BI 8 LECTURE 4

DNA APPROACHES: HOW WE KNOW WHAT WE KNOW

Ellen Rothenberg

14 January 2016

Reading: from Alberts Ch. 8

Central concept: DNA or RNA polymer length as an identifying feature

- RNA has intrinsically distinctive length: a particular gene's RNA is only copied from a discrete fraction of genome
  - After splicing, most eukaryotic RNAs are ~600-4000 nt long
  - Each gene's RNA has distinctive length
  - Helps to define that RNA: presence, intactness, & abundance
- DNA is intrinsically extremely long (100's of million nucleotide pairs in higher animals) with no breaks between genes
  - Thousands of genes are linked together physically along one duplex of DNA
  - However, you can convert local *sequence* information into another dimension of *length*
  - Sequence specific cutting
  - Sequence specific replication

Complexity: Meaning vs. mass
DNA as biochemical "stuff" or DNA as a particular code?

- Mass of total DNA, total base pairs in a sample
  - 1 bp: mw ~660

  … That is, 1 Mole bp: ~660g

  (Remember, Avogadro's #: ~6×10^{23} molecules/mole)
  - 10^{15} bp are 1.1 μg of DNA

- ***Crude mass has no sequence distinction***
- How much of this DNA consists of sequence of interest?

- Mass of a human genome: 3×10^{9} bp, ~3.3 pg (3.3×10^{-12}g) DNA per haploid genome (~6.6 pg/cell)

- For any unique coding sequence of 3 kb, mass per cell is ~3.3×10^{-18} g DNA/haploid genome…
- one *millionth* of genome

How can you locate, purify particular sequences??
Main elements of "DNA wrangling" strategy

- Use base complementarity as a method of identification for specific sequences: ***hybridization with specific probes***
- Use polymer length and structure (linear vs. circular) as a way to separate different classes of DNA or RNA *physically*
- Use DNA synthesizing enzymes to copy DNA, even from RNA
- Sequence-specific DNA cutting enzymes: enables dissection based on *sequence*, read out as *length*
- Use requirement for primers in DNA synthesis to target DNA copying to desired sequences

How do you “get” the DNA for the probe???

- To make a short probe for a sequence you already know, order it from a company and they will send it to you
- In small genome organisms, can start with restriction enzyme digested fragment of genomic DNA
- To make a probe for sequences in a particular RNA, make artificially replicated DNA from the purified RNA (cDNA)
- For others, need a way to obtain *homogeneous, large quantities* of particular sequence (reproducibly)

Restriction enzymes: highly reproducible sequence-based cutting of DNA with “worm’s eye view” specificity

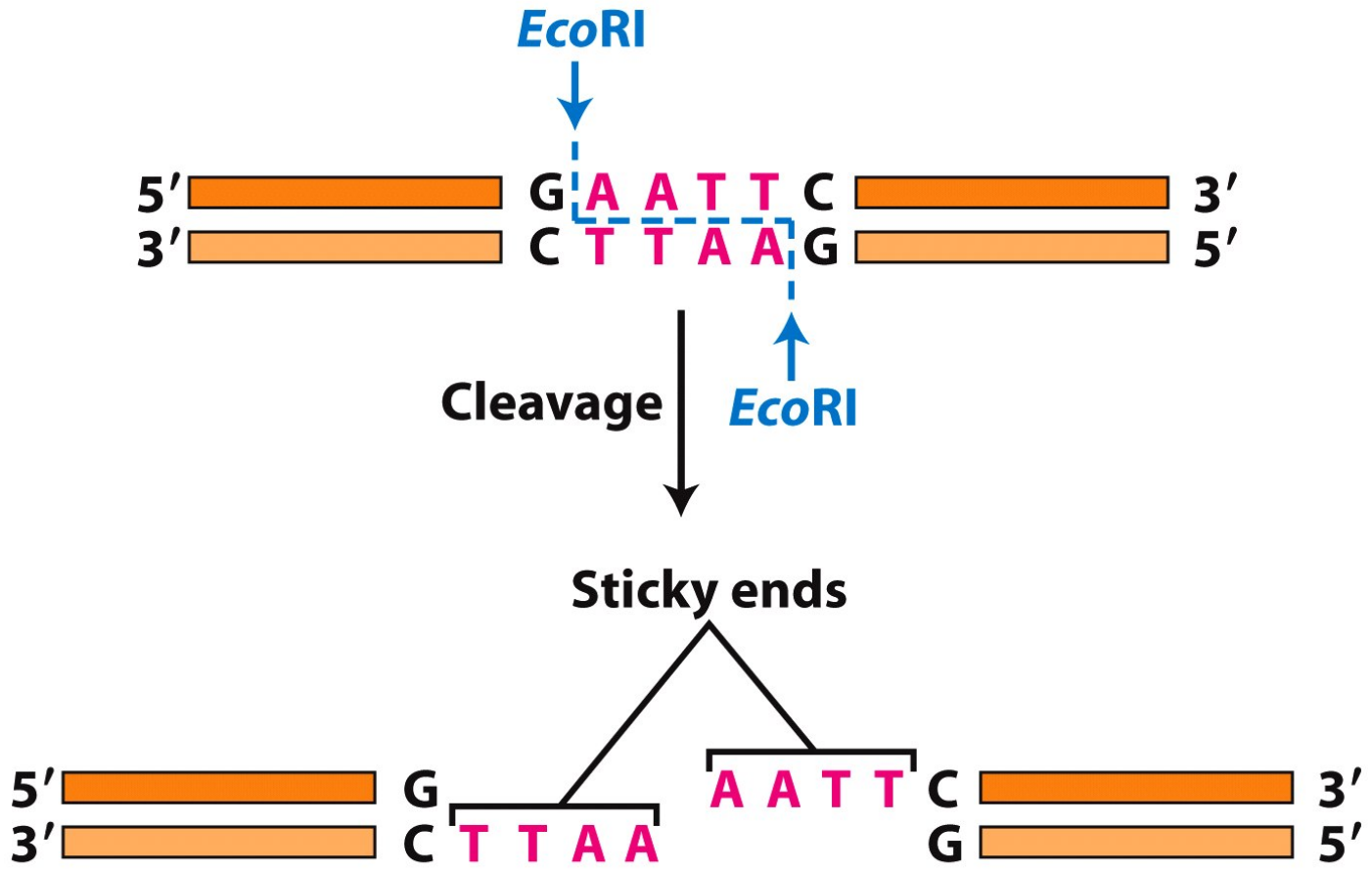
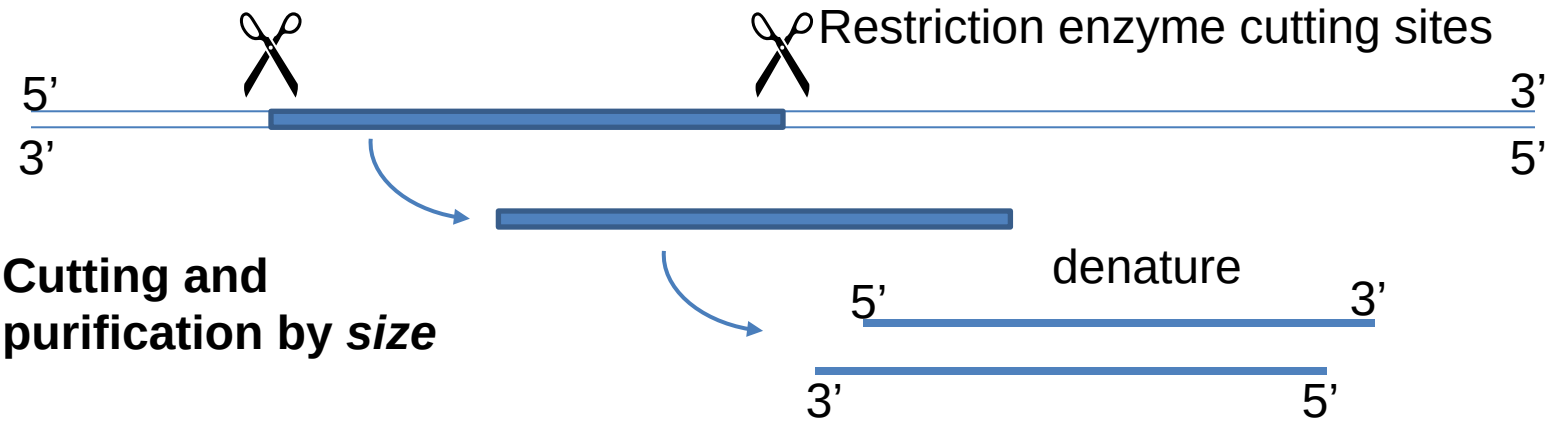


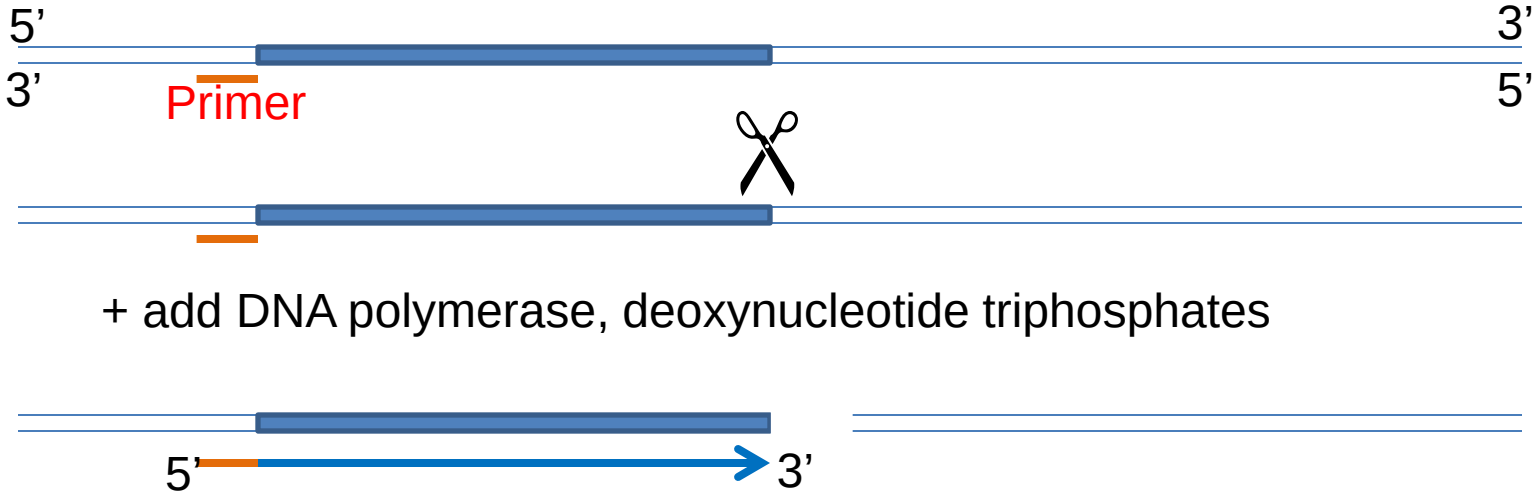
Figure 5-11
Molecular Cell Biology, Sixth Edition

Probes: sections of DNA with defined sequence and endpoints, that you can track through hybridization



Cutting and purification by *size*

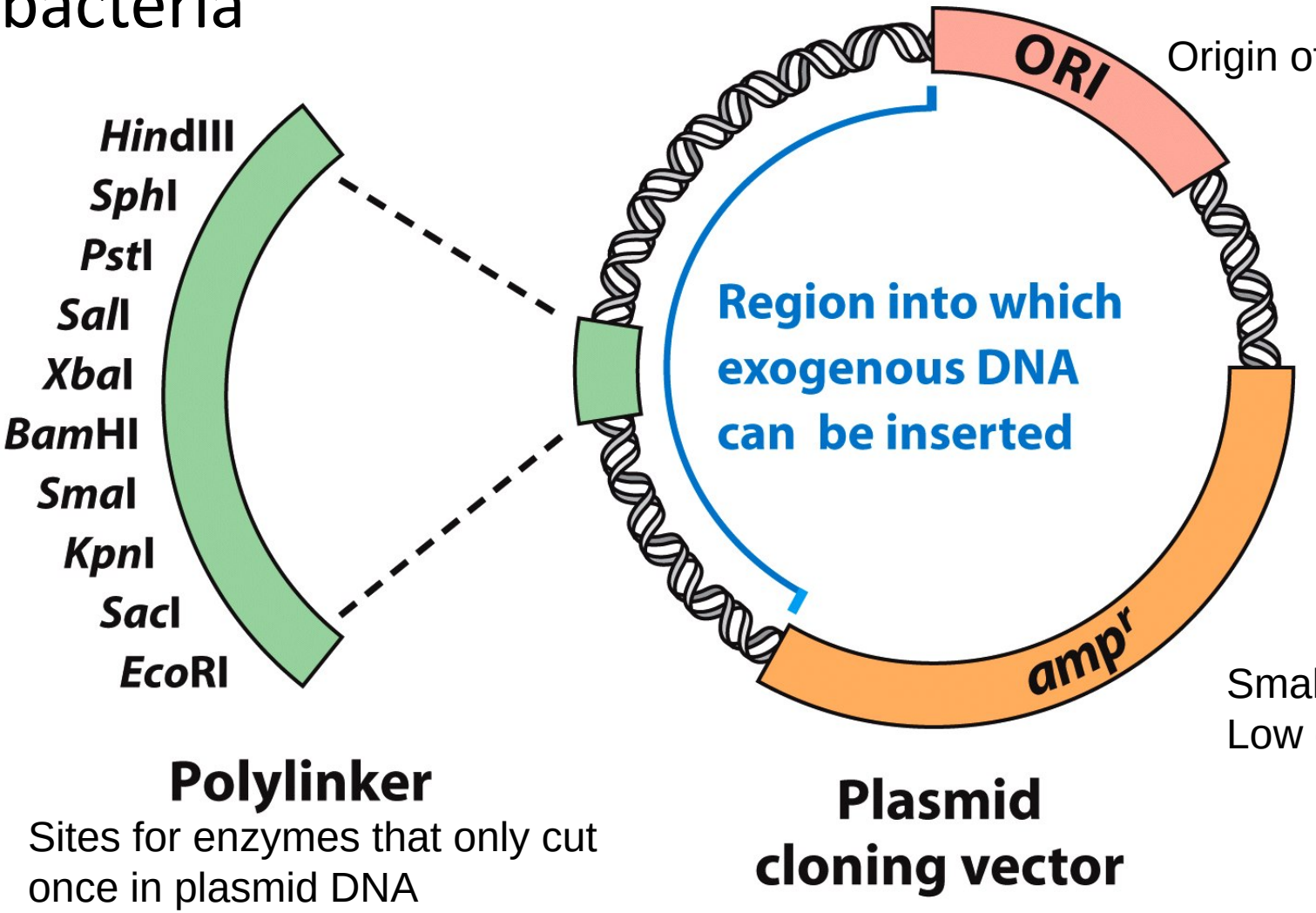
Purification by synthesis



Cloning: for purification, immortalization, and searching for a rare sequence

- Ability to use bacteria (or enzymes derived from them) to replicate 10^3- $>10^6$ copies of a single molecule
- Spread out individual bacteria to form separate colonies: physical separation, one colony = one cloned DNA sequence
  - Individual colonies can be grown up to massive numbers of copies
  - DNA sequence purification of large enough numbers of clones in parallel to look across whole genomes

Plasmids: small circular DNA molecules that are replicated by host DNA replication machinery in bacteria

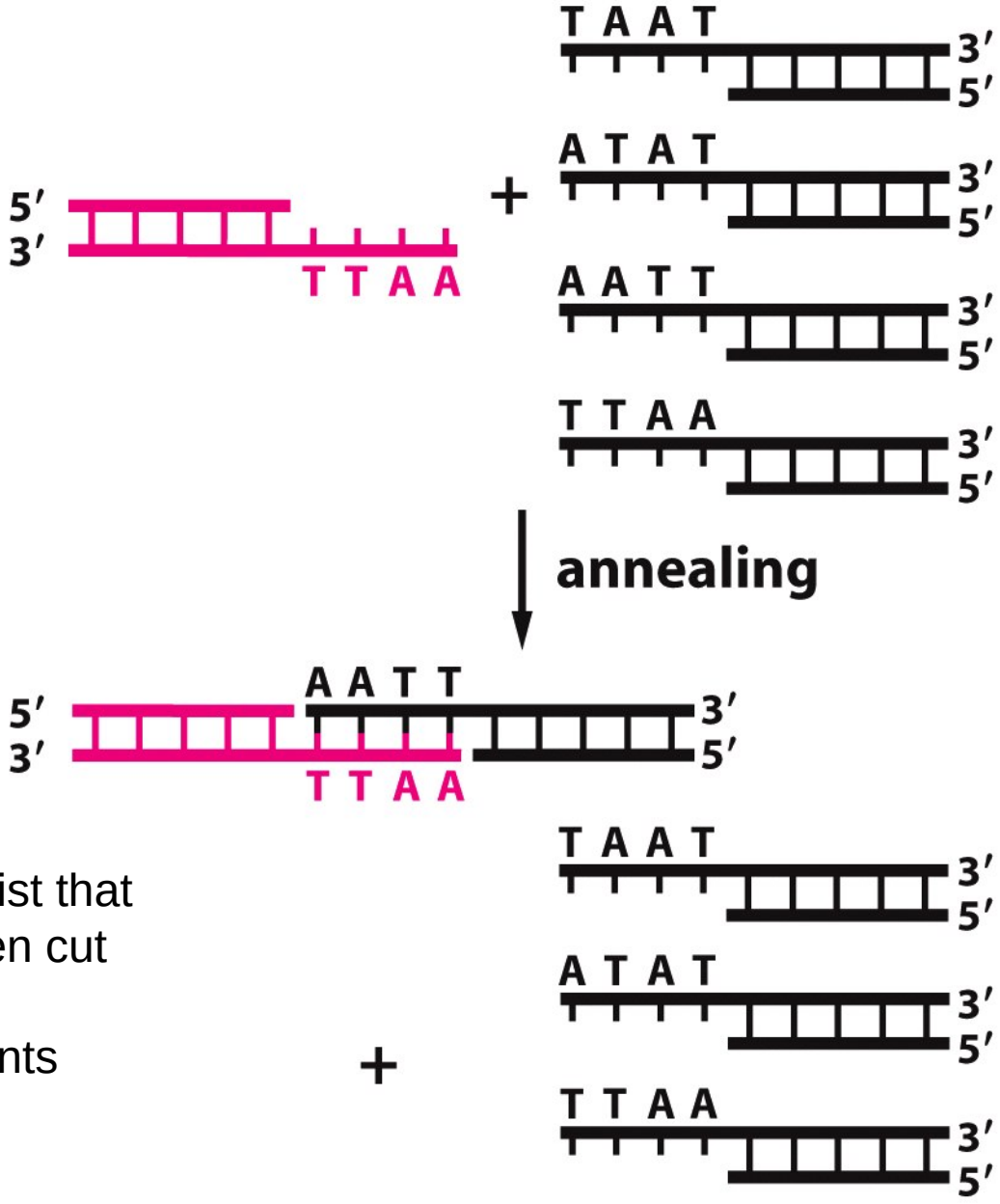


Origin of replication in bacteria

Small DNA: Low complexity

Figure 5-13
Molecular Cell Biology, Sixth Edition

Restriction enzyme-cleaved DNA with "sticky ends" can reanneal specifically to DNA with complementary sticky ends



… And enzymes (LIGASES) exist that can join DNA ends that had been cut by the DNA cutting enzymes, combining them to form new joints

Figure 8-32 *Molecular Biology of the Cell* (© Garland Science 2008)

Bacteria as DNA purification and amplification tools

One colony: one cloned DNA sequence, purified and immortally self-propagating

An array of individual colonies: an array of geometrically ordered, individually purified, immortally self-propagating DNA sequences

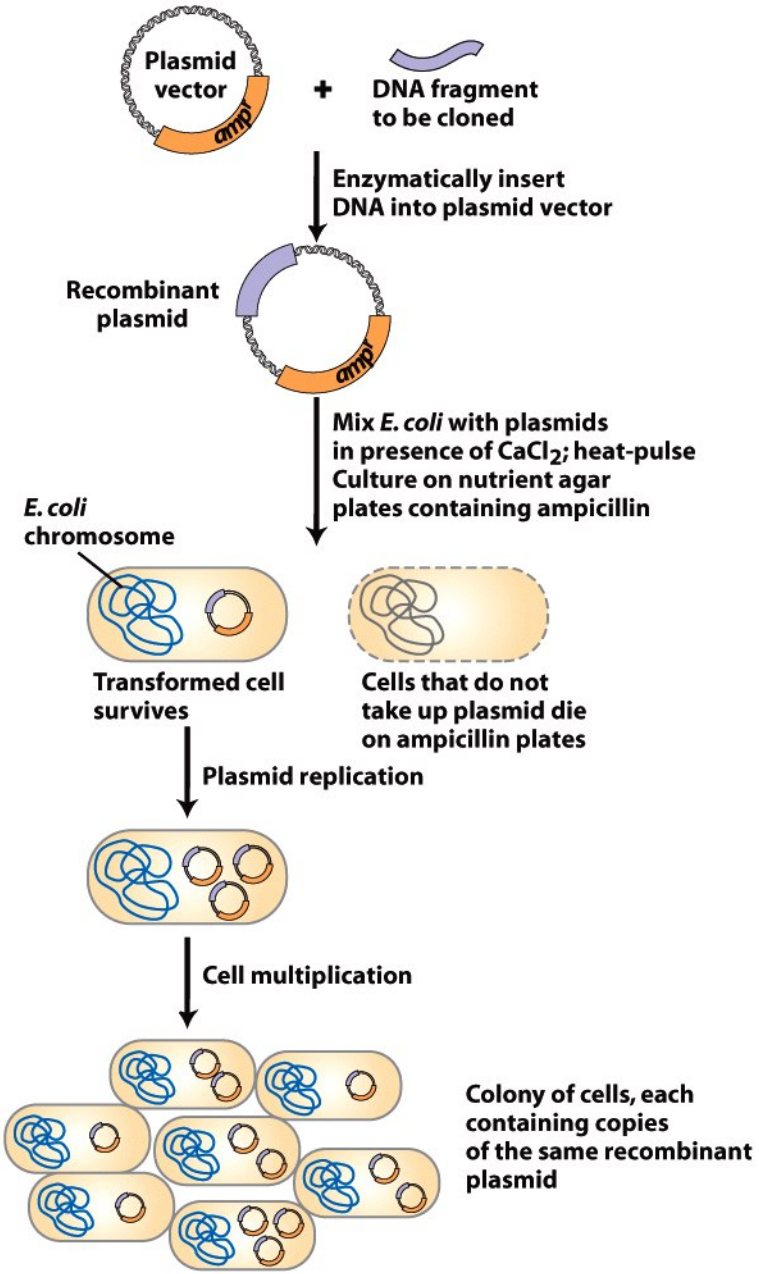


Figure 5-14
Molecular Cell Biology, Sixth Edition

Polymerase chain reaction uses enzymes that can enable us to “clone” a particular DNA sequence without using a live bacterium

- Depended on discovery of special, heat-tolerant DNA polymerases
- Depended on technology to make artificial DNA sequences at will – primers to start the DNA copying reaction in vitro
- Depended on knowing the sequence to make primers to!
- Depended on all these things being cheap

Making DNA, in vitro or in vivo, always depends on adding nucleotides to pre-existing primers

Corollary: only sequences for which the right complementary primers are present will get copied…..
*So **primer sequence design** determines what DNA will get replicated*
High specificity even in a highly complex sample!

Requirement #1: polymerase
Requirement #2: the right specific primers in enormous, vast excess

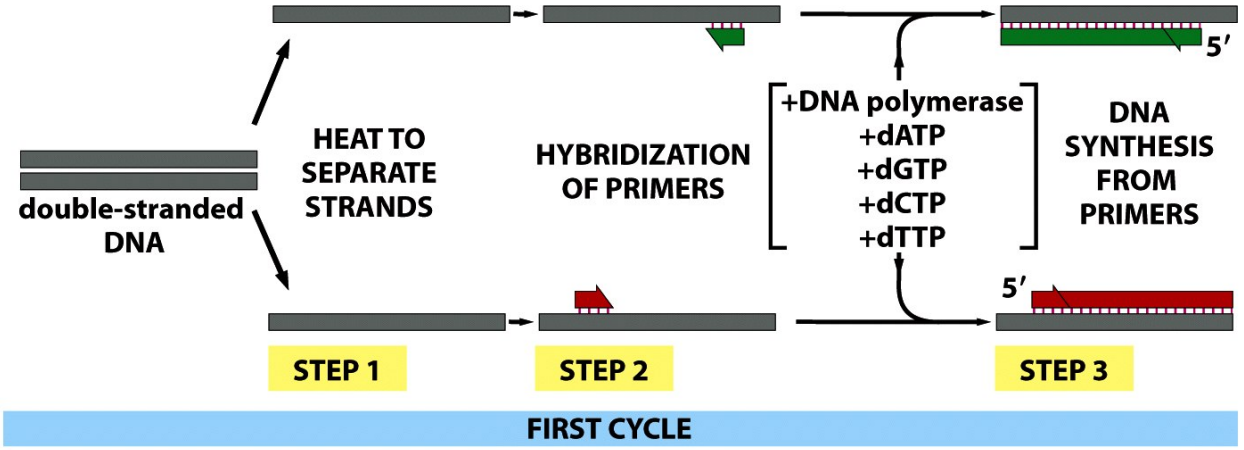


Figure 8-45a *Molecular Biology of the Cell* (© Garland Science 2008)

Exponential amplification as long as primers are still in excess… >2^{30}-fold amplification of the gene you want to isolate is possible

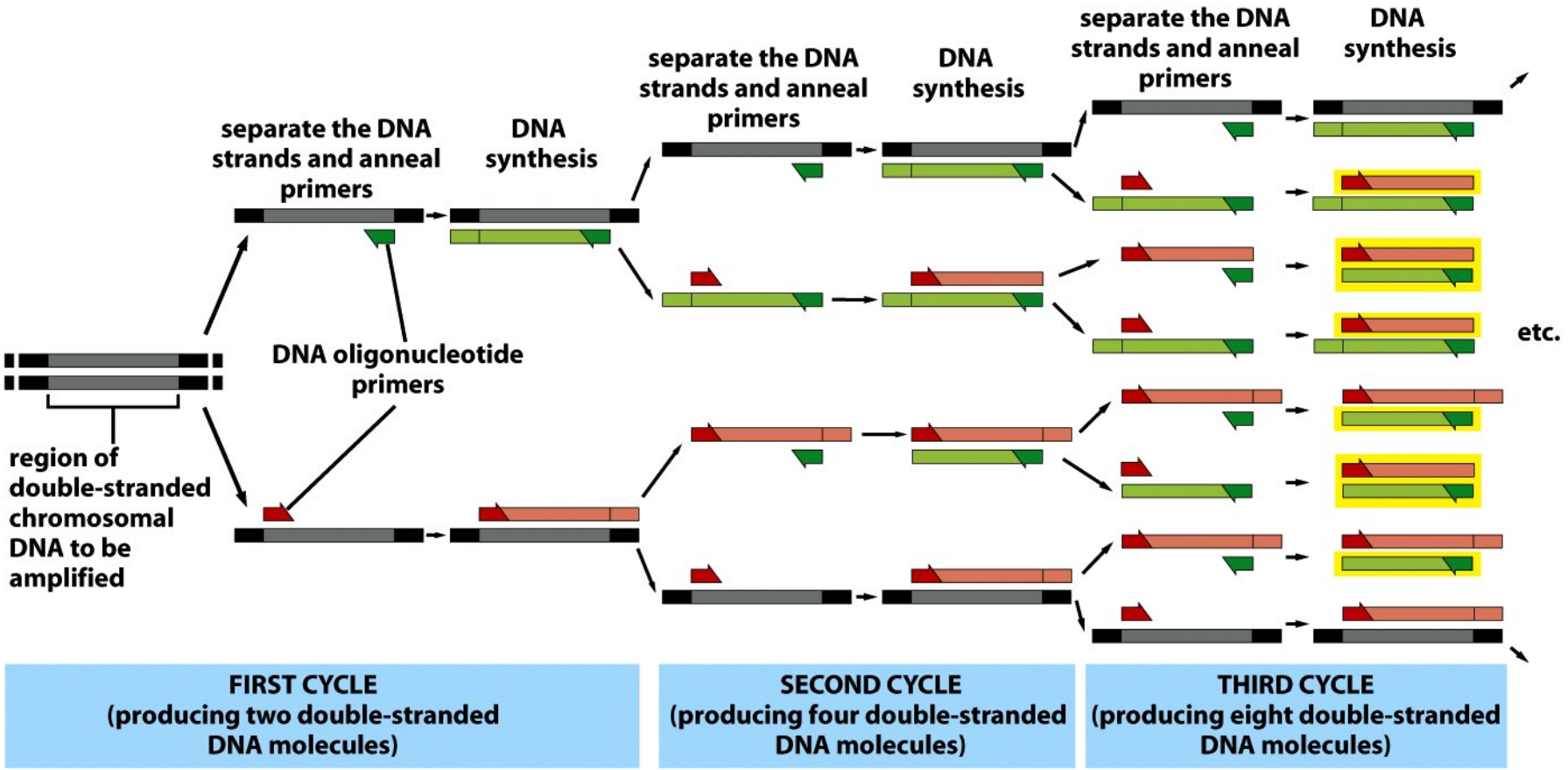


Figure 8-45b *Molecular Biology of the Cell* (© Garland Science 2008)

Features built into PCR primers can add artificial restriction sites to the ends of a DNA sequence

Now you have millions or billions of copies of a pure sequence already modified to make it easy to study and manipulate

You can control it, as well as study it

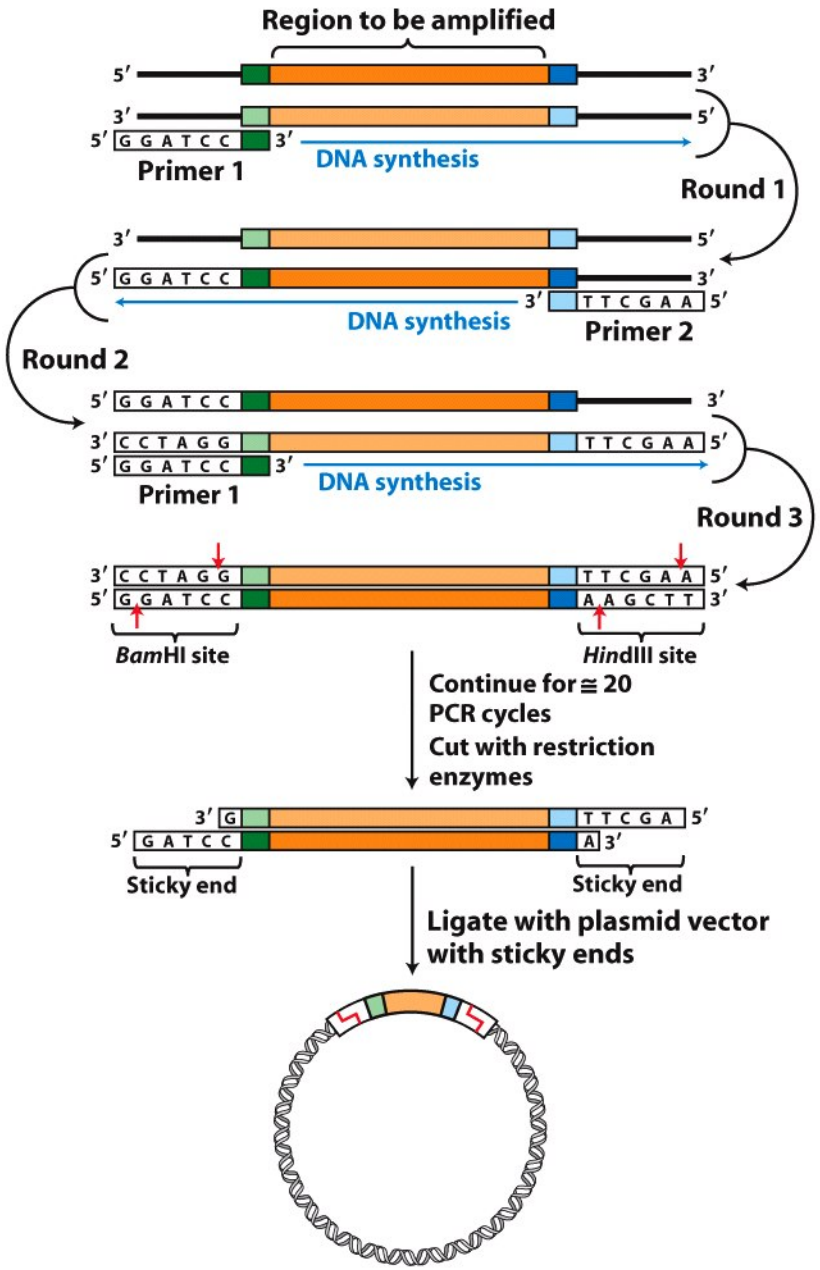


Figure 5-24
Molecular Cell Biology, Sixth Edition

Labeling is the easy part if you have the rest.

Make a copy of the DNA you care about using chemically tagged or radiolabeled nucleotide components

label is now incorporated into a “probe”

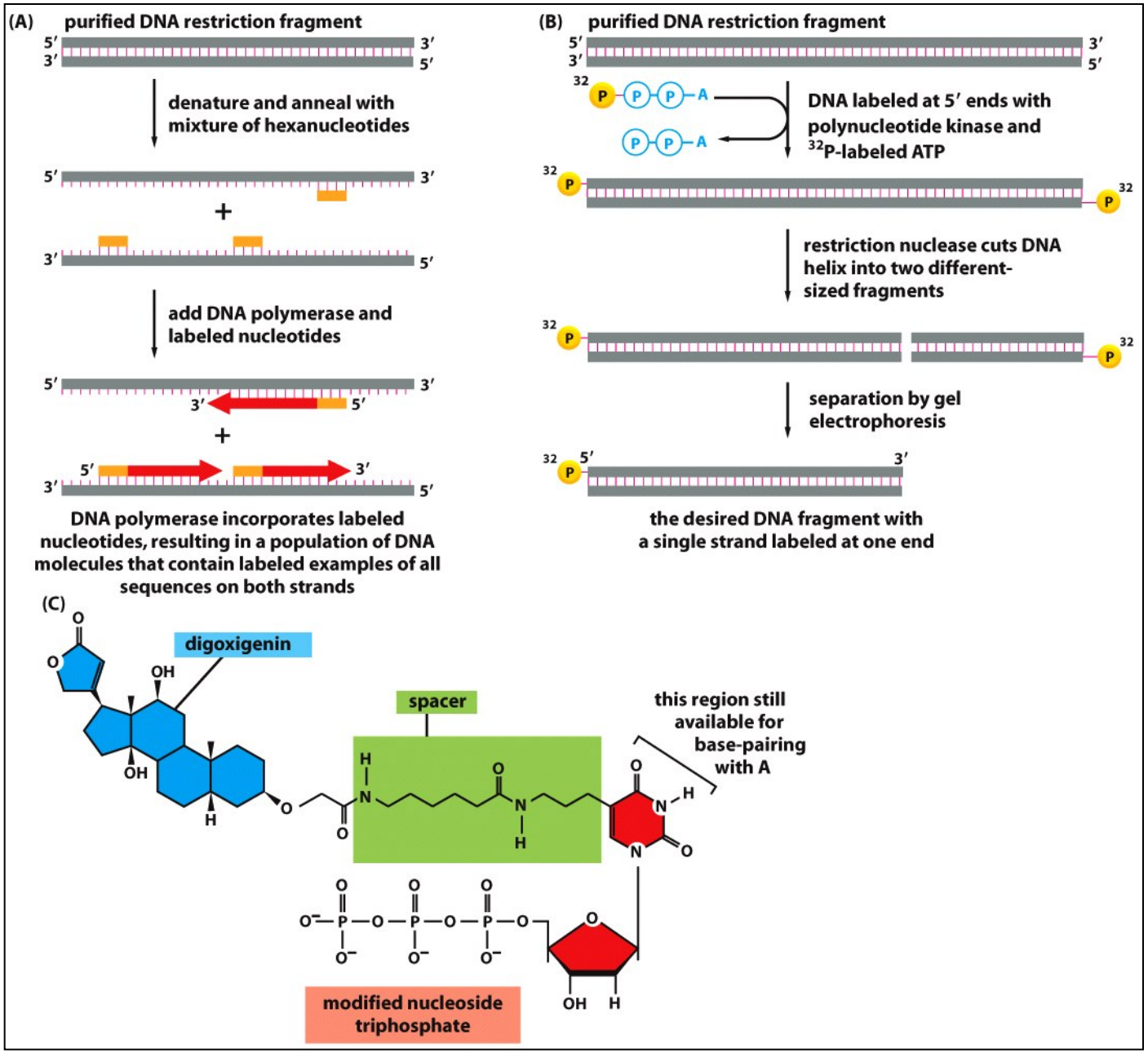


Figure 8-34 *Molecular Biology of the Cell* (© Garland Science 2008)

Combination of sequence location together with DNA or RNA length is a powerful analytical tool

- Infer structure of DNA or RNA relative to defined sequence “landmarks”

- Read DNA sequence by length/ composition relationship from a fixed starting point

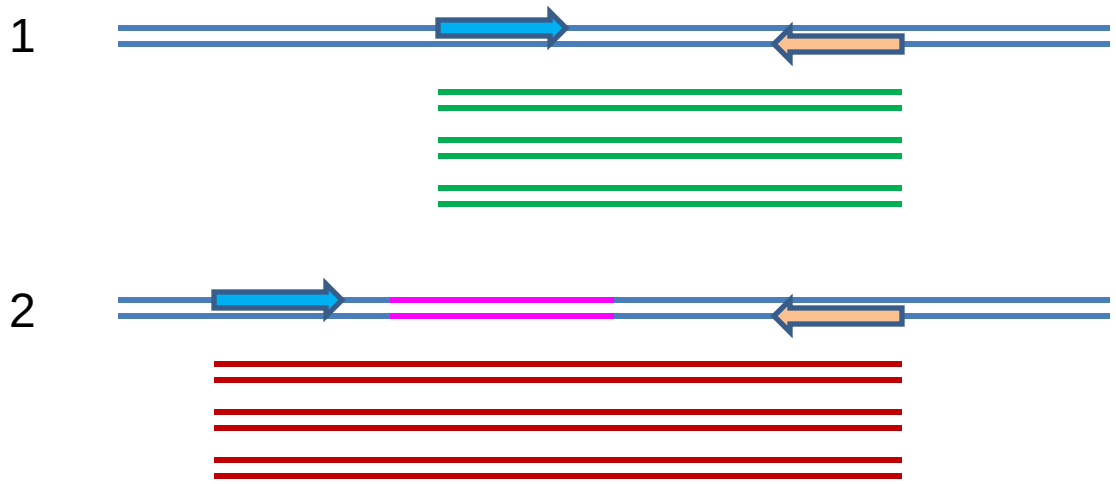
- Need to physically separate molecules of different sizes (gel electrophoresis)
- Need to define a known endpoint by restriction digestion or primer hybridization

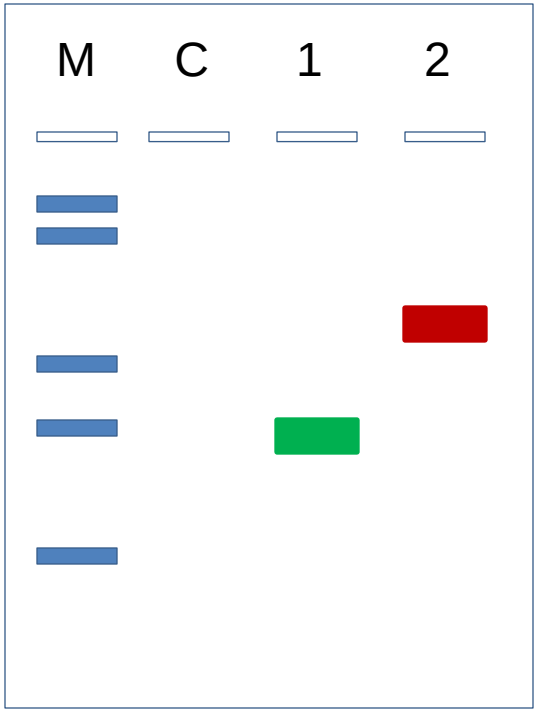
Analytical use of PCR:
Using size plus primer sequence complementarity with PCR to dissect DNA/RNA structures

1

2

Using size plus primer sequence complementarity with PCR to dissect DNA/RNA structures

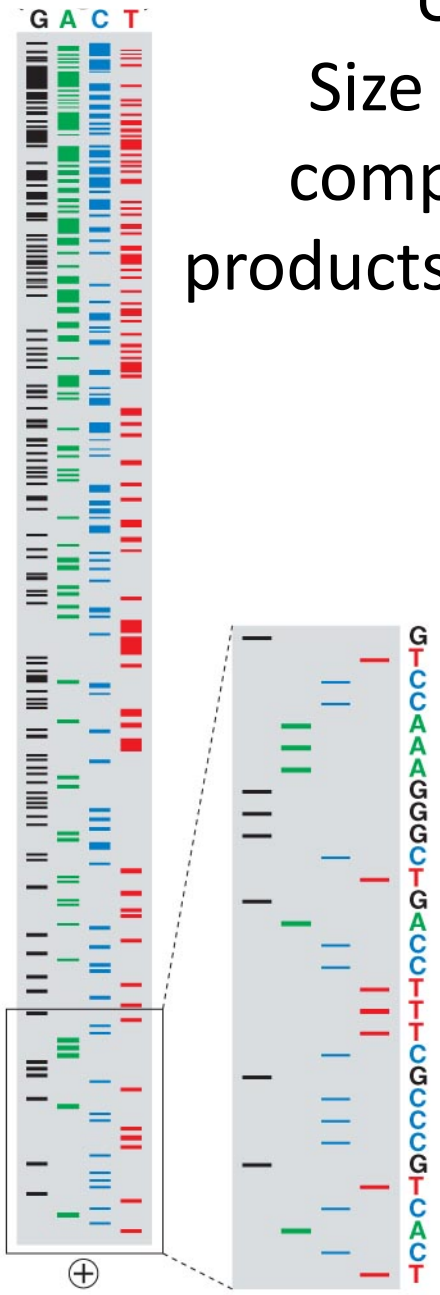




Infer the presence of the insertion in gene 2 based on size difference of the products amplified from the same primers, even before you know what the inserted sequence is

USE OF SIZE IN EARLY SEQUENCING:

Size separation is the basis for converting a complex mixture of aborted DNA replication products into an orderly "ladder" of DNA sequence



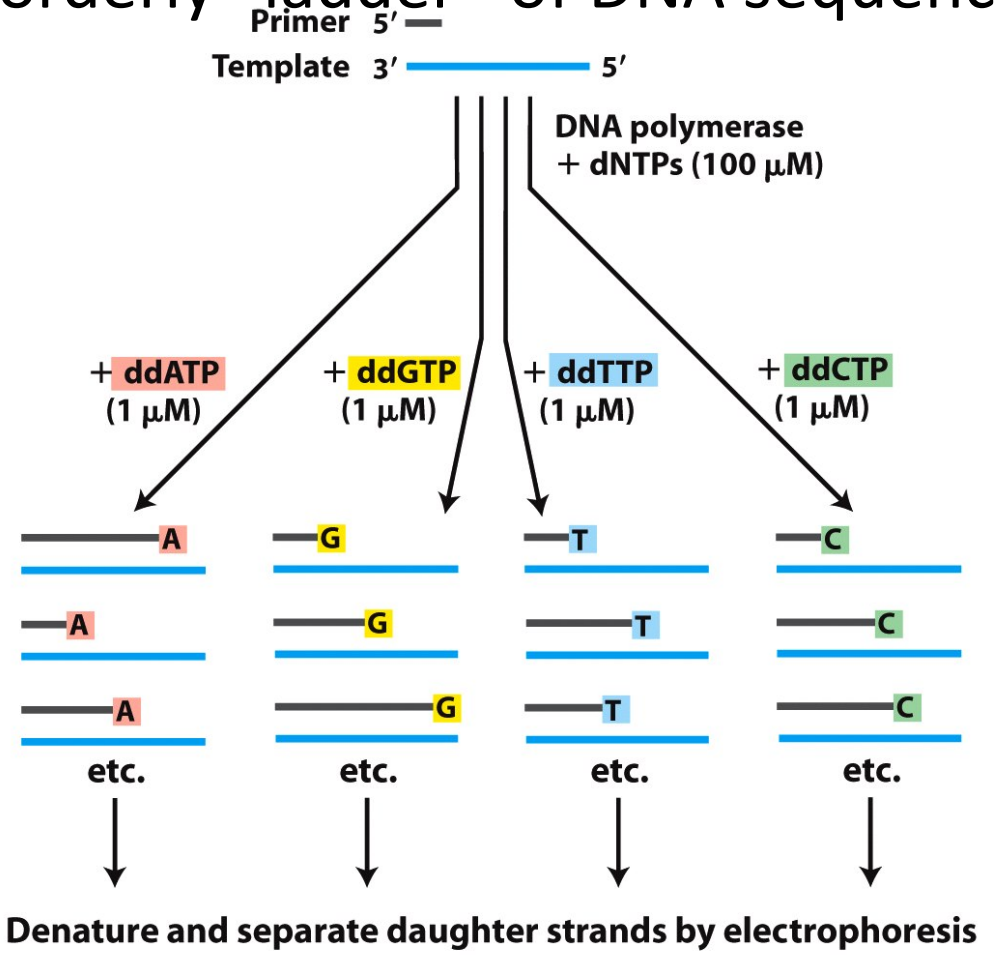


Figure 5-21b
Molecular Cell Biology, Sixth Edition

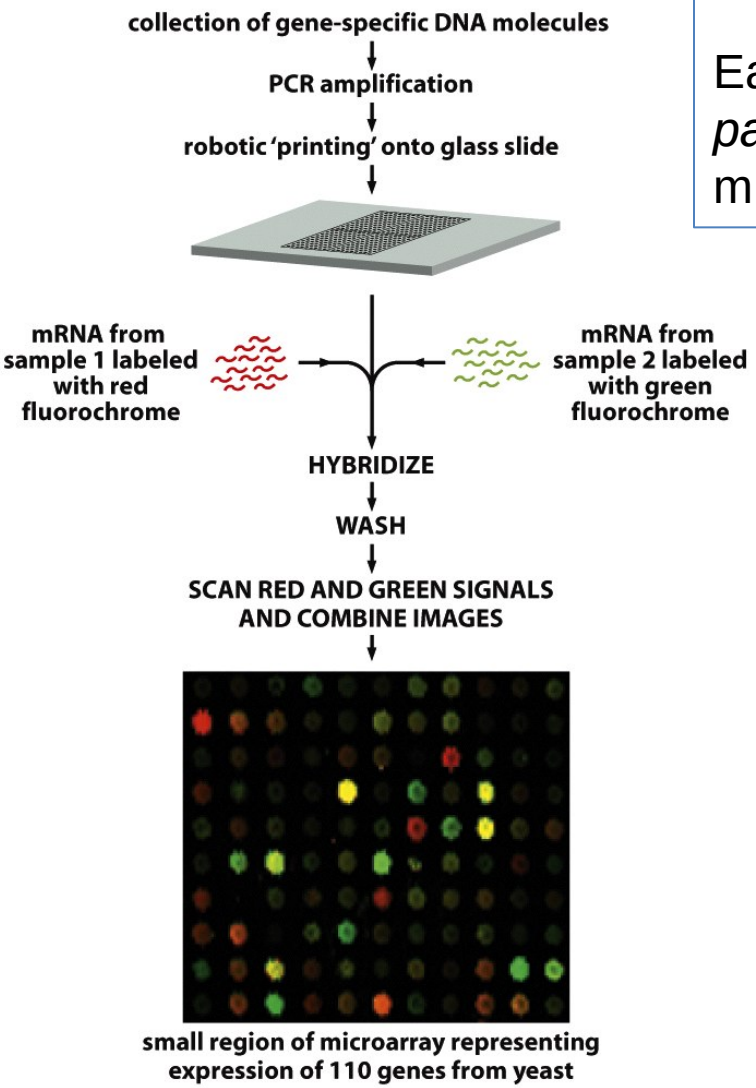
Analyzing the *expressed* parts of the genomes in different cells or conditions

- The RNA is what is different between cell types in an organism or conditions of response to a stimulus… DNA stays the same. So RNA often holds the answer to the question one is asking.
- But RNA has "issues":
  - Not very stable compared to DNA (easy to degrade randomly without intending to!)
  - Not double-stranded, and not as easy to cut at specific sequences or rejoin by enzymes as a result
  - Not many ways to promote RNA to be replicated to high purity and high copy number by itself
- Solution: make it into a DNA copy

Common methods of measuring RNA expression

Once mRNA sequences are known, probes for mRNAs of interest can be designed and synthesized… MICROARRAYS

to use hybridization to measure levels of multiple RNA species in biological samples at once



Short probes: Oligonucleotides

Each detect one *part* of target mRNA

Figure 8-73 *Molecular Biology of the Cell* (© Garland Science 2008)

But…RNA-seq and microarray hybridization may not tell you which isoforms are made… or even if they exist

Structures come from cDNA clones

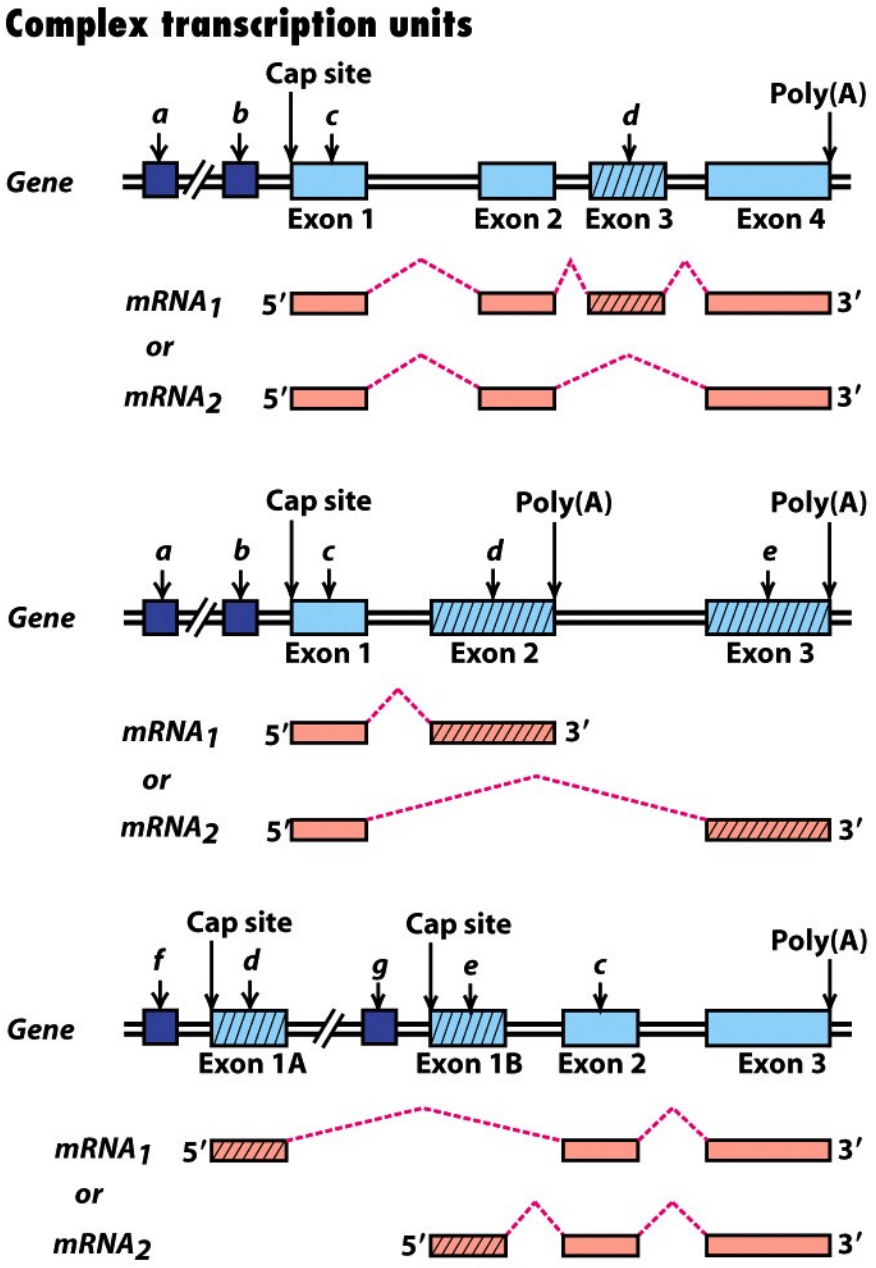


Figure 6-3b
Molecular Cell Biology, Sixth Edition

Central concept: DNA or RNA polymer length as an identifying feature

- RNA has intrinsically distinctive length: a particular gene's RNA is only copied from a discrete fraction of genome
  - After splicing, most eukaryotic RNAs are only 600-4000 nucleotides long
  - Each gene's RNA has distinctive length
- Separation of a mixture of RNAs based on size gives a second dimension of identification in addition to reaction with a probe
- DNA is intrinsically extremely long (~10^8 nucleotide pairs in higher animals) with no physical breaks between genes

Analyzing expression of RNAs of different structures

- Cloned, long probe bridging feature to assay
  - ssDNA form
  - RNA form can also be made
- Hybridize with target RNA
- Measure not only *amounts* but also *length(s)* of hybridized probe fragments
- Can use position of label in probe to get more info…
  - End labeling
  - Uniform labeling

Classic intron mapping with a single-stranded DNA probe

Shows presence of sequences from both exons in RNA and lengths of sequence expressed from both

This hybridization to labeled genomic DNA does not directly show that these sequences are linked in RNA

But hybridization of cloned cDNA to genomic DNA – the reverse experiment -- will show that they were linked in RNA! (Consider also how you could use PCR…)

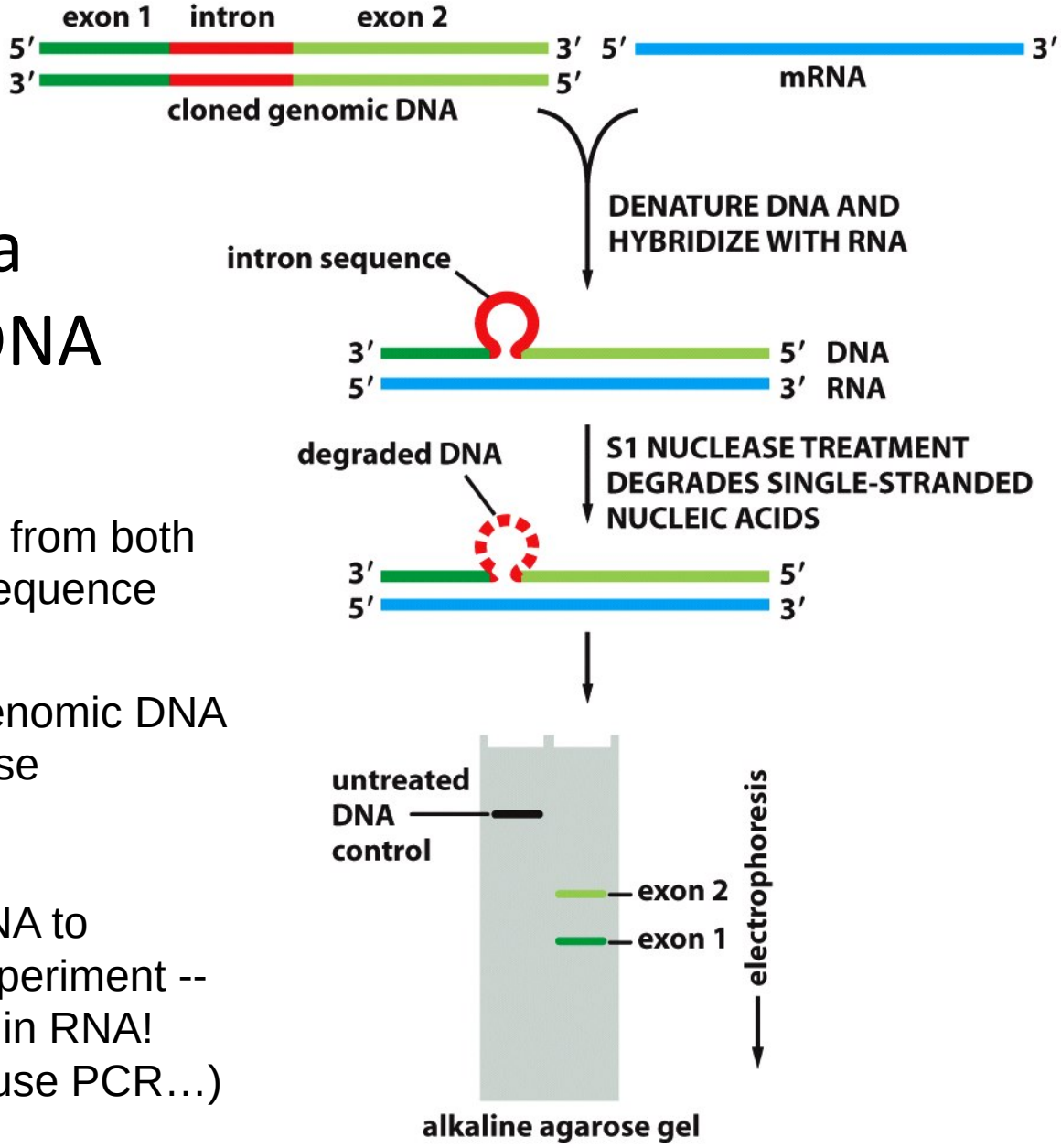


Figure 8-37 *Molecular Biology of the Cell* (© Garland Science 2008)

Defining a probe is defining the biological question

- Known gene of interest & known sequence → an artificial DNA probe can be synthesized to order (cheap now)
- A known gene or artificial DNA copy of an RNA molecule (cDNA) can be used as a probe in its entirety
  - (even if you *don't* know its sequence yet)
  - Cloning the gene (to be described) helps produce billions of copies of the same gene to make purification and labeling very easy
- Special expression features of a gene that is NOT yet identified can be used to find a probe that will enable the gene to be identified and isolated
  - *without prior knowledge of sequence*
  - DIFFERENTIAL LIBRARY SCREENING → new gene identification

Both genomic and cDNA (mRNA proxy) DNA can be cloned in libraries

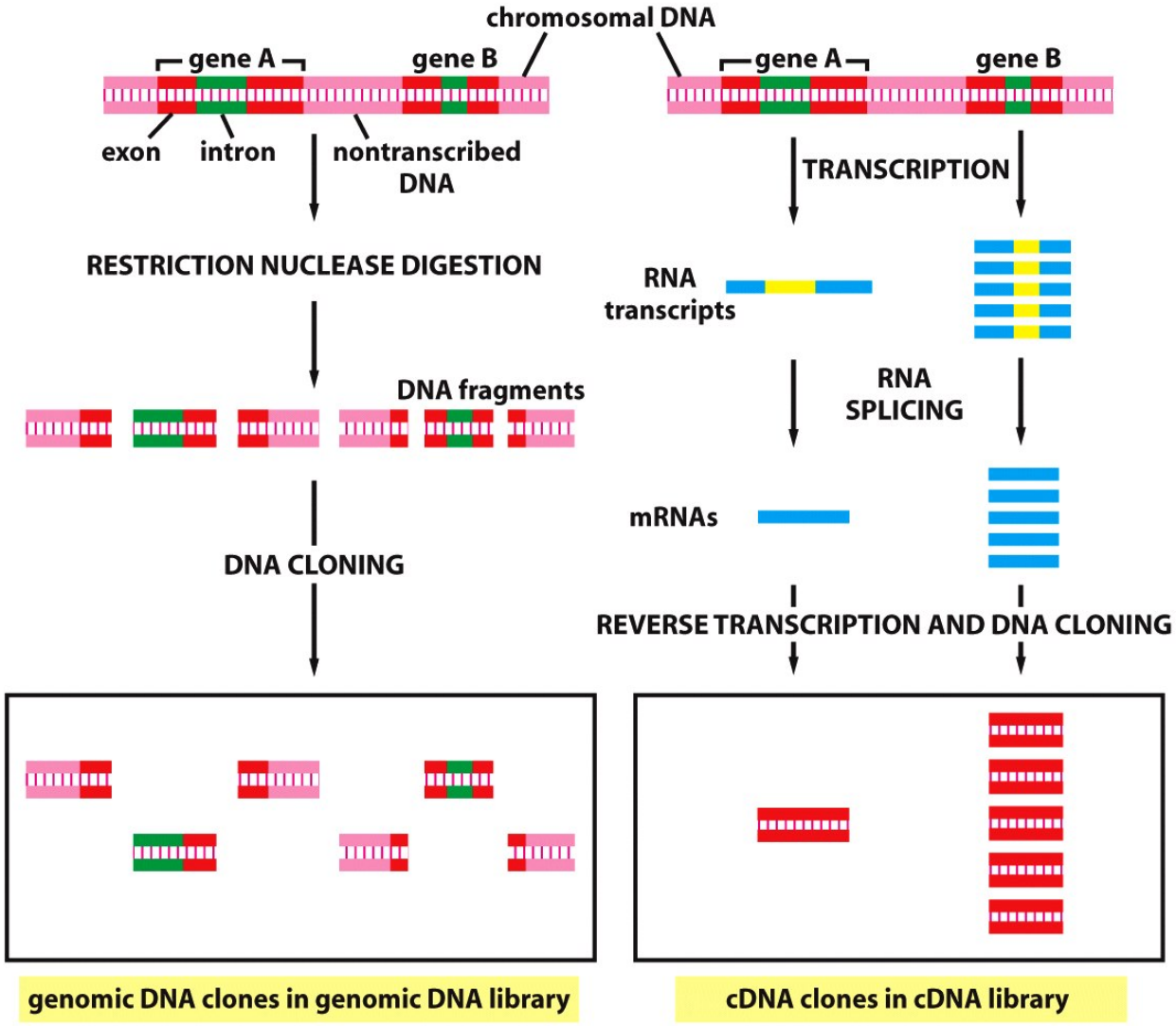
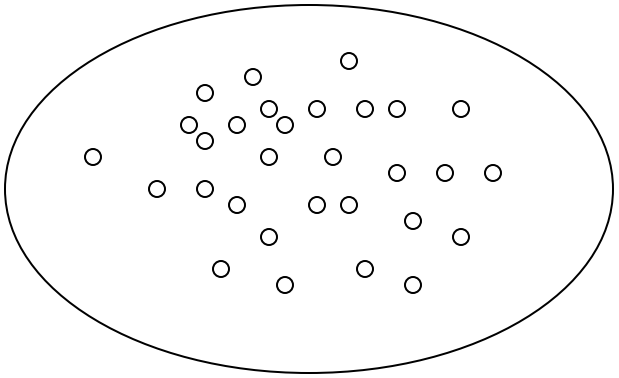


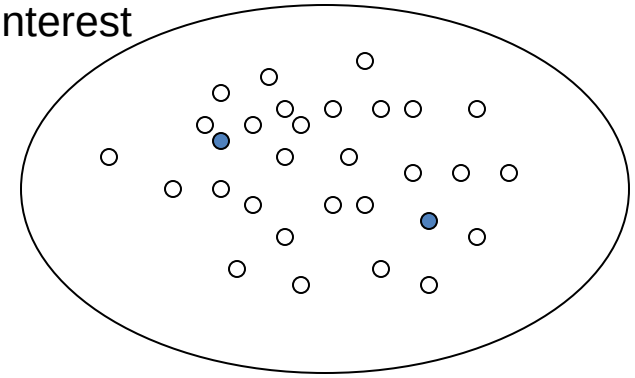
Figure 8-44 *Molecular Biology of the Cell* (© Garland Science 2008)

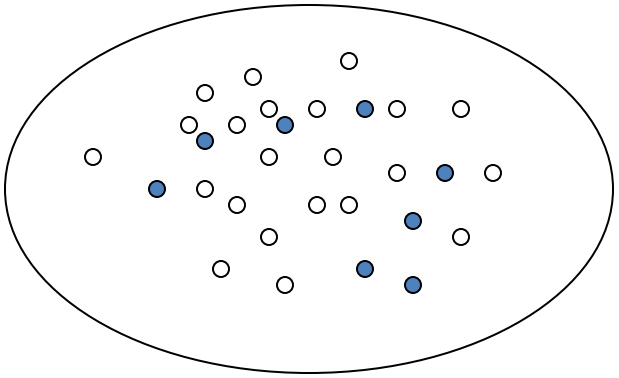
Cloning and screening a genomic or cDNA library with specific probes is the basis of specific gene discovery

Bacterial library colony DNA screened with negative control probe

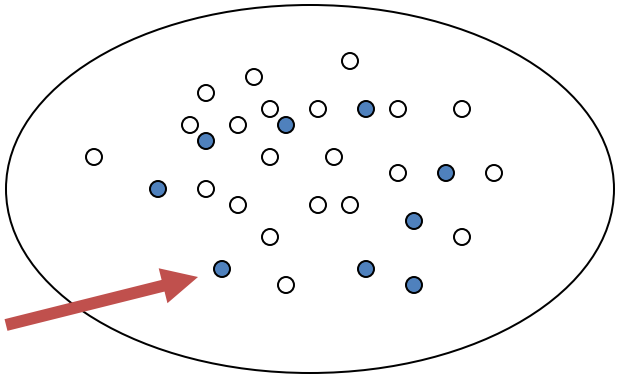


Bacterial library colony DNA screened with probe for a specific gene of interest





Bacterial library colony DNA screened with total cDNA from unstimulated cell (control cell)



Bacterial library colony DNA screened with total cDNA from stimulated cell (***stim specific cDNA in library can be identified***)

What we use molecular technology for

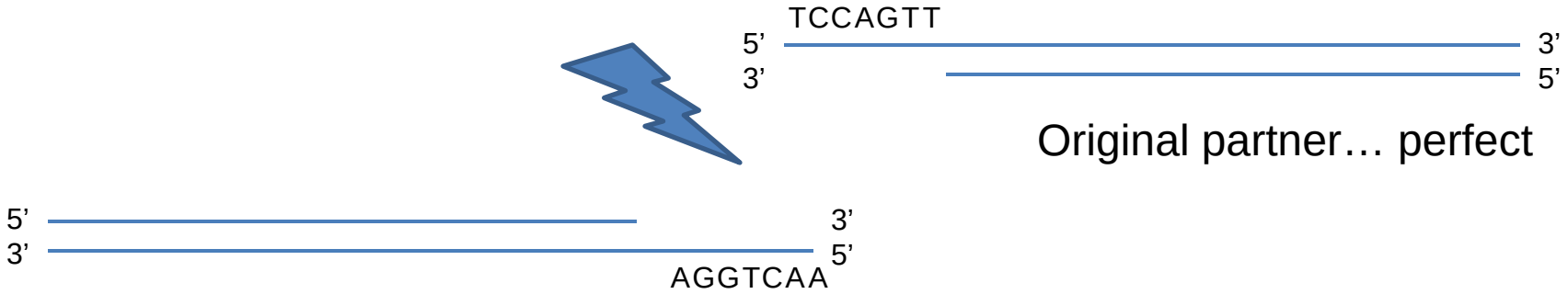
- Finding NEW genes that are expressed in particular cell types
  - DNA (genomically stored) versions
  - RNA (expressed) versions
- Finding all the genes an organism's genome includes and how they are organized
- Measuring the expression of different genes quantitatively
- Determining the sequence of a gene or genome
  - See what proteins and RNA structures it encodes
- Engineering ways to put one gene or gene variant under experimentally determined control
- Testing the effect on a cell or organism when a specific gene or gene regulatory sequence is deleted or mutated

Modifying genomes of cells

- Use sequence-specific recombination to introduce desired mutations into target DNA plasmids …or cellular genomes

- Recombination is a key natural aspect of DNA maintenance in cells as well as an artificial result of DNA cleavage and ligation in vitro
- Starts like restriction digestion with a nick or staggered break in the DNA
- Local homology promotes rejoining (to be discussed in detail later)
- But recombination can introduce new sequences or delete original sequences

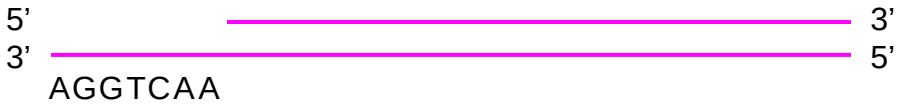
Sequence homology, even over short distance, can enhance DNA break repair



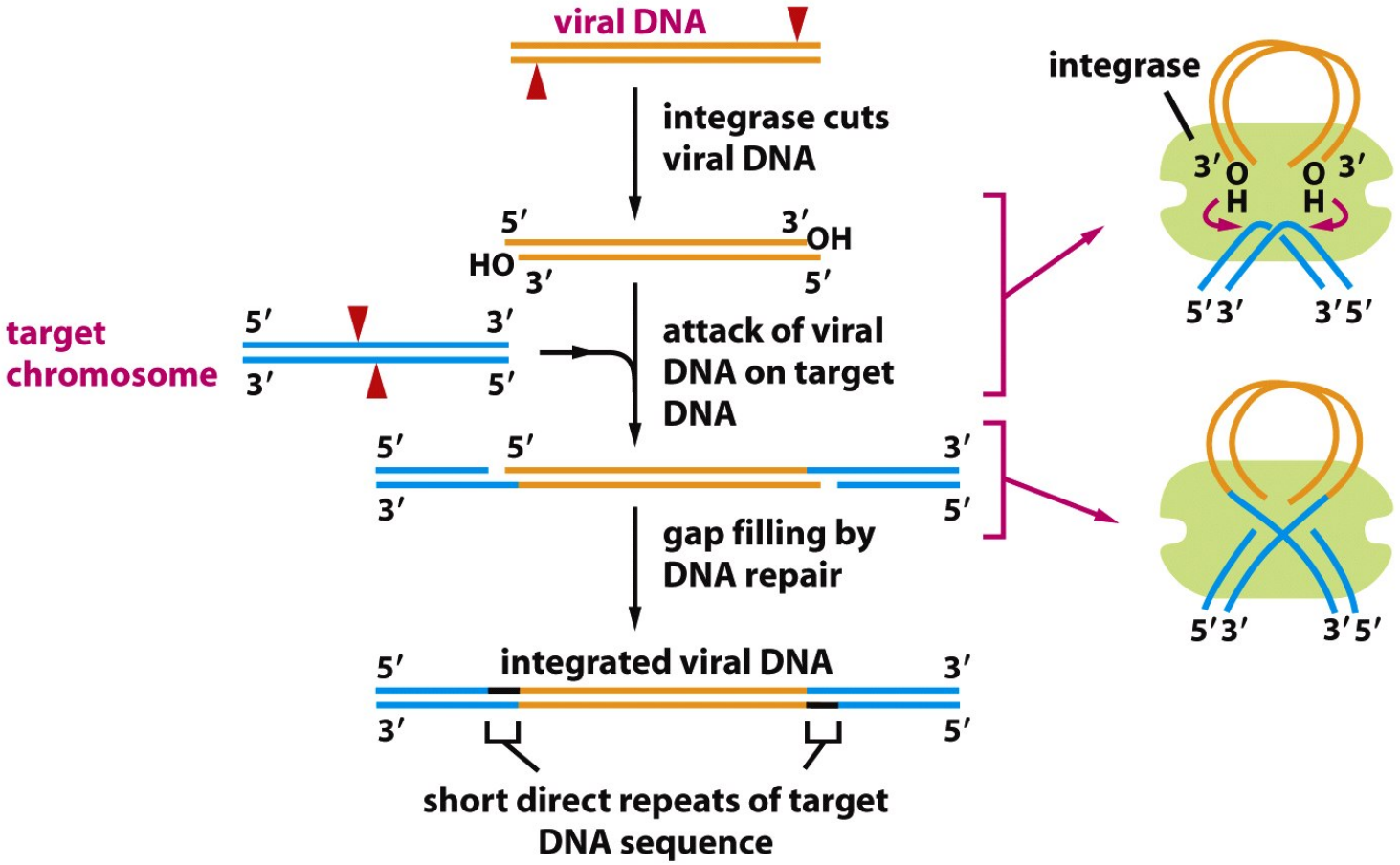
Poor partner for resealing

Imperfect … but adequate!

Poor partner for resealing



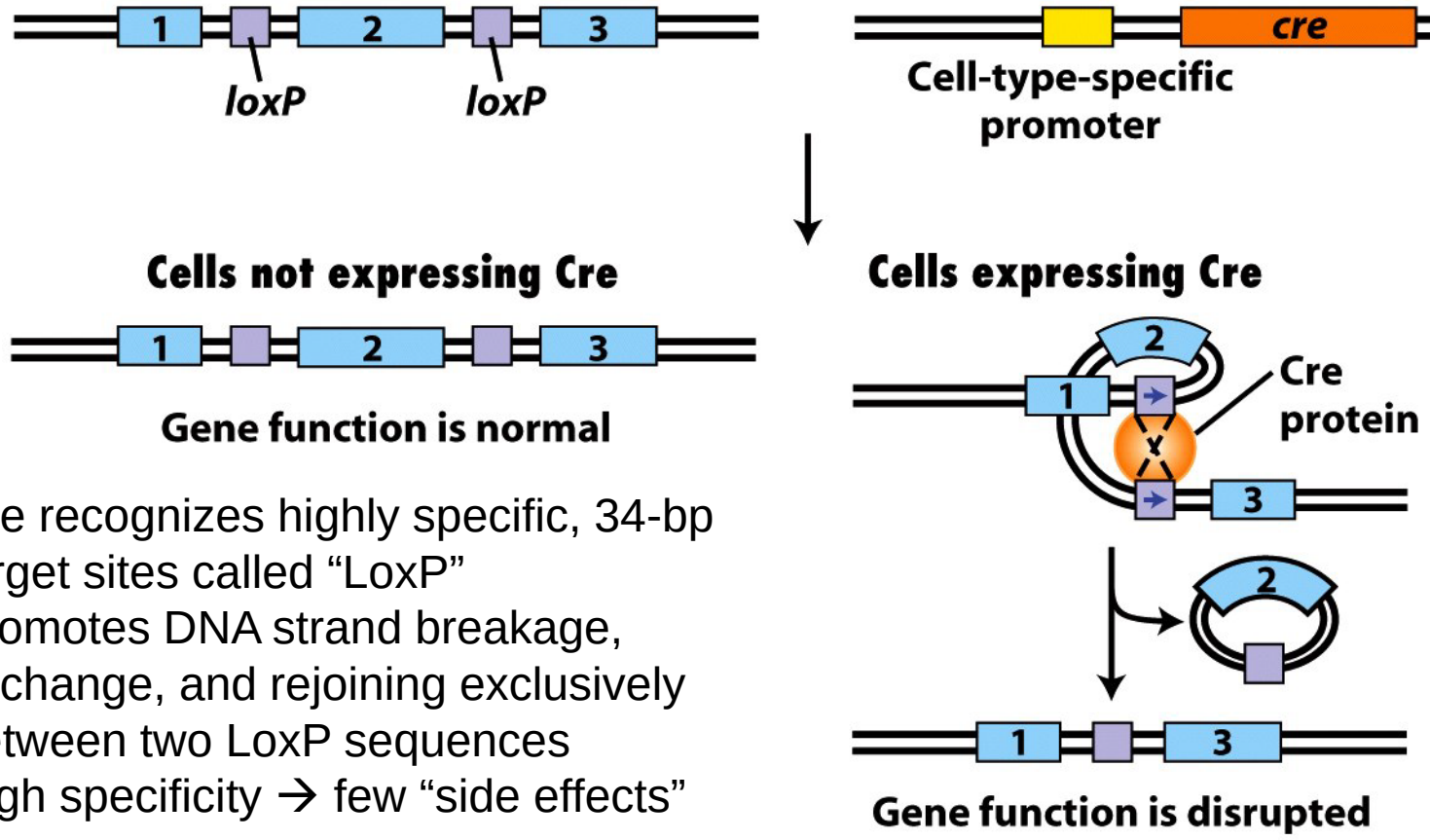
Viruses in prokaryotes and eukaryotes modify host genomes "for a living": they encode their own equivalents of restriction enzymes



Integrases: high specificity for non-disruptive sites in viral DNA, varying specificity for sites in host genome

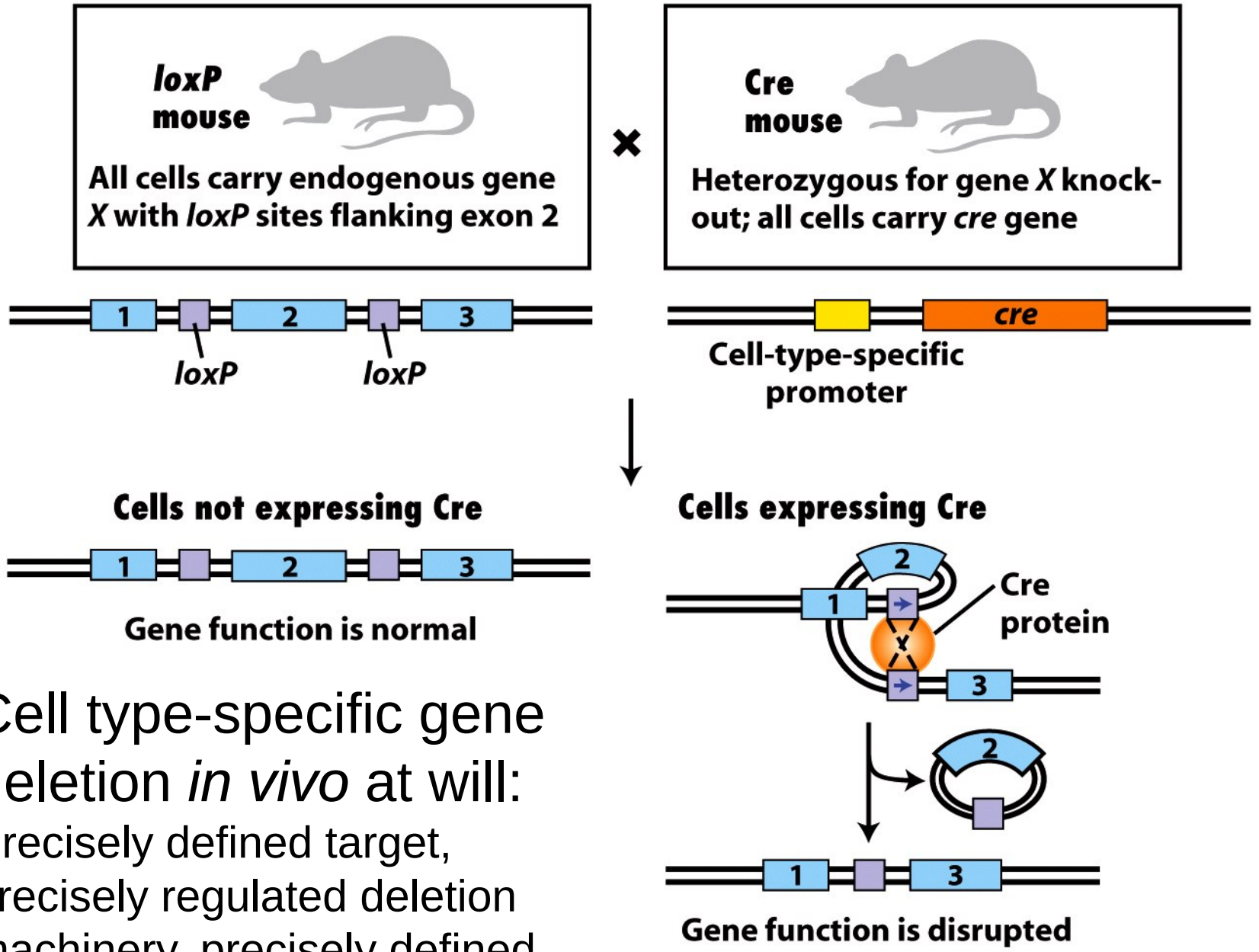
Figure 5-73 *Molecular Biology of the Cell* (© Garland Science 2008)

Cre: A particularly useful recombination enzyme – site-specific cutting and rejoining from a single enzyme



Cre recognizes highly specific, 34-bp target sites called "LoxP"
Promotes DNA strand breakage, exchange, and rejoining exclusively between two LoxP sequences
High specificity → few "side effects"

Figure 5-42
Molecular Cell Biology, Sixth Edition



Cell type-specific gene deletion *in vivo* at will:

Precisely defined target, precisely regulated deletion machinery, precisely defined structure for end product