BI 8 LECTURE 5

MORE ON HOW WE KNOW WHAT WE KNOW… AND INTRO TO THE PROTEIN CODE

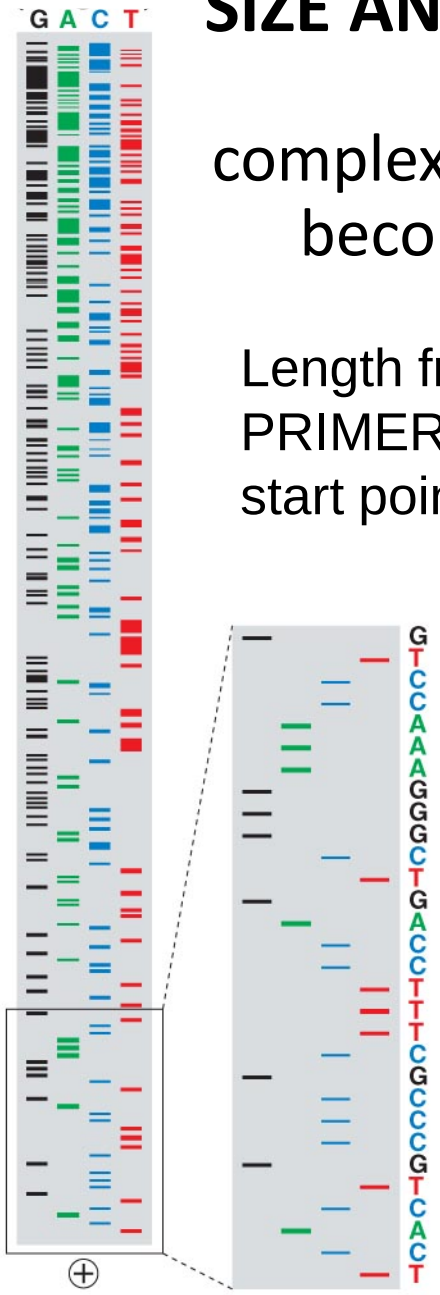
Ellen Rothenberg

19 January 2016

SIZE AND PURIFICATION BY SYNTHESIS: BASIS OF EARLY SEQUENCING

complex mixture of aborted DNA replication products becomes an orderly “ladder” of DNA sequence

Length from specific PRIMER site defines start point by ***synthesis***



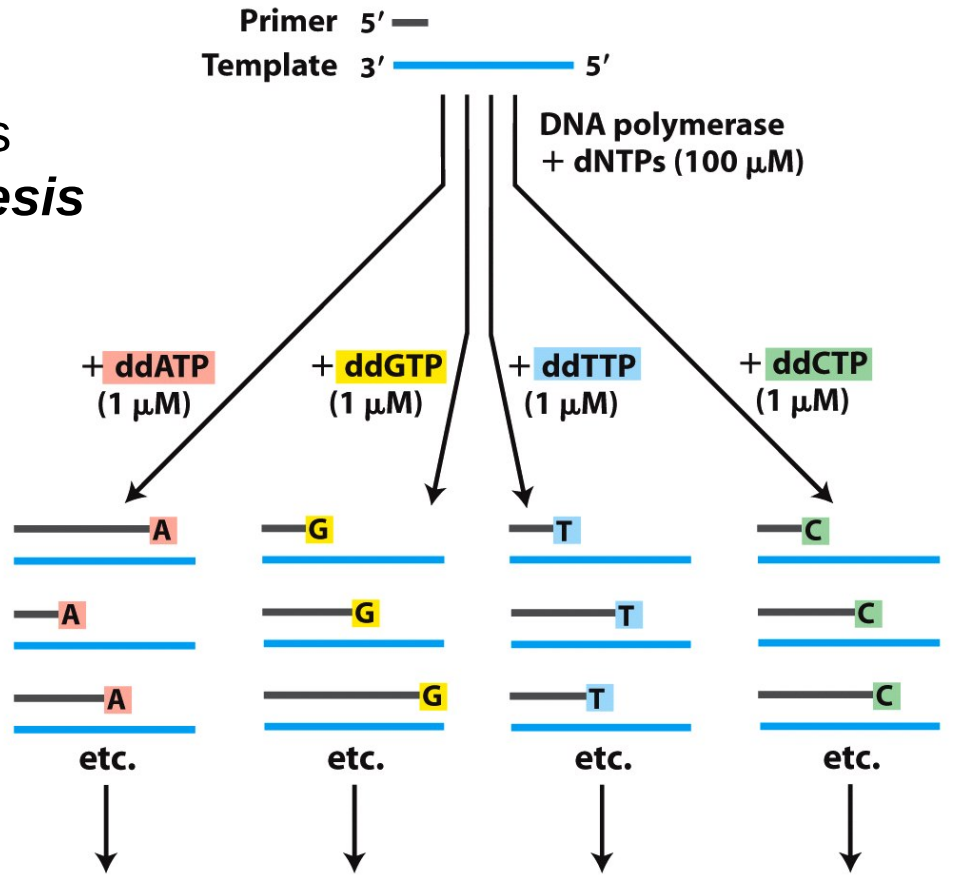


Figure 5-21b
Molecular Cell Biology, Sixth Edition

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

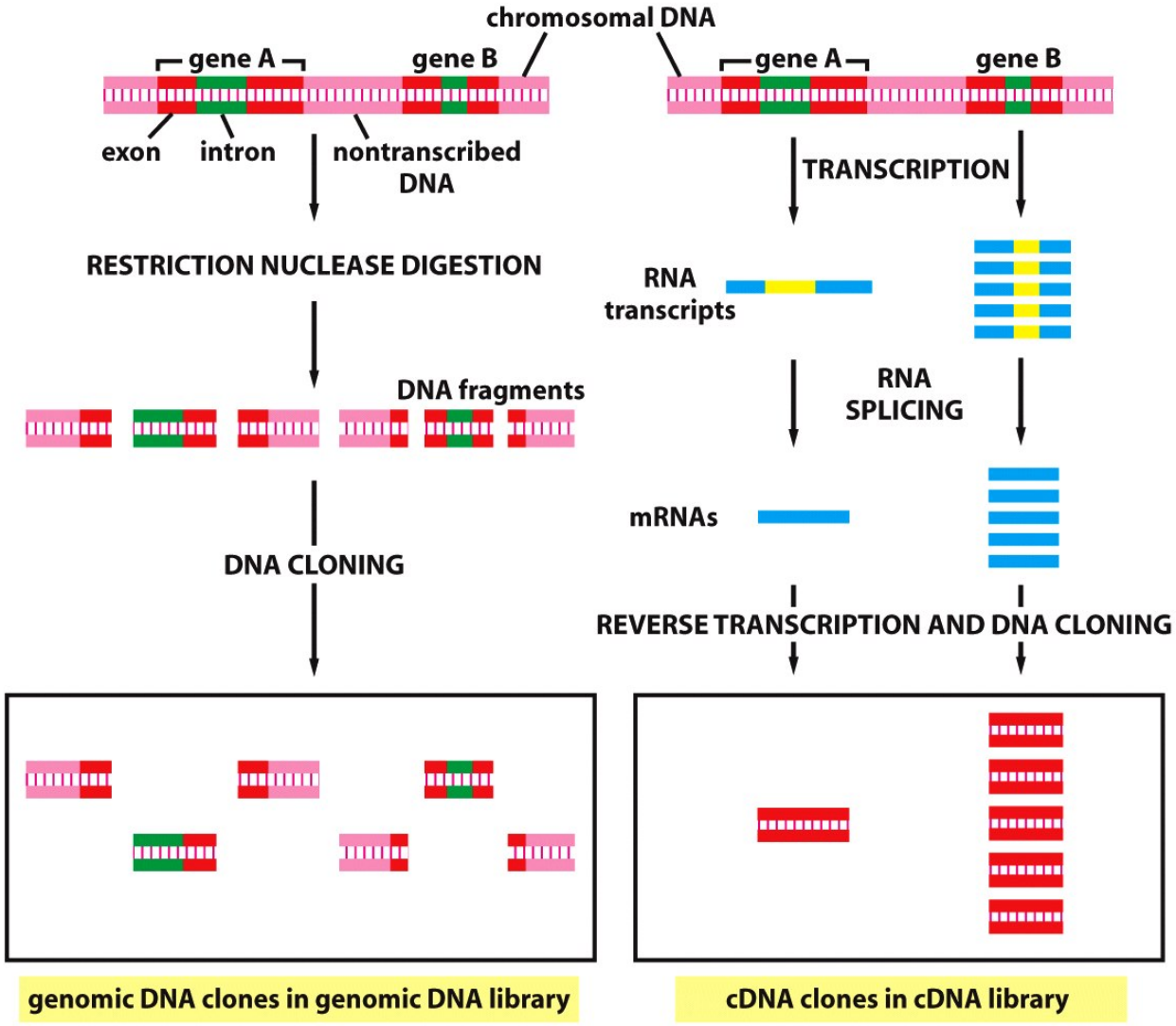
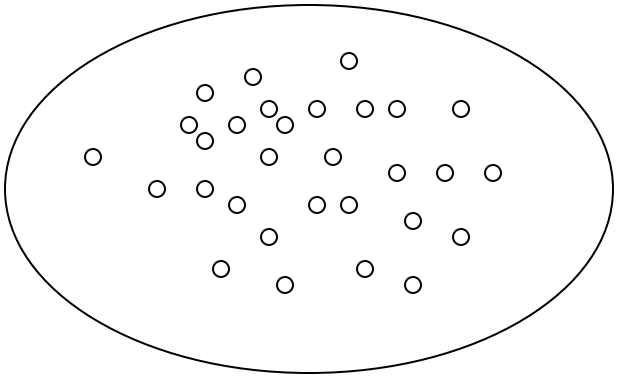


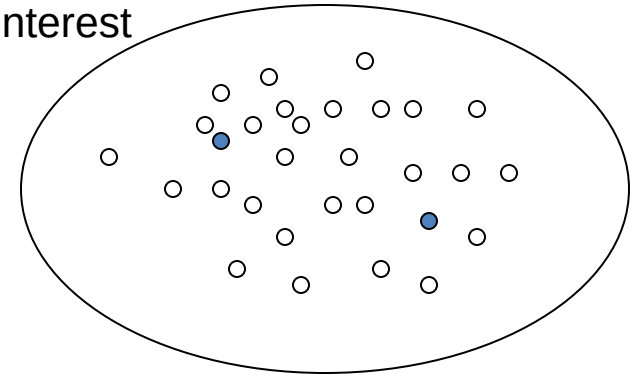
Figure 8-44 *Molecular Biology of the Cell* (© Garland Science 2008)

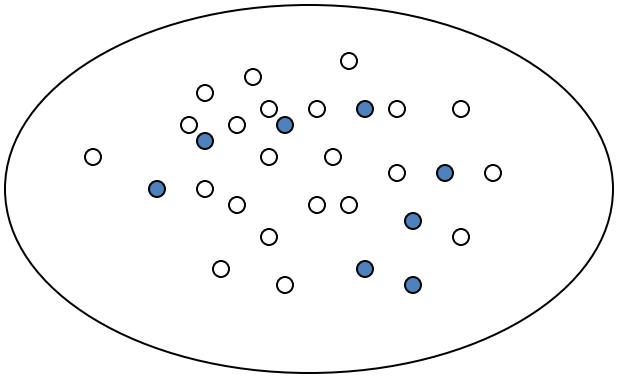
Cloning and screening a genomic or cDNA library with specific probes is the basis of specific gene discovery

Bacterial library colony DNA screened with negative control probe

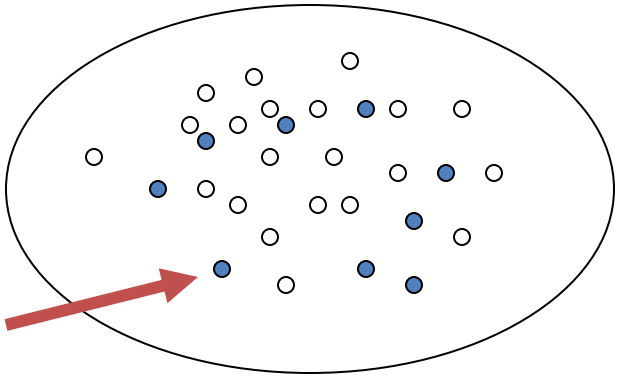


Bacterial library colony DNA screened with probe for a specific gene of interest





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***)

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.

- Can test by hybridization either to RNA itself or to cDNA copied from the RNA

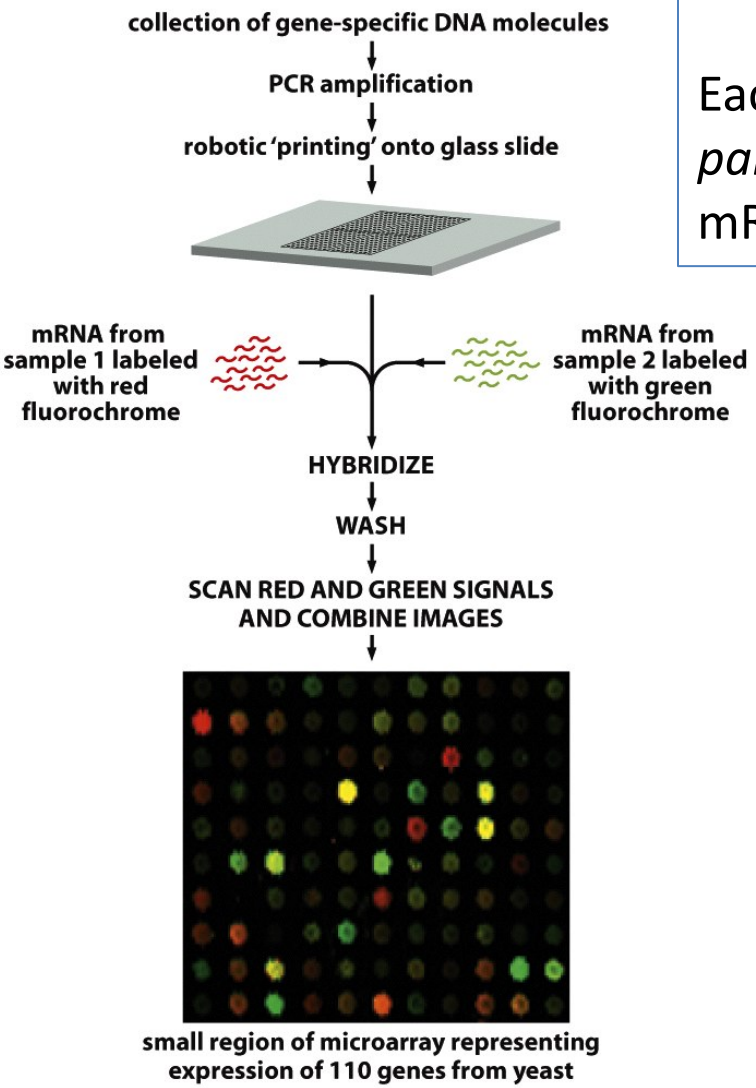
- Can probe bulk RNA samples, RNA samples separated by size, or cDNA clones separated as separate bacterial colonies

- Analyze structure and frequency of particular RNA molecules in population

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

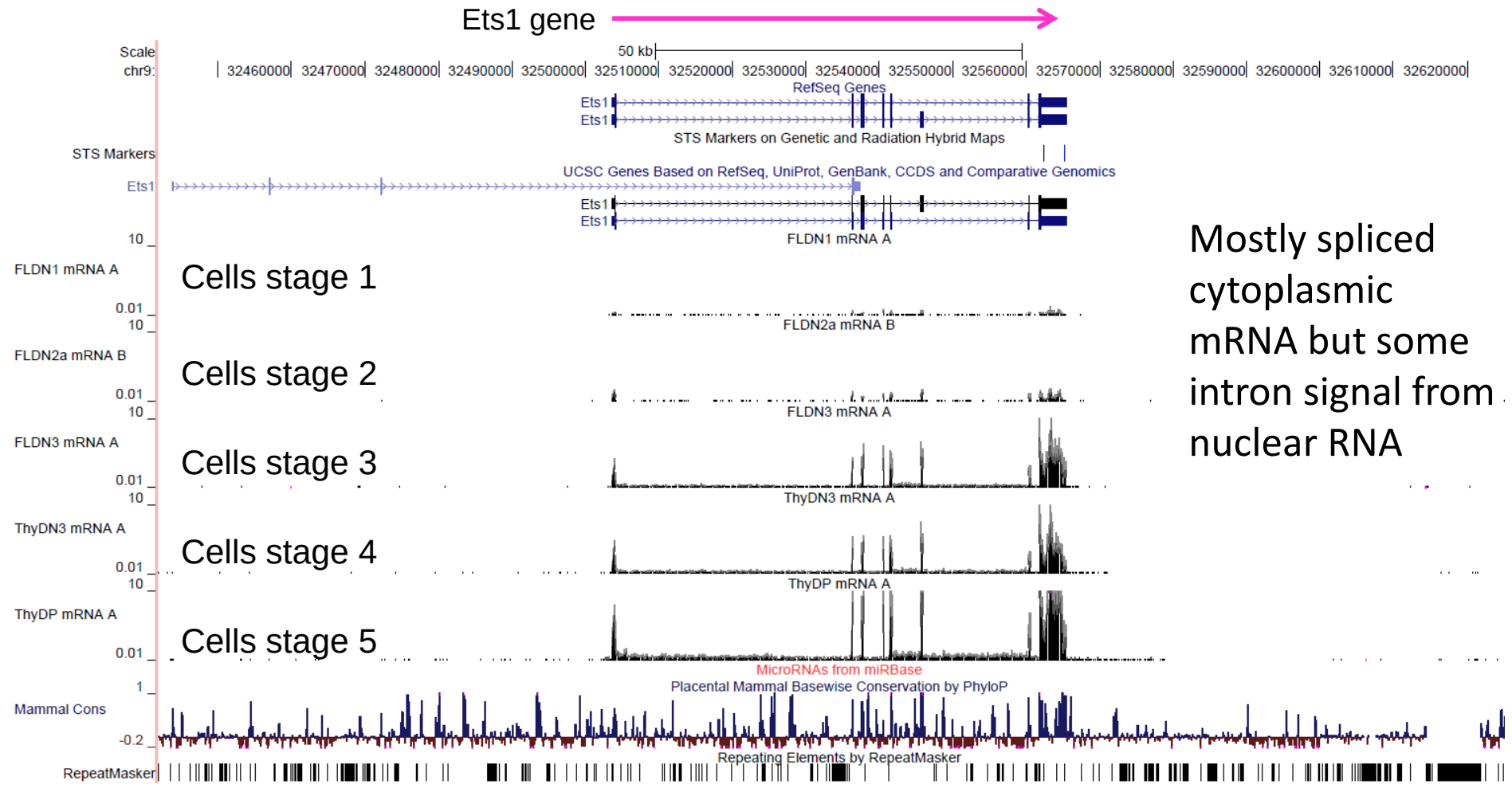


Short probes: Oligonucleotides

Each detect one *part* of target mRNA

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

Identification of expressed RNAs by bulk cDNA sequencing & alignment to genome: RNA-seq



Mostly spliced cytoplasmic mRNA but some intron signal from nuclear RNA

- Easy to see changes in level and which exons are expressed overall

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

Structures come from cDNA clones

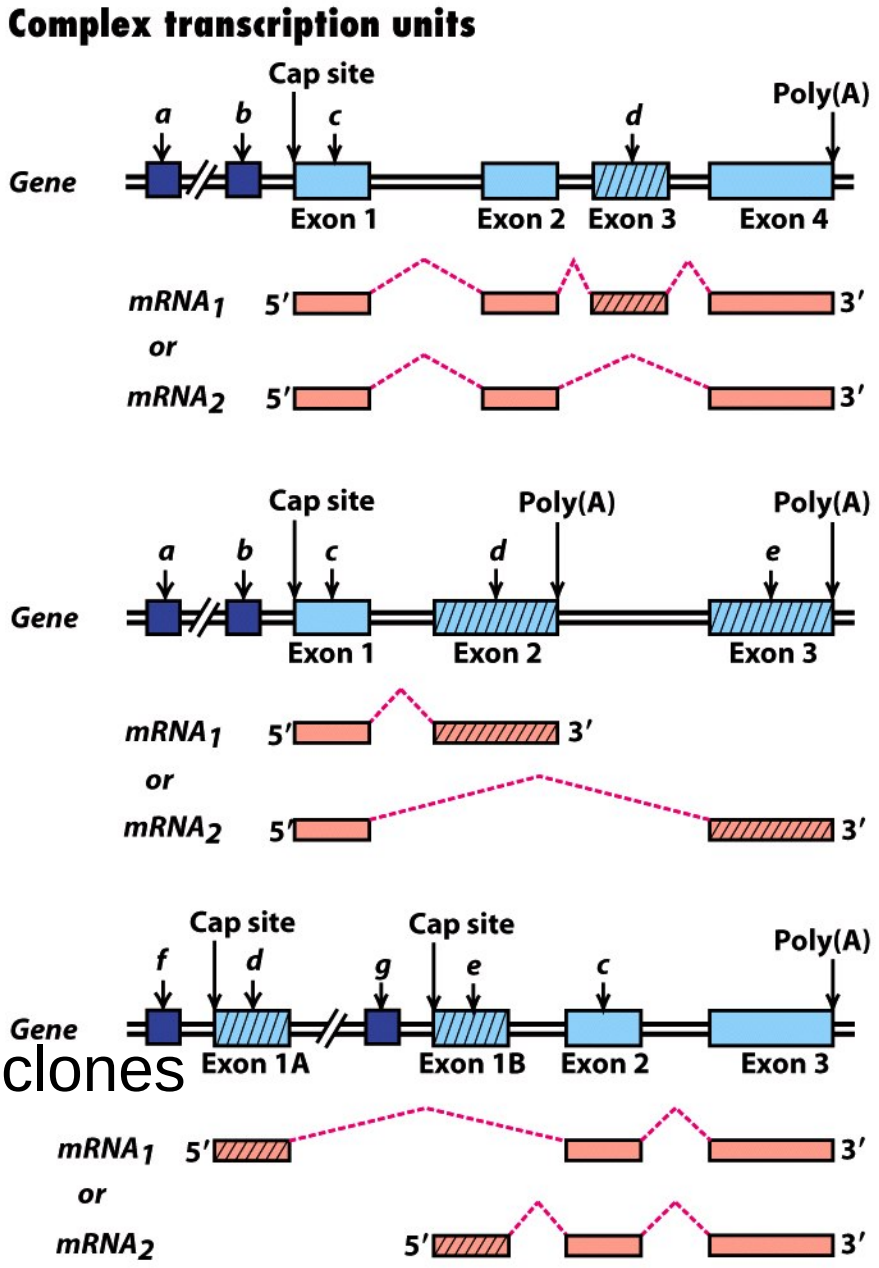


Figure 6-3b
Molecular Cell Biology, Sixth Edition

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…)

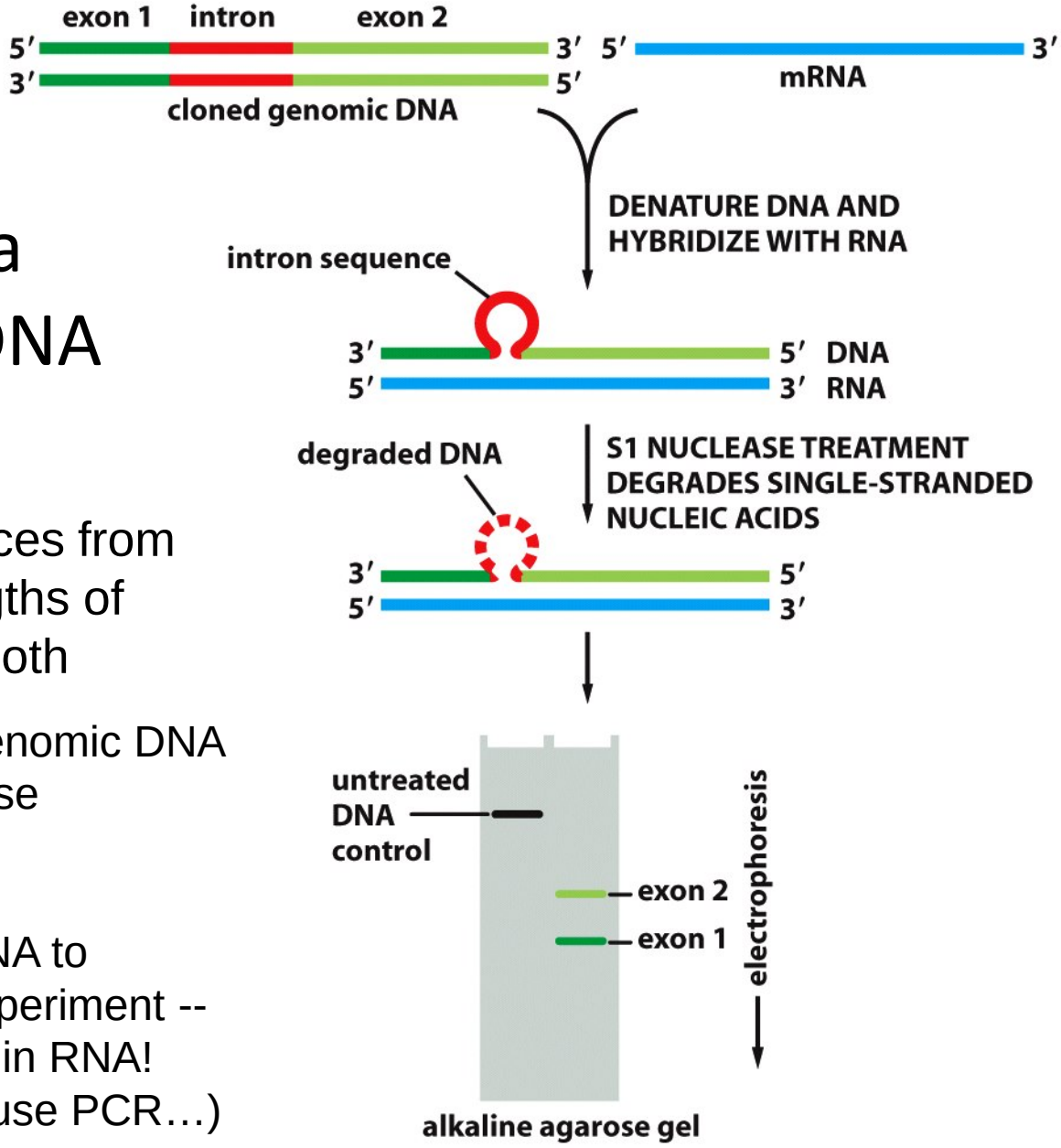
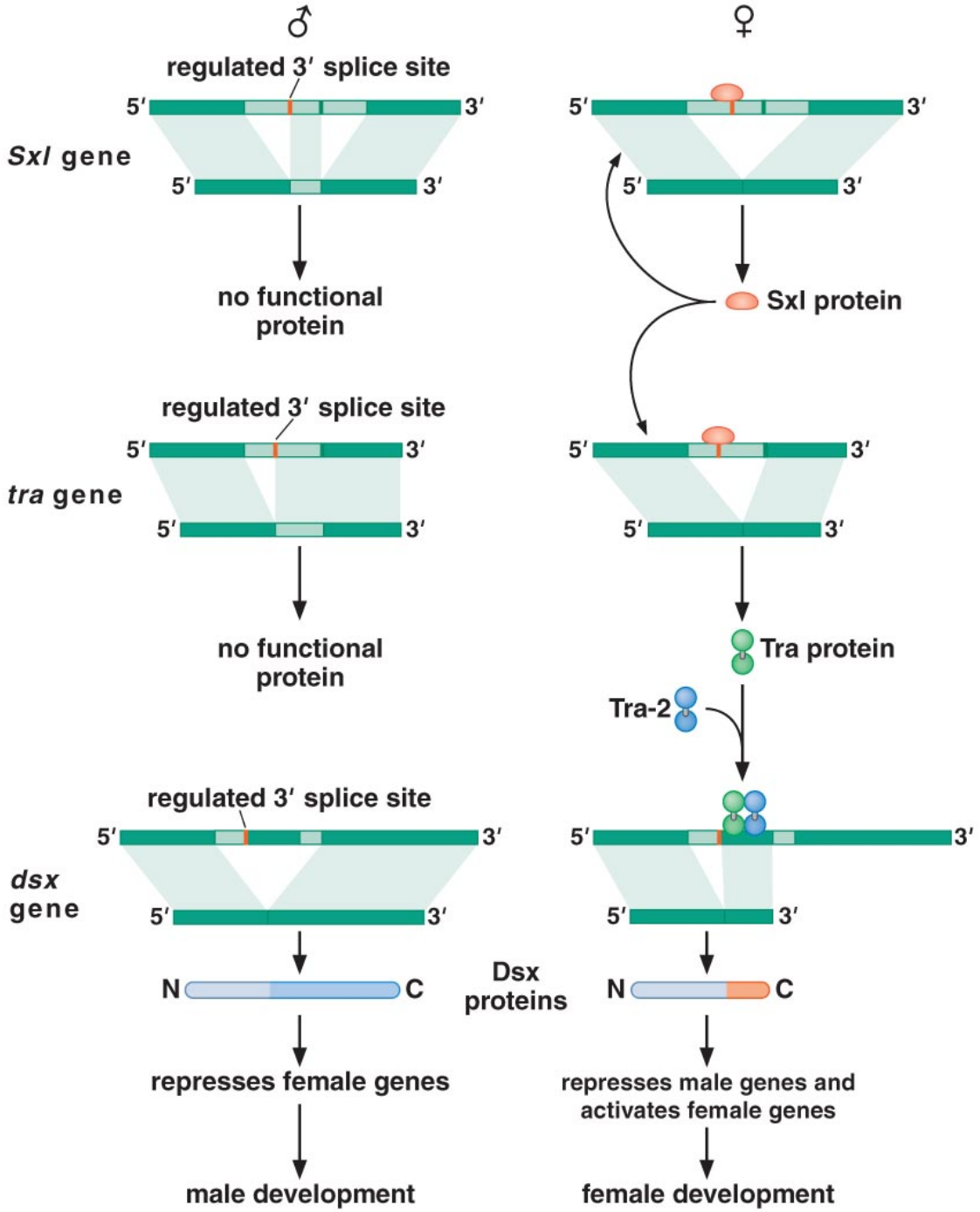


Figure 8-37 *Molecular Biology of the Cell* (© Garland Science 2008)

Splicing matters! Famous case: Drosophila sex determination

Splicing of internal exons can be regulated by SA site accessibility: binding of protein to RNA at SA site can control intron inclusion or exclusion… by blocking RNA to RNA attack



A powerful combination: regulating splicing by regulating choice of polyadenylation site:

No splice acceptor → no clipping at splice donor site either

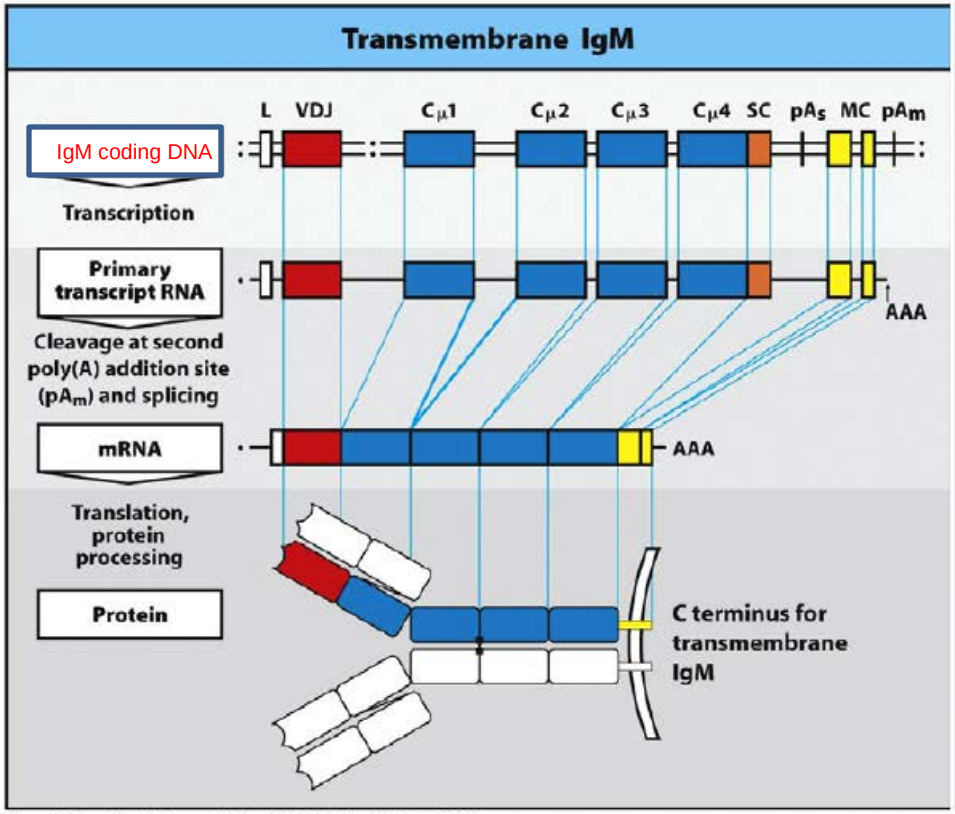


Figure 4-19 part 1 of 2 Immunobiology, 7ed. (© Garland Science 2008)

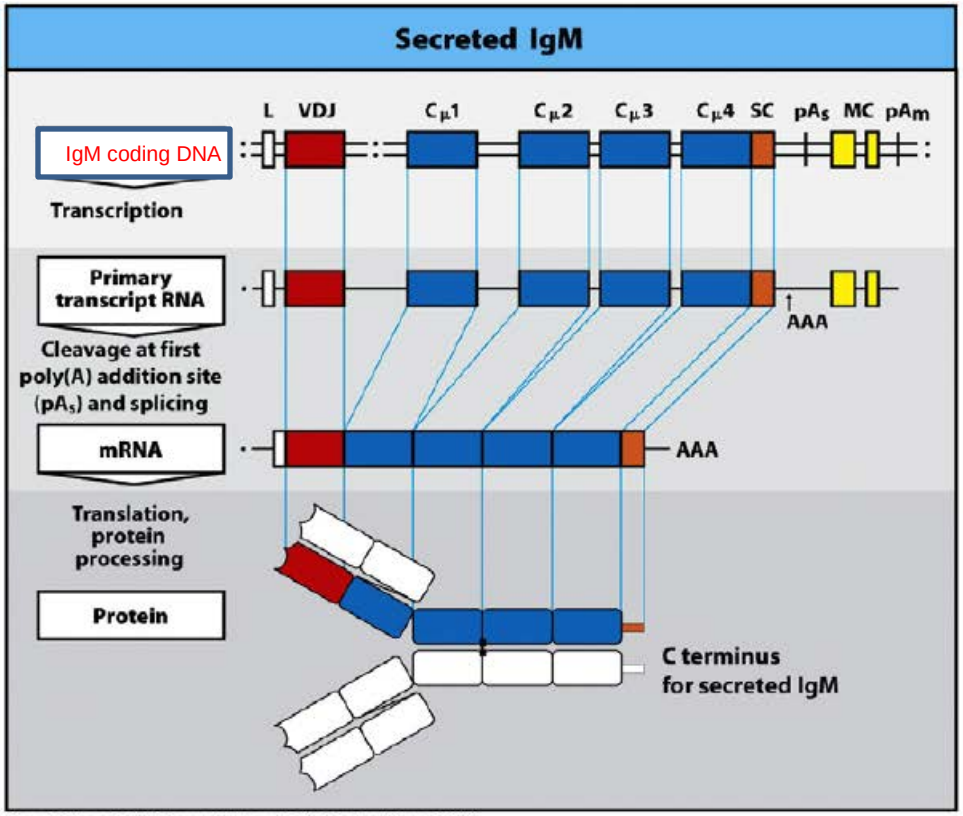


Figure 4-19 part 2 of 2 Immunobiology, 7ed. (© Garland Science 2008)

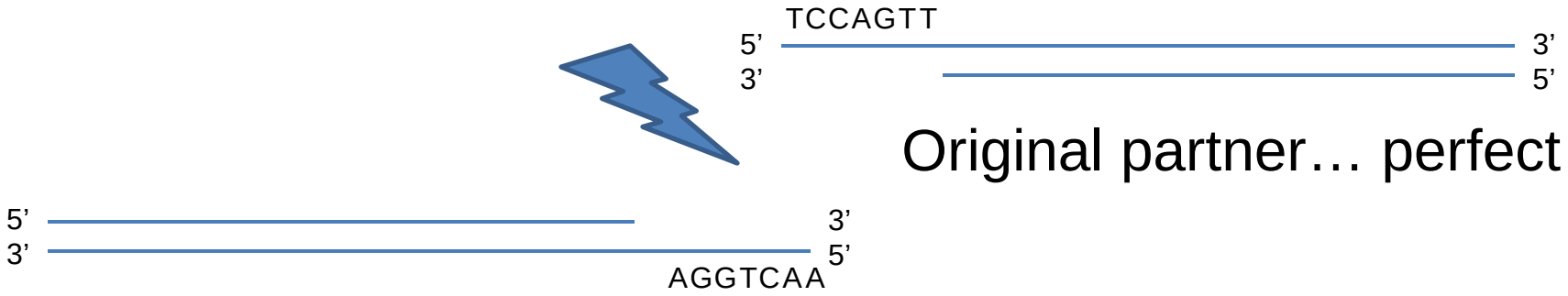
Same gene can code for either a membrane-bound or a secreted protein depending on ***choice of poly(A) site***… either before or at the end of last possible coding exon

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

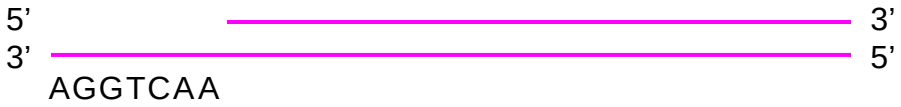
Sequence homology, even over short distance, can enhance DNA break repair



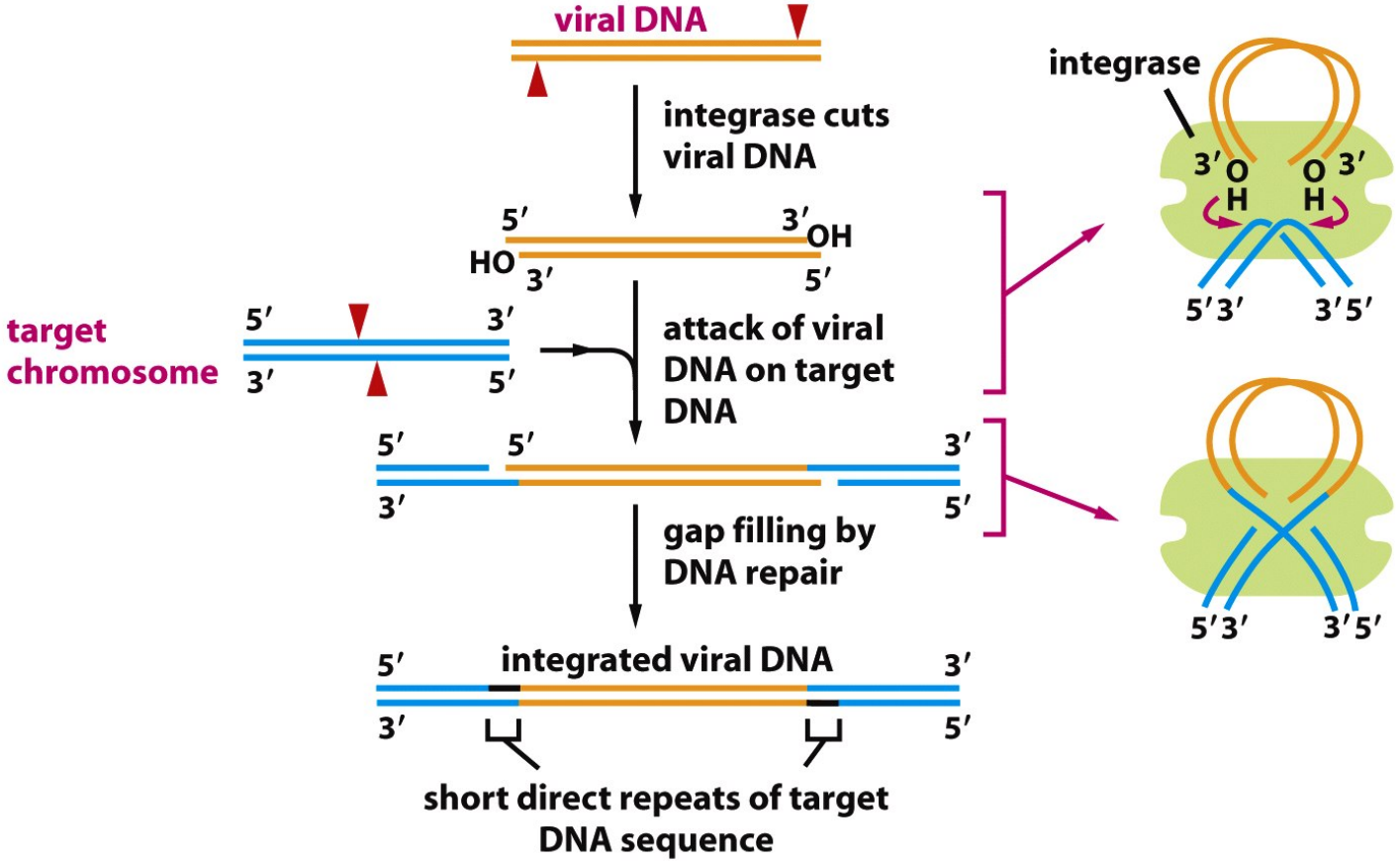
Poor partner for resealing

Imperfect … but adequate!

Poor partner for resealing



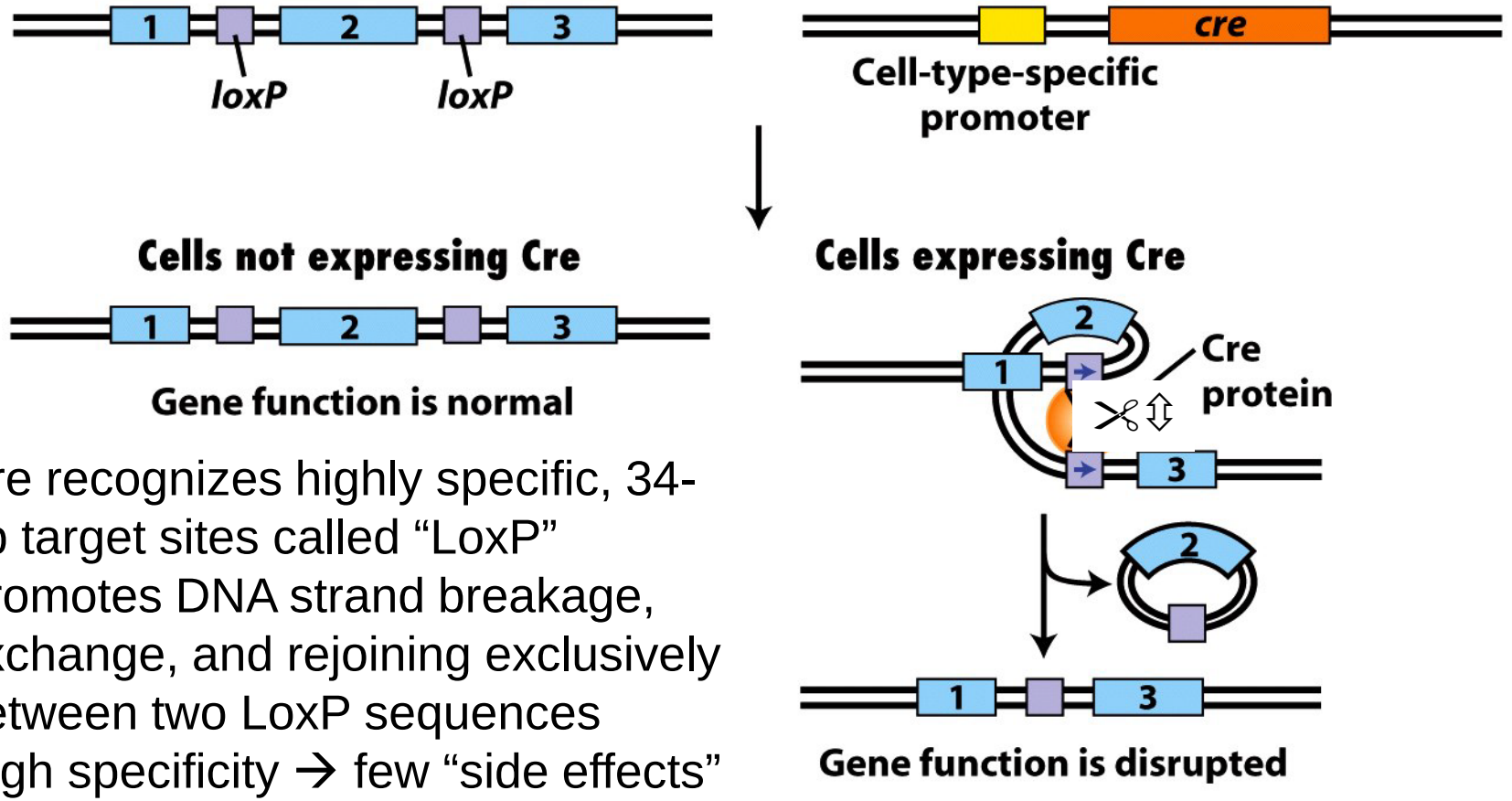
Viruses in prokaryotes and eukaryotes modify host genomes "for a living": they encode their own equivalents of restriction enzymes



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

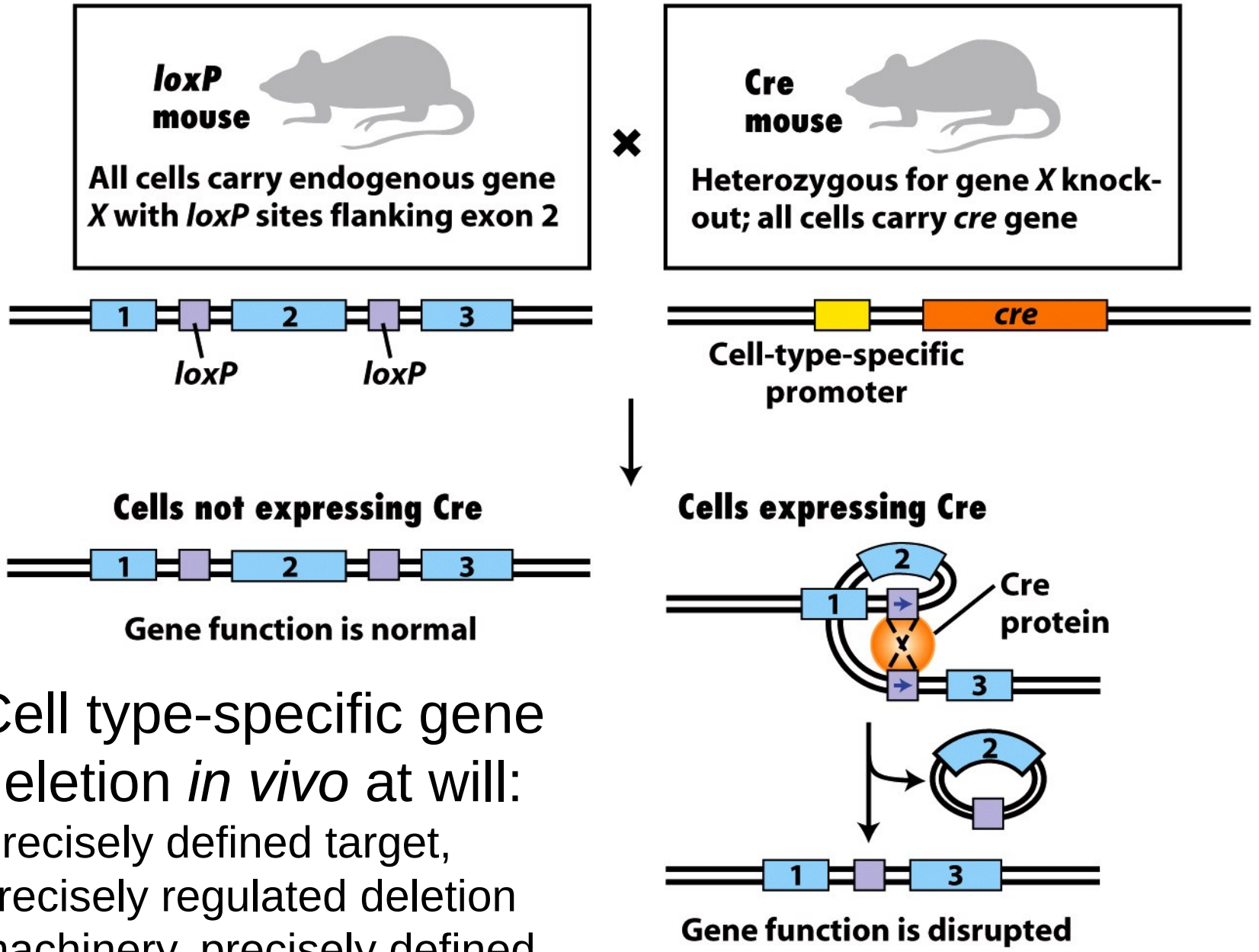
Figure 5-73 *Molecular Biology of the Cell* (© Garland Science 2008)

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



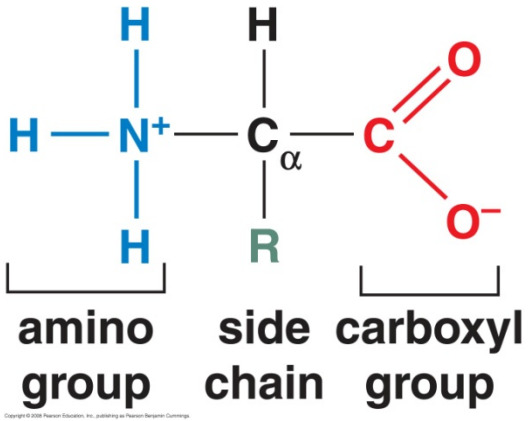
Cell type-specific gene deletion *in vivo* at will:

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

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

Translating from one code to another
Protein basics: subunits are amino acids



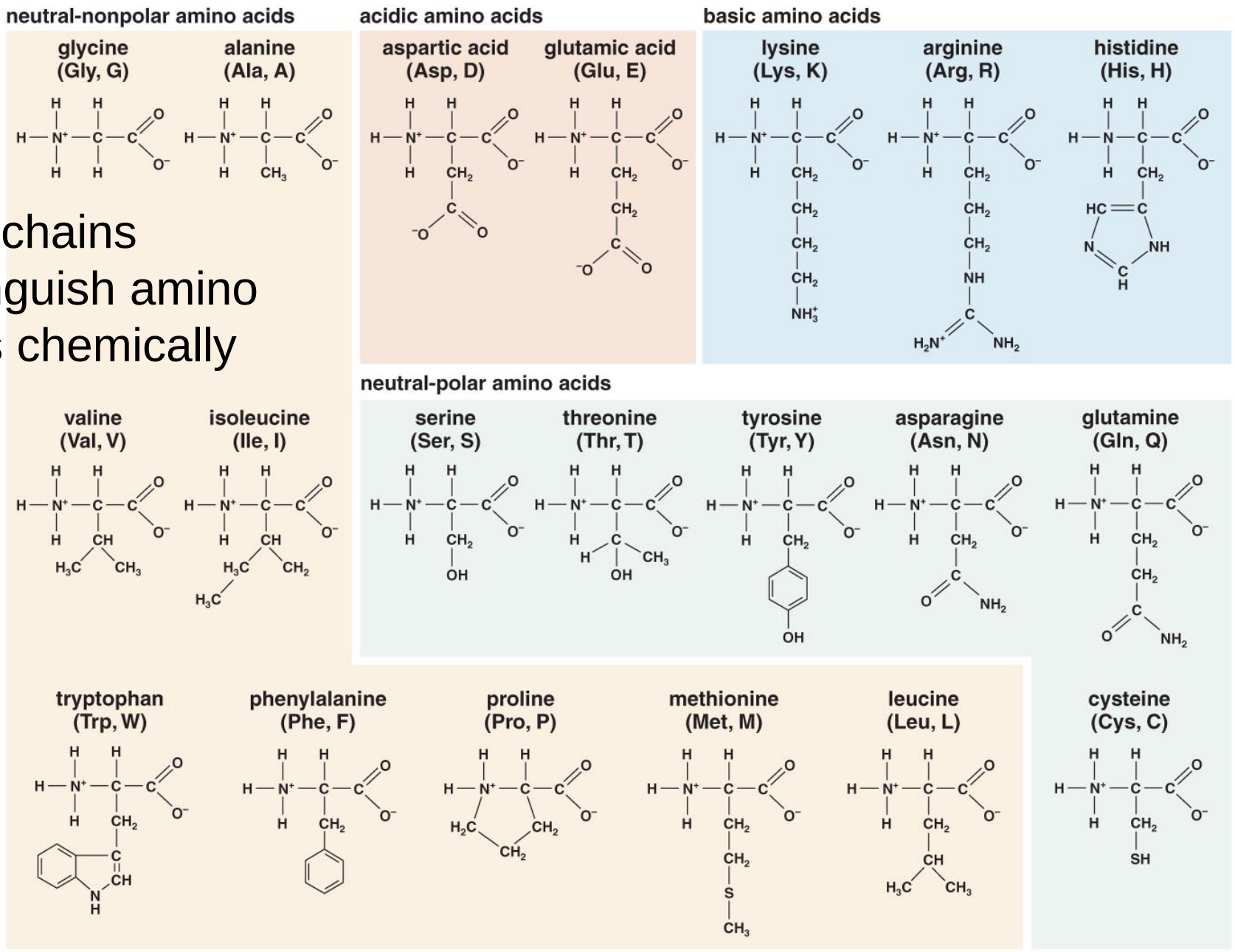
Unlike –NH2 in nucleotides, these *free* amino groups and free carboxyl groups are ionized at neutral pH

Coding is a qualitative transformation, nucleotide triplets translated to amino acids

No structural homology

Only polarity of polymerization is similar: what happens to the end of the chain depends on what comes before it

Side chains distinguish amino acids chemically



Formation of amide bonds between carboxyl groups of one amino acid and amino group of next amino acid polymerizes protein chain… "N" to "C"

Progressive, vectorial polymerization

Read from RNA template, 5' to 3'

N terminal amino acid is first, keeps free NH_3^+

Other residues added sequentially to its carboxy end

Last amino acid added: "C terminal"

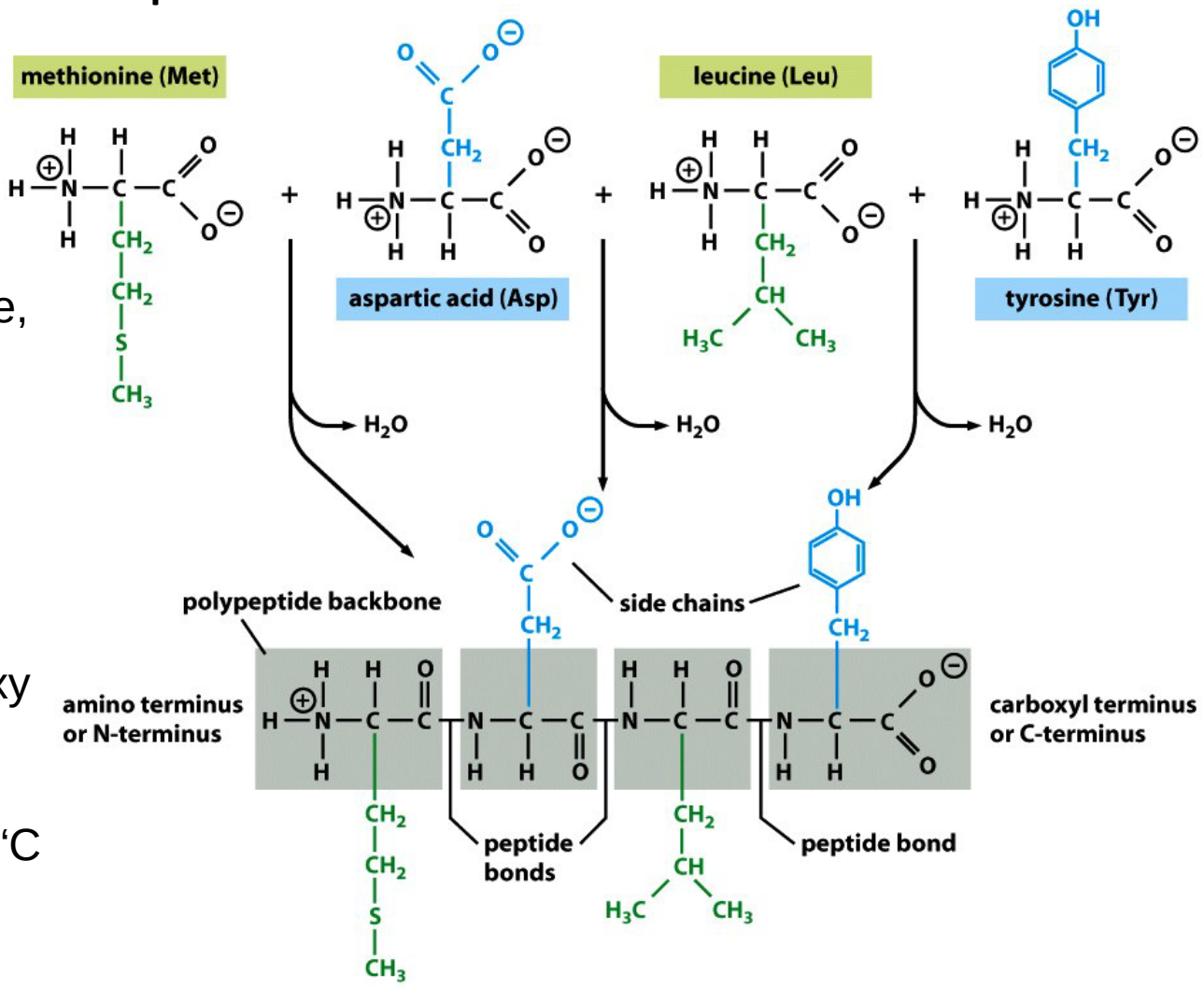


Figure 3-1 *Molecular Biology of the Cell* (© Garland Science 2008)

THREE MATTERS… in protein coding regions

You must know where to start counting and stay in register

4^3 = 64 triplets
20 amino acids + stop signals

Unique codon for starting protein chains; also used for internal Met

GAUAU=?

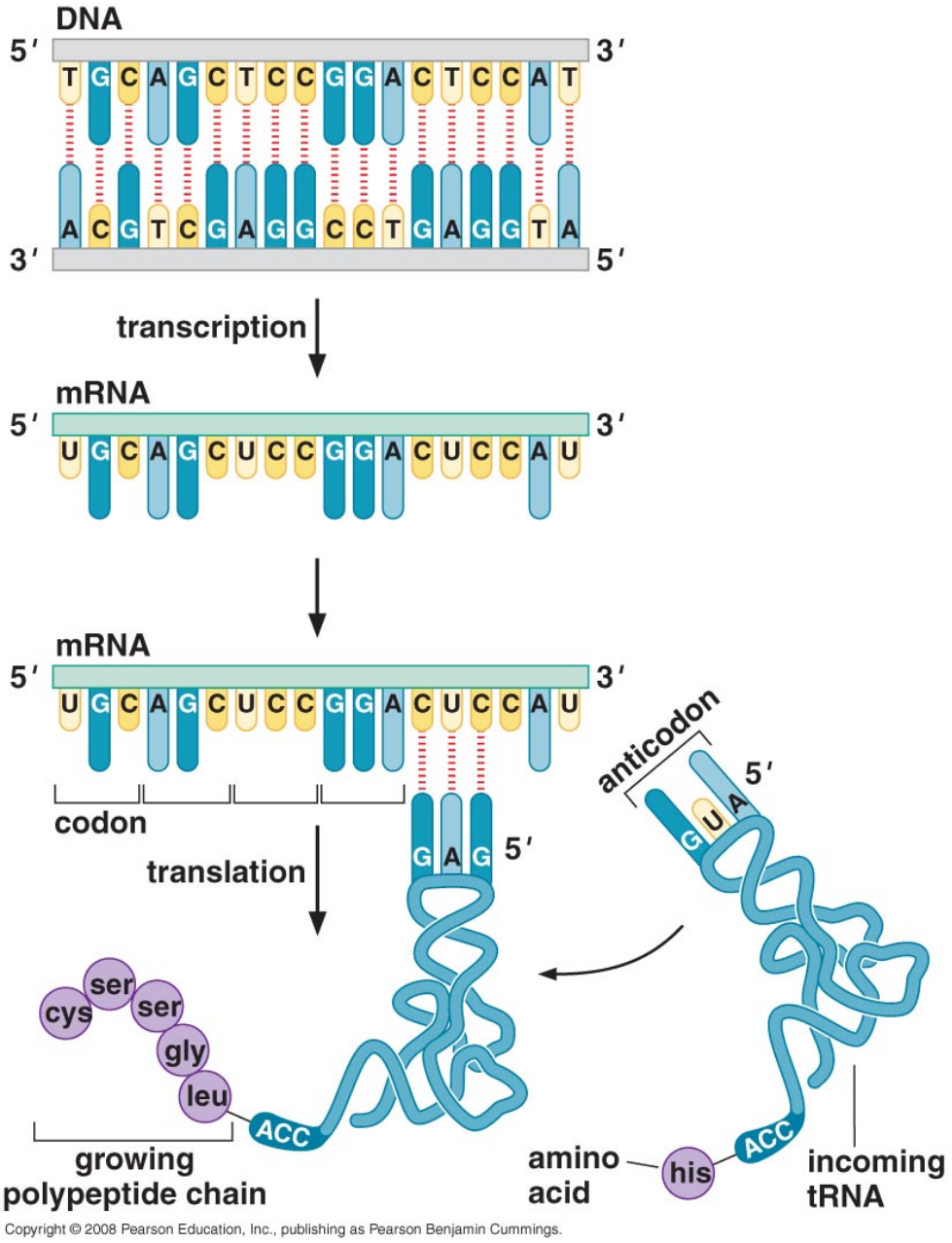
TABLE 15-1 The Genetic Code

first position (5′ end)	second position: U	second position: C	second position: A	second position: G	third position (3′ end)
U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	U
	UUC Phe	UCC Ser	UAC Tyr	UGC Cys	C
	UUA Leu	UCA Ser	UAA* stop	UGA* stop	A
	UUG Leu	UCG Ser	UAG* stop	UGG Trp	G
C	CUU Leu	CCU Pro	CAU His	CGU Arg	U
	CUC Leu	CCC Pro	CAC His	CGC Arg	C
	CUA Leu	CCA Pro	CAA Gln	CGA Arg	A
	CUG Leu	CCG Pro	CAG Gln	CGG Arg	G
A	AUU Ile	ACU Thr	AAU Asn	AGU Ser	U
	AUC Ile	ACC Thr	AAC Asn	AGC Ser	C
	AUA Ile	ACA Thr	AAA Lys	AGA Arg	A
	AUG† Met	ACG Thr	AAG Lys	AGG Arg	G
G	GUU Val	GCU Ala	GAU Asp	GGU Gly	U
	GUC Val	GCC Ala	GAC Asp	GGC Gly	C
	GUA Val	GCA Ala	GAA Glu	GGA Gly	A
	GUG Val	GCG Ala	GAG Glu	GGG Gly	G

* Chain-terminating or "nonsense" codons.

† Also used in bacteria to specify the initiator formyl-Met-tRNA^{fMet}.

Converting mRNA sequence to protein sequence requires triplet decoding adaptor molecules: tRNA



… to be continued. Next session's reading:
Alberts Ch. 6: pp. 333-366.
Now,

QUIZ 1

- 20 min, closed book