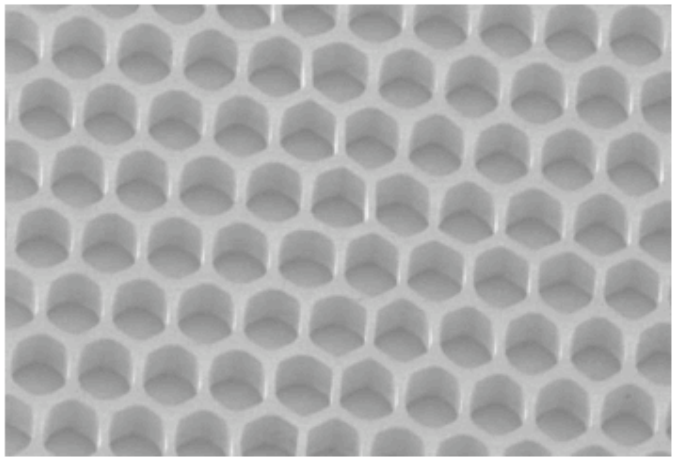
Pyrogram



Ronaghi M. Pyrosequencing sheds light on DNA sequencing. Genome Res 2001

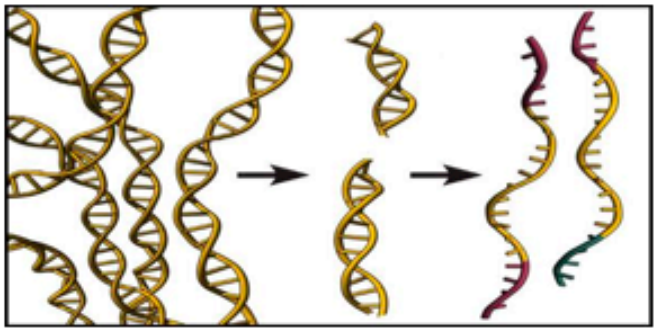
454 LifeSciences Sequencer

PicoTiter™ Plates

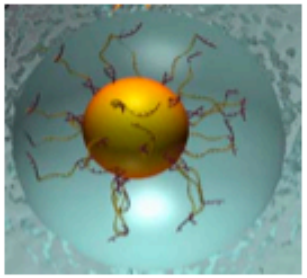


- Multiple optical fibers are fused to form an optical array
- Selective removal of core material leaves wells that serve as 'test tubes'
- Reactions occurring in the 'test tubes' can be monitored optically, through the remaining fiber
- Well diameter: 3µ - 250µ, 44µ typical.
- Plate contains 1.6 million wells

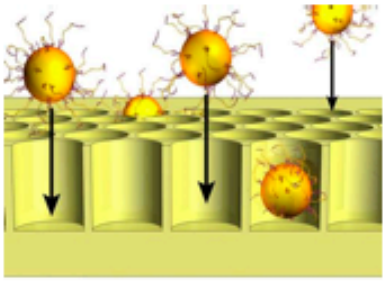
454 LifeSciences Sequencer - Process Overview



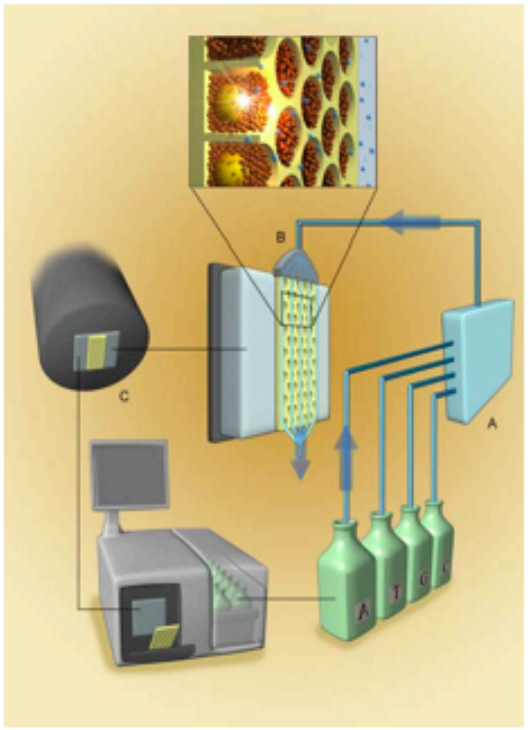
1) Prepare Adapter Ligated ssDNA Library



2) Clonal Amplification on 28 μ beads

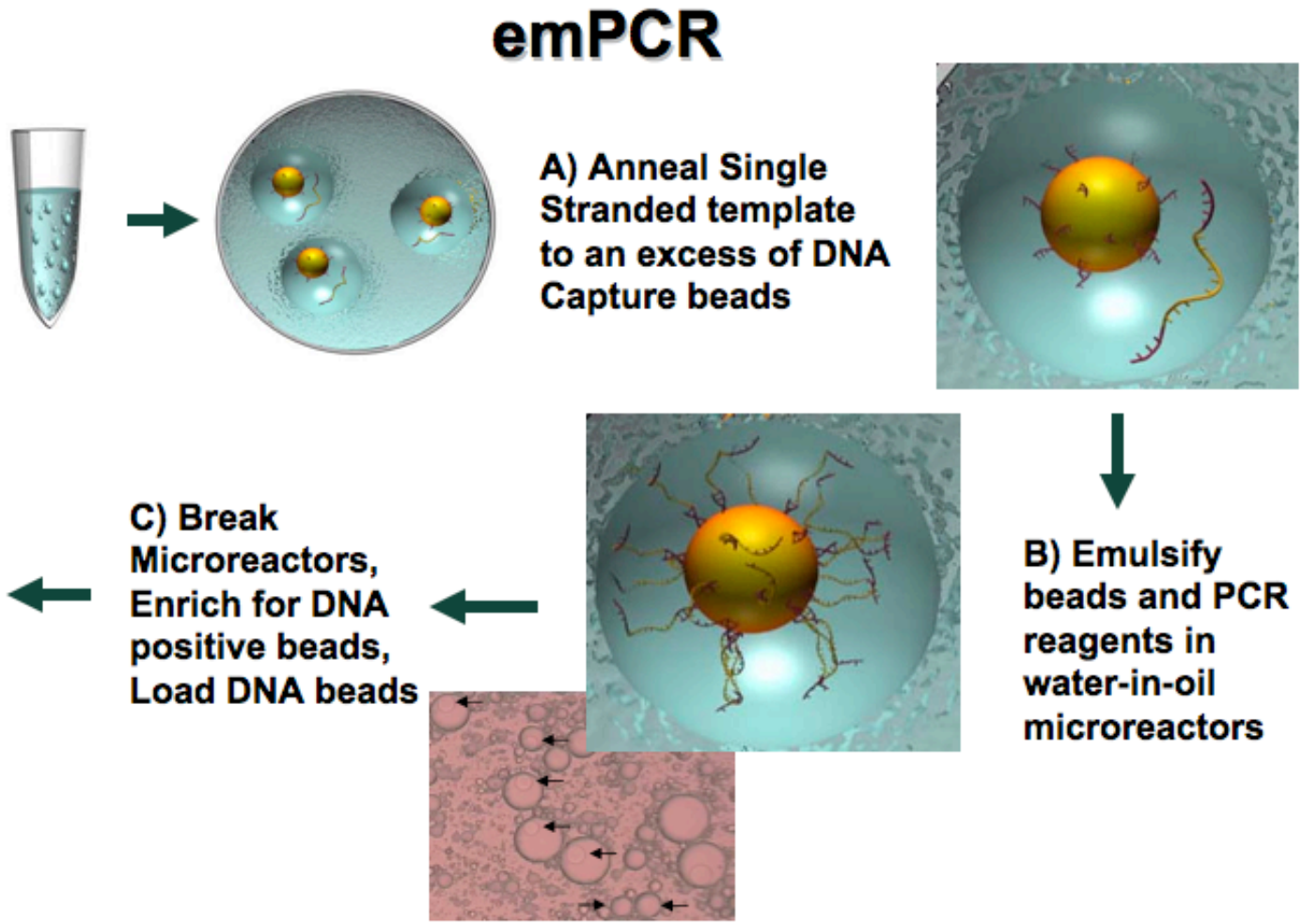


3) Load beads and enzymes in PicoTiter Plate™

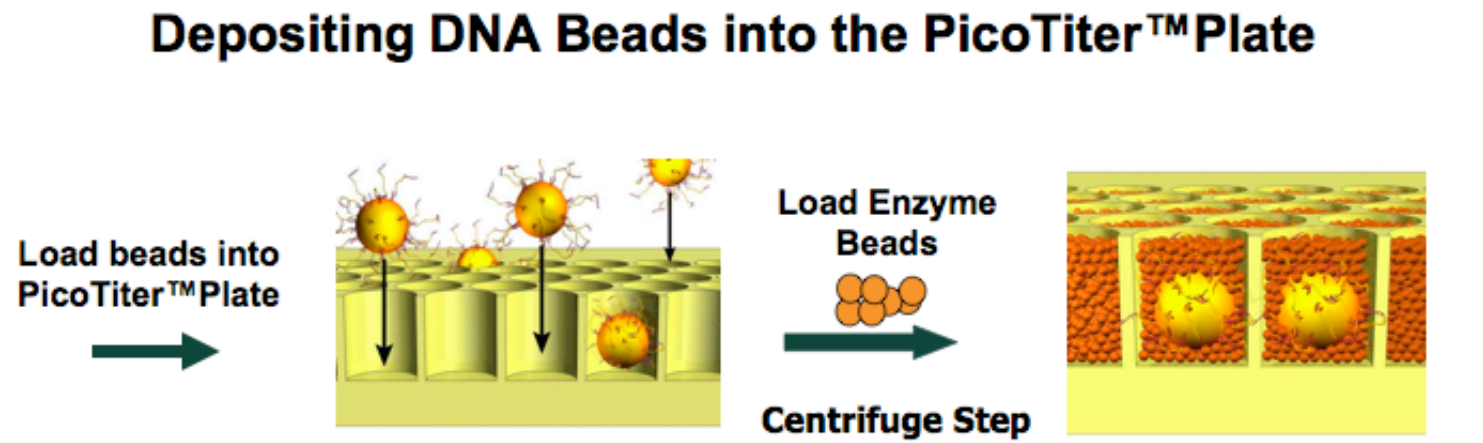


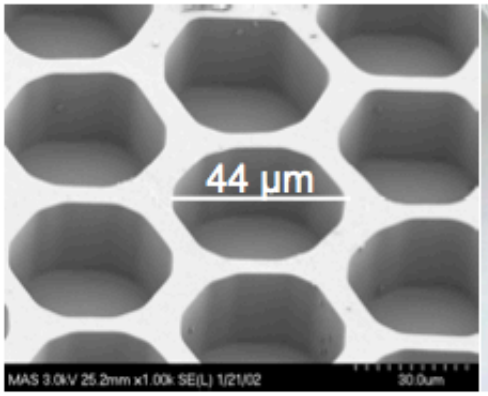
4) Perform Sequencing by synthesis on the 454 Instrument

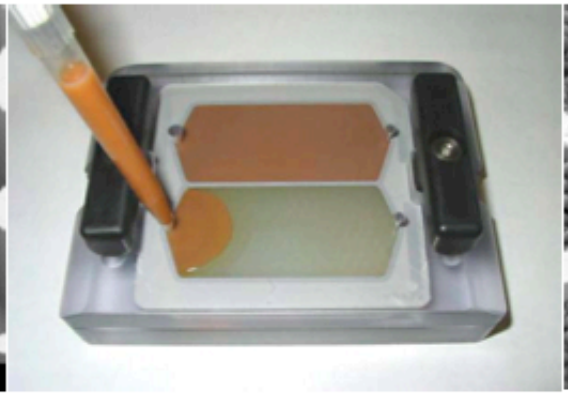
454 LifeSciences Sequencer - Process Overview

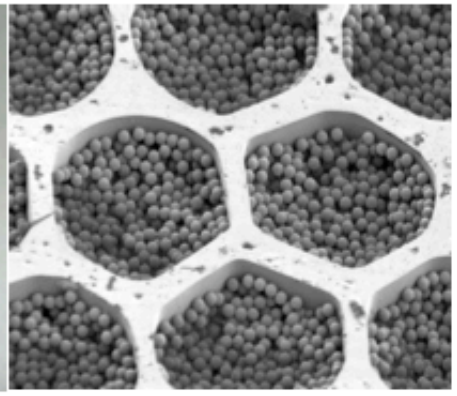


454 LifeSciences Sequencer



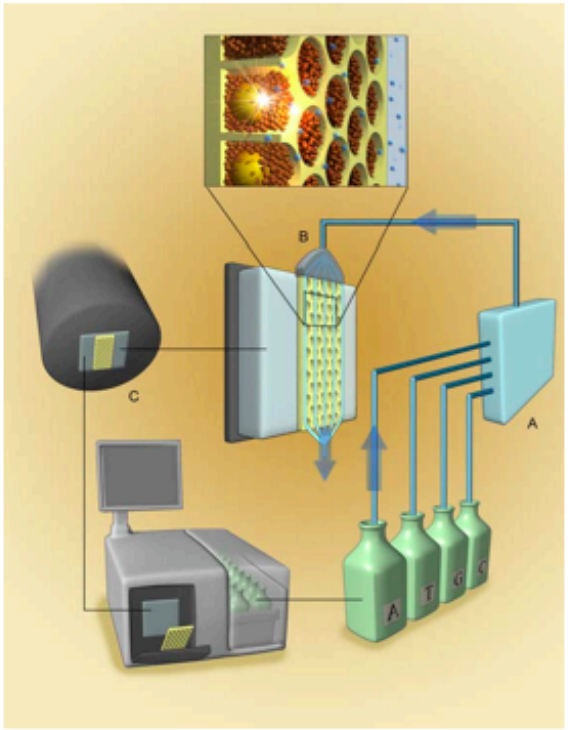


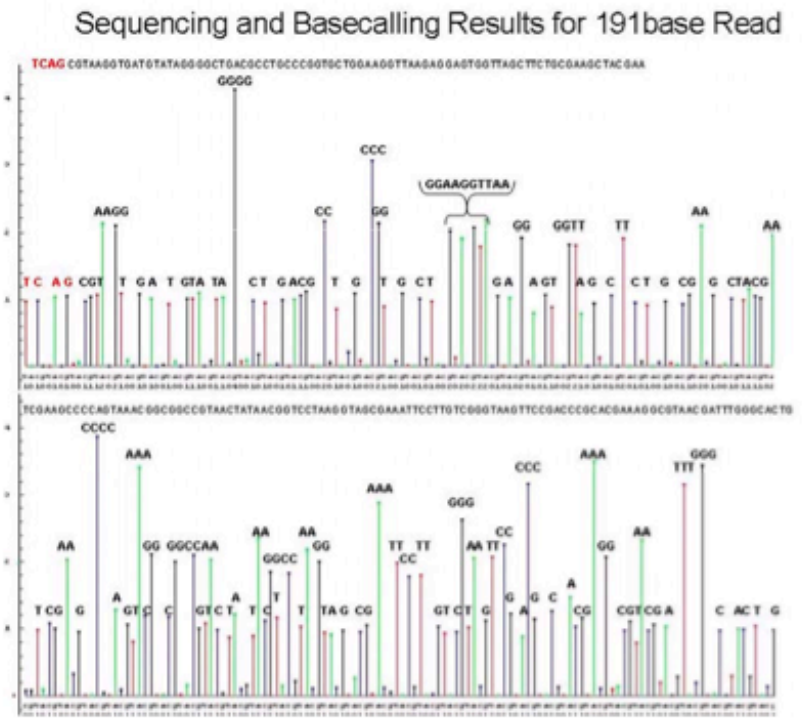




454 LifeSciences Sequencer

454 Technology - Sequencing Instrument





454 LifeSciences Sequencer

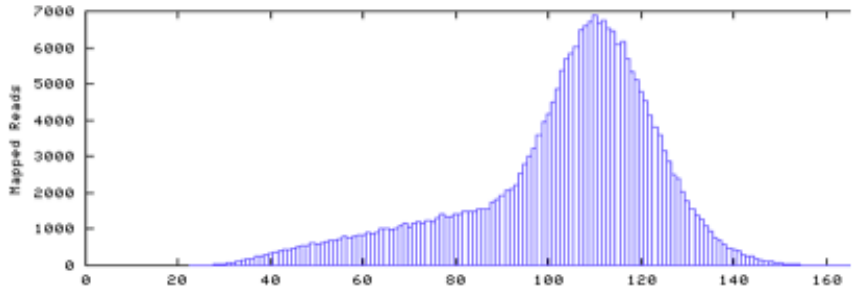
Read Results: Run Statistics

	# of Reads	# of Bases	Coverage
***C. jejuni* (NCTC 11168)** 1.6MB Genome 31% GC Content 1 Run 24.5 Million Bases	226,429	24,521,527	13.1
***Escherichia coli* (K12)** 4.6MB Genome 51% GC Content 4 Runs 86 Million Bases	904,858	86,135,933	18.6
***H. salinarum* (NRC-1)** 2.0MB Genome 68% GC Content 1 Run 21 Million Bases	213,774	21,309,326	10.4

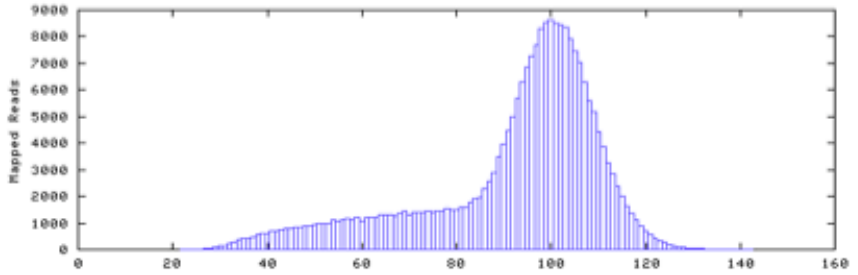
454 LifeSciences Sequencer

Read Results: Read Length

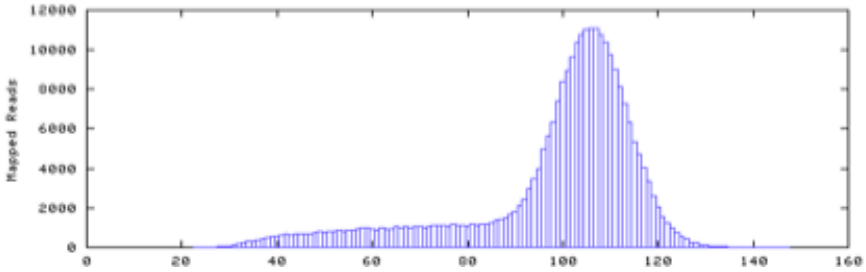
***C. jejuni* (NCTC 11168)**
1.6MB Genome
31% GC Content
1 Run
24.5 Million Bases



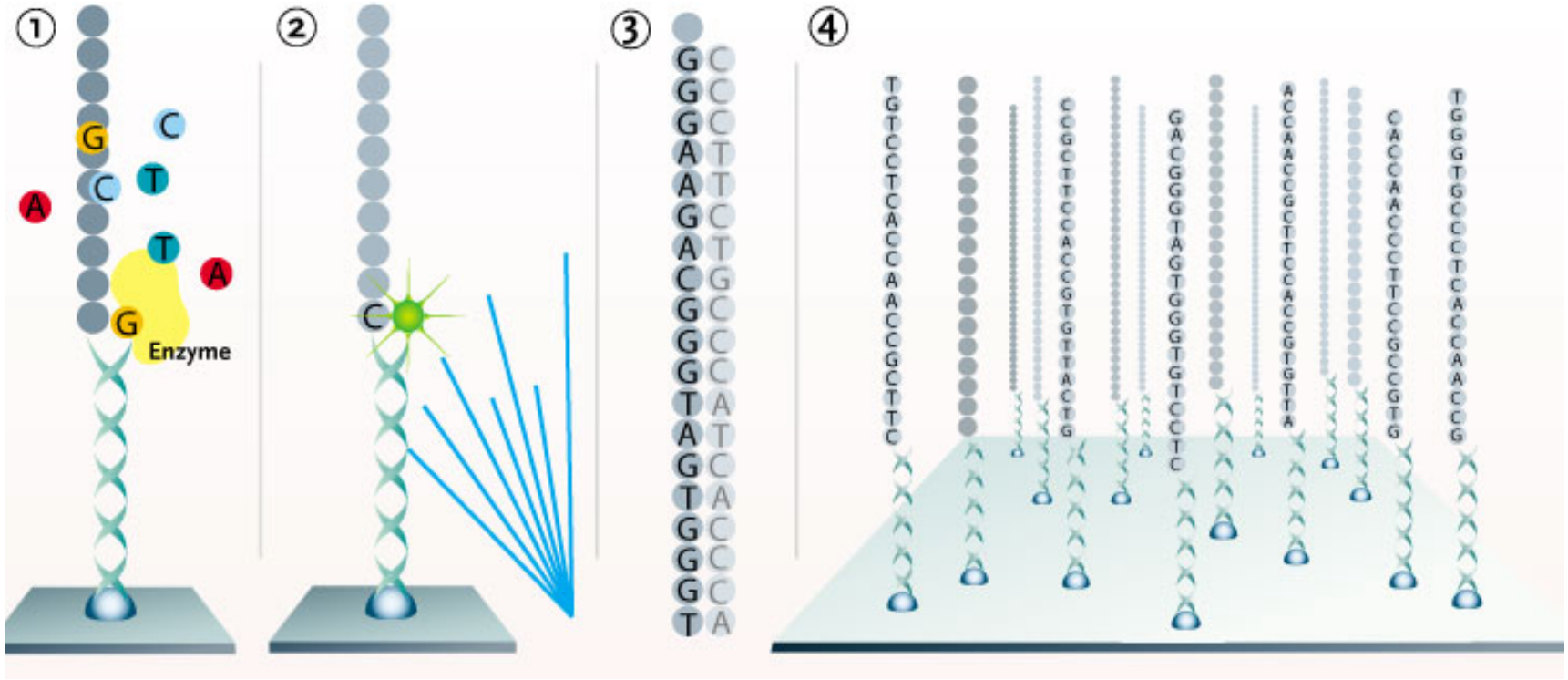
***Escherichia coli* (K12)**
4.6MB Genome
51% GC Content
4 Runs
86 Million Bases



***H. salinarum* (NRC-1)**
2.0MB Genome
68% GC Content
1 Run
21 Million Bases



Solexa Sequencing



Genome Projects

1976/79	First viral genome – MS2/fX174
1980	Mitochondrion
1982	First shotgun sequenced genome – Bacteriophage lambda
1995	First prokaryotic genome – H. influenzae
1996	First unicellular eukaryotic genome – Yeast
1998	The first multicellular eukaryotic genome – C.elegans
2000	Drosophila melanogaster
2000	Arabidopsis thaliana
2001	Human Genome
2002	Mouse Genome

The Genome OnLine Database knows of 958 genome sequencing projects, of which 169 are completed

Favourite and Model Organisms

Multicellular Animals

Mammals

Human	3.5 Gb
Mouse	3.2 Gb
Cow	3.0 Gb
Dog	2.8 Gb
Rat	3.1 Gb
Chimp	3.5 Gb
Pig	3.0 Gb

Fish

Puffer Fish	0.4 Gb
Zebra Fish	1.9 Gb

Insects

Drosophila	165 Mb
Honey Bee	270 Mb
Yellow Fever Mosquito	780 Mb
Malaria Mosquito	278 Mb

Birds

Chicken	1.2 Gb

Frog

Xenopus Laevis	1.7 Gb

Nematodes

Caenorhabdites elegans	100 Mb
Caenorhabdites briggsae	80 Mb

Sea Urchin

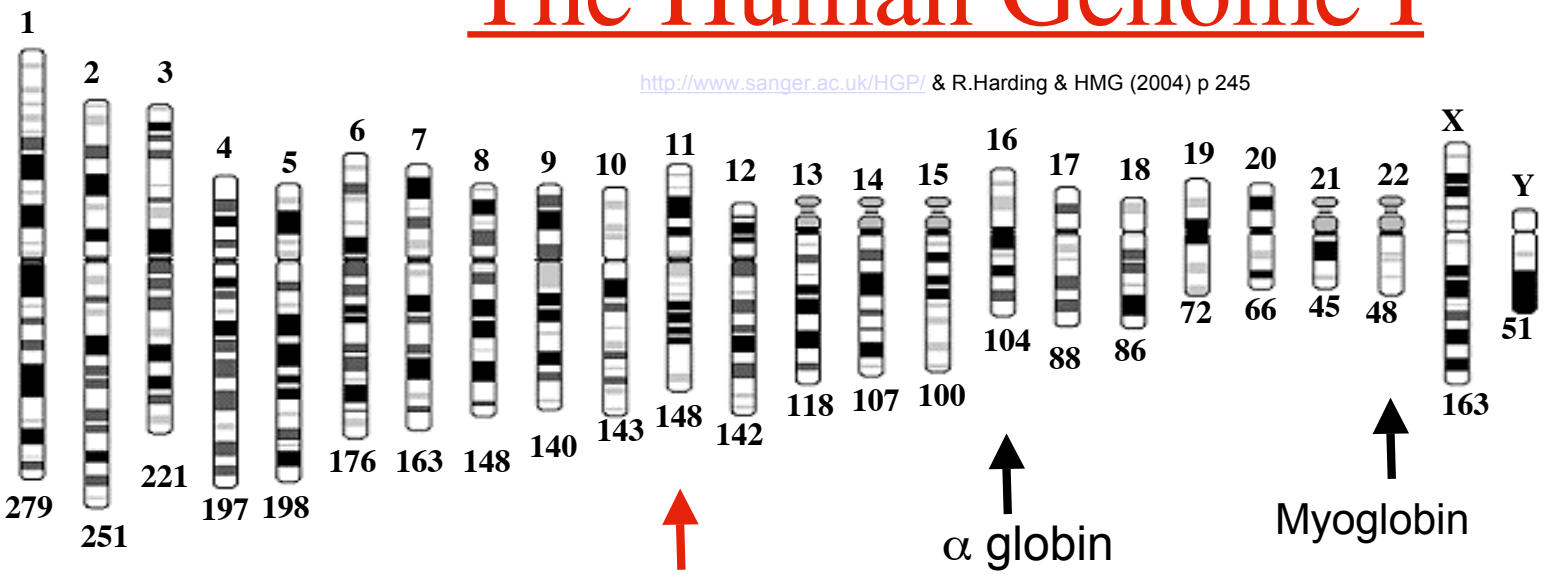
Strongylocentrotus purpuratus	800 Mb

Multicellular Plants

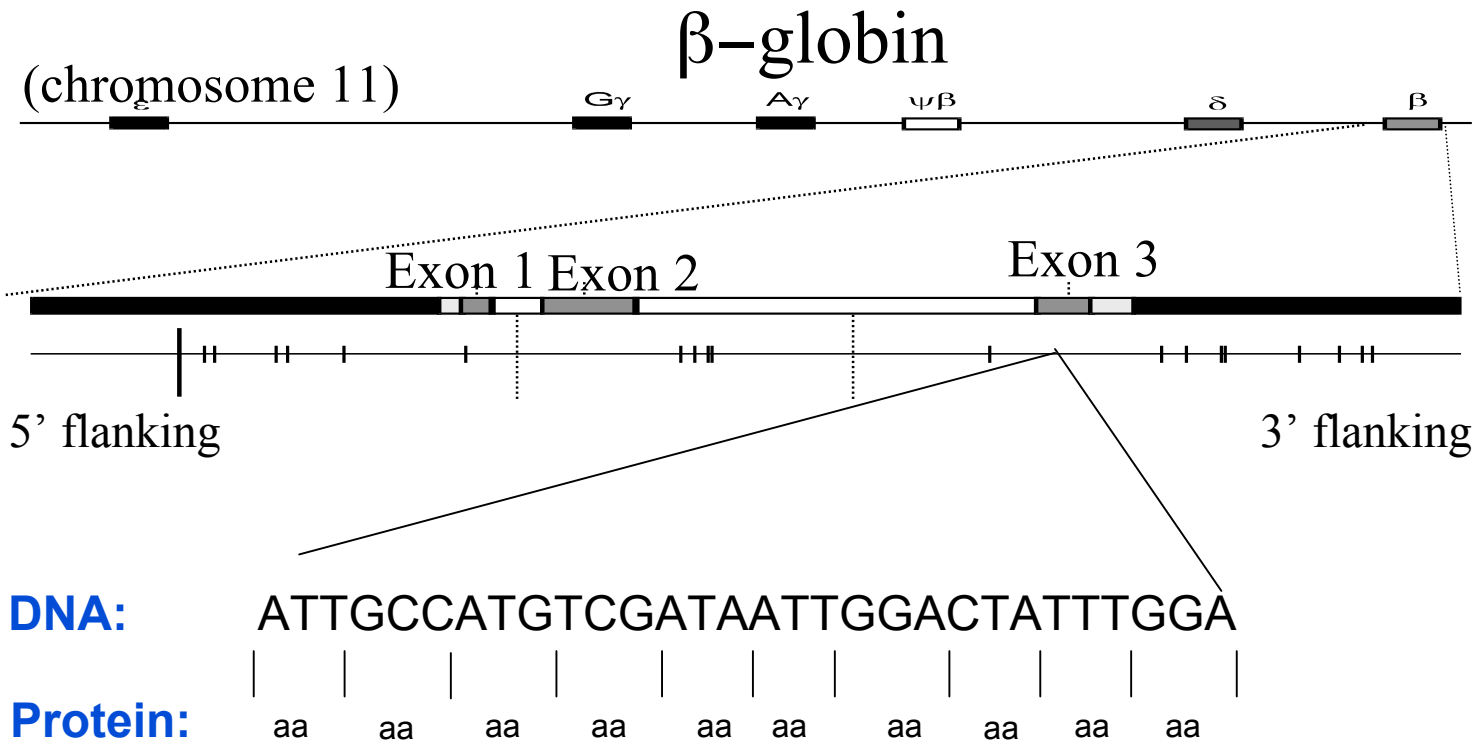
Arabidopsis thaliana	125 Mb
Rice	430 Mb

Strachan and Read (2004) Chapter 8

The Human Genome I



http://www.sanger.ac.uk/HGP/ & R.Harding & HMG (2004) p 245



mitochondria

.016

$3.2*10^9$ bp

*5.000

$6*10^4$ bp

*20

$3*10^3$ bp

$*10^3$

30 bp

The Human Genome II

http://www.sanger.ac.uk/HGP/

	Nuclear Genome	Mitochondria
Highly conserved - coding	1.5%	93%
Highly conserved - other	3.5%	5%
Transposon based repeats	45 %	-
Heterochromatin	6.6%	-
Other non-conserved	44 %	2%
	Mendelian inheritance	Maternal inheritance
	1 (typically)	Possibly thousands
	Recombination	No recombination
Gene Density:	1/130 kb	2 kb

Pseudogenes: 20000

Processed Pseudogenes

The Human Genome III

http://www.sanger.ac.uk/HGP/

Gene families

Clustered

α-globins (7), growth hormone (5), Class I HLA heavy chain (20),….

Dispersed

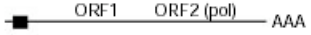
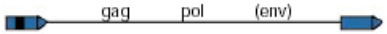
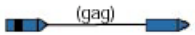
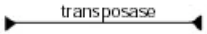
Pyruvate dehydrogenase (2), Aldolase (5), PAX (>12),..

Clustered and Dispersed

HOX (38 – 4), Histones (61 – 2), Olfactory receptors (>900 – 25),…

Transposons

Classes of interspersed repeat in the human genome

			Length	Copy number	Fraction of genome
LINEs	Autonomous		6–8 kb	850,000	21%
SINEs	Non-autonomous		100–300 bp	1,500,000	13%
Retrovirus-like elements	Autonomous		6–11 kb	450,000	8%
	Non-autonomous		1.5–3 kb		
DNA transposon fossils	Autonomous		2–3 kb	300,000	3%
	Non-autonomous		80–3,000 bp		

Strachan and Read (2004) Chapter 9 + Lander et al.(2001)

Genes and Gene Structures I

- Presently estimated Gene Number: 24.000 (reference:)
- Average Gene Size: 27 kb
- The largest gene: Dystrophin 2.4 Mb - 0.6% coding – 16 hours to transcribe.
- The shortest gene: $tRNA^{TYR}$ 100% coding
- Largest exon: ApoB exon 26 is 7.6 kb Smallest: <10bp
- Average exon number: 9
- Largest exon number: Titin 363 Smallest: 1
- Largest intron: WWOX intron 8 is 800 kb Smallest: 10s of bp
- Largest polypeptide: Titin 38.138 smallest: tens – small hormones.
- Intronless Genes: mitochondrial genes, many RNA genes, Interferons, Histones,..

Jobling, Hurles & Tyler-Smith (2004) HEG p 29 + HMG chapt. 9

Genes and Gene Structures II

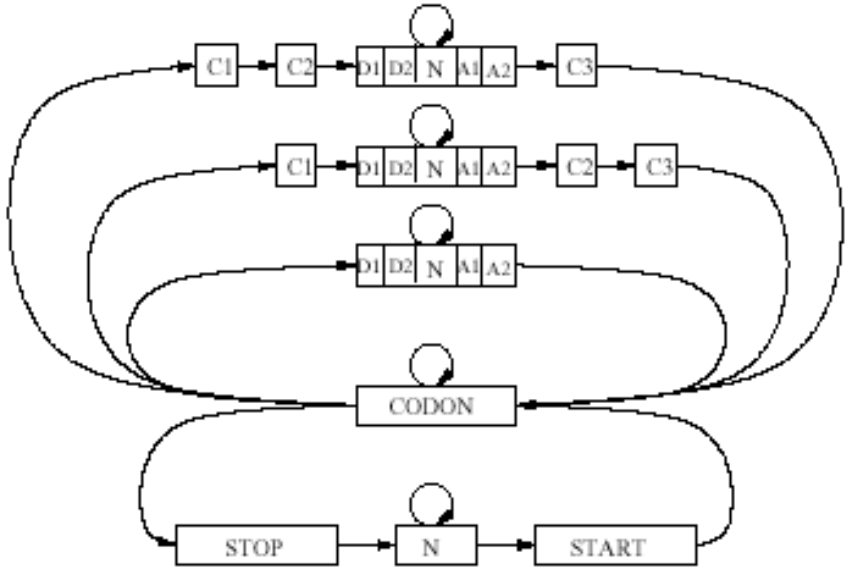
Genes within Genes:

Intron 26 of neurofibromatosis type I (NF1) contains 3 internal (2 exons) genes in the opposite direction.

Overlapping Genes:

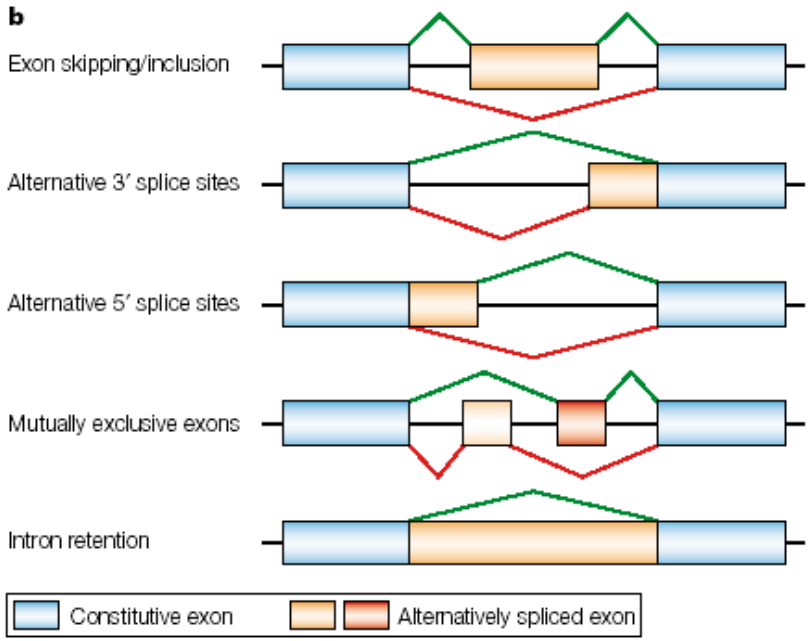
Class III region of HLA

Simple Eukaryotic



Strachan and Read (2004) Chapter 9 p 258

Alternative Splicing



1. A challenge to automated annotation.
2. How widespread is it?
3. Is it always functional?
4. How does it evolve?

Cartegni,L. et al.(2002) “Listening to Silence and understanding nonsense: Exonic mutations that affect splicing” Nature Reviews Genetics 3.4.285-

RNAs in the Genome

~200	snoRNA	small nucleolar, over 100 types - RNA modification and processing
~100	snRNA	small nuclear - involved in splicing
~200	miRNA	very small ~22bp , regulation
~175	28S,5.8S,5S	large cytosolic subunit
~175	18S	small mitochondrial subunit
~250	5S	large mitochondrial subunit
>500	tRNA	transfer RNA
>1500	Antisense RNA	> 1500 types
≈10000	non-coding RNA	regulatory functions?

Strachan and Read (2004) p.247 F9.4

Applications of Ultra-Low-Cost Sequencing

- Sequencing individual human genomes as component of preventive care.
- Genotype-phenotype associations
- Comprehensive gene expression profiling in vitro and in situ at all stages of development of a multicellular organism
- Comprehensive analysis of mutations present in cancer clones.
- Mitochondrial heteroplasmy
- Microbial diversity (metagenomic studies)

Shendure N. Advanced sequencing technology: methods and goals. Nat Rev Gen 2004

Oligonucleotide Array Synthesis

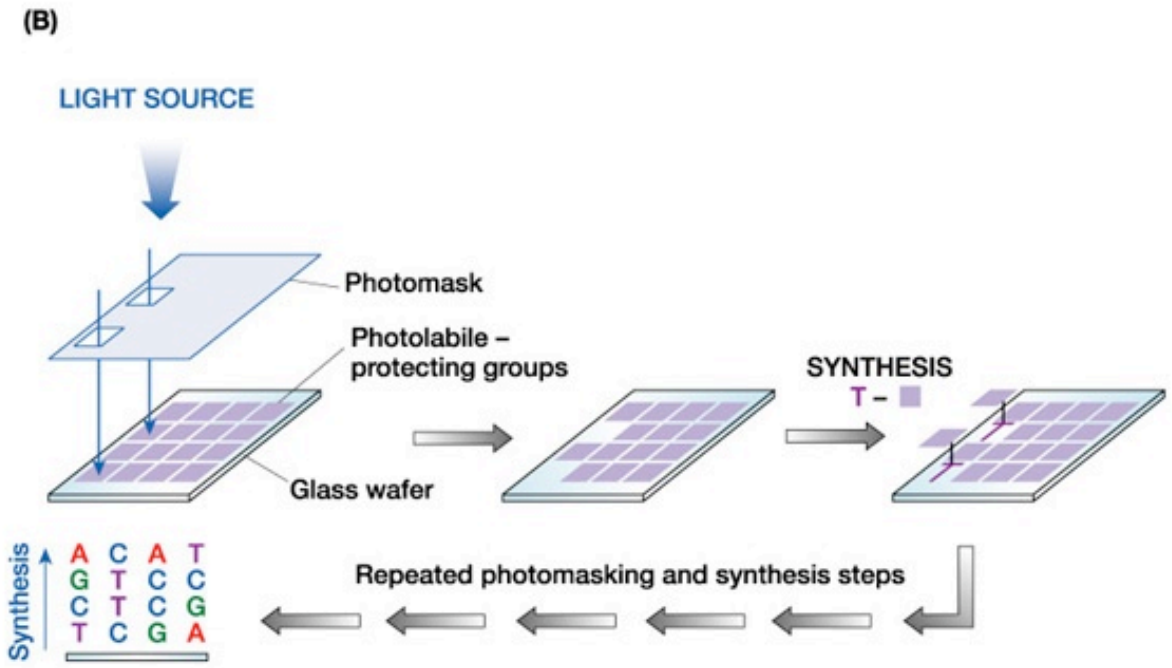
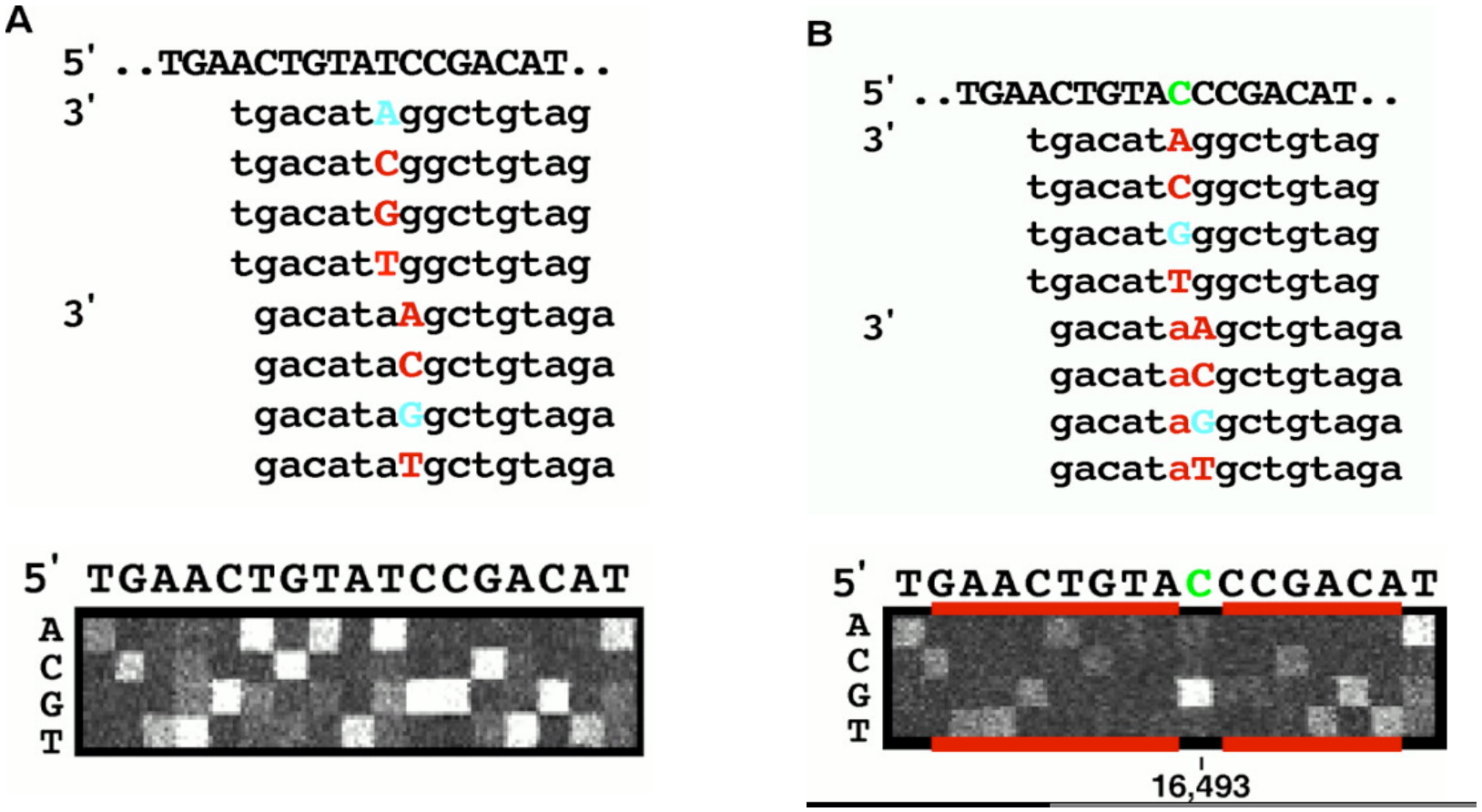


Figure 6-18 part 2 of 2 Human Molecular Genetics, 3/e. (© Garland Science 2004)

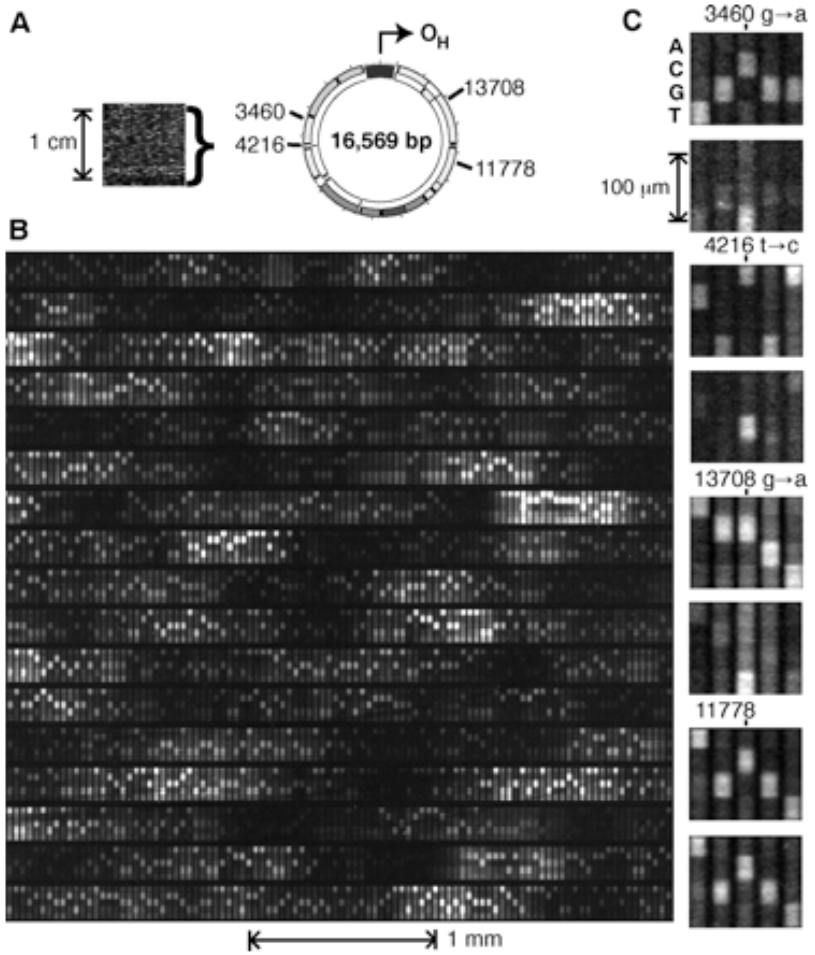
- Photolithography
- In situ synthesis
- Photolabile protective groups (photomasking(

High-Density Oligonucleotide Arrays



Chee M. Assessing genetic information with high-density oligonucleotide arrays. Science 1996

High-Density Oligonucleotide Arrays



Chee M. Assessing genetic information with high-density oligonucleotide arrays. Science 1996