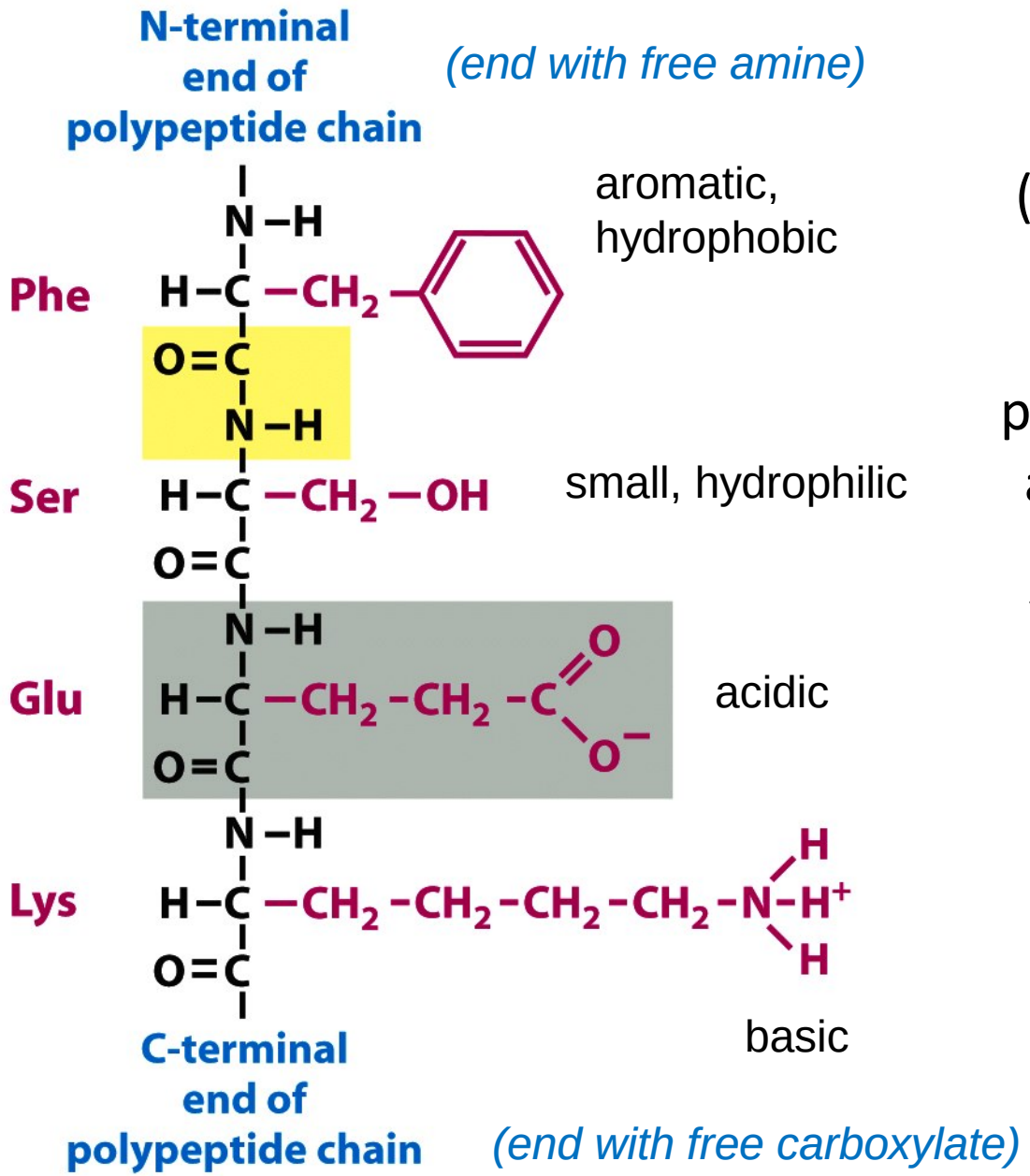
BI 8 Lecture 7
Protein structure, functional analysis, and evolution

Ellen Rothenberg

26 January 2016

Reading: Ch. 3, pp. 109-134; panel 3-1



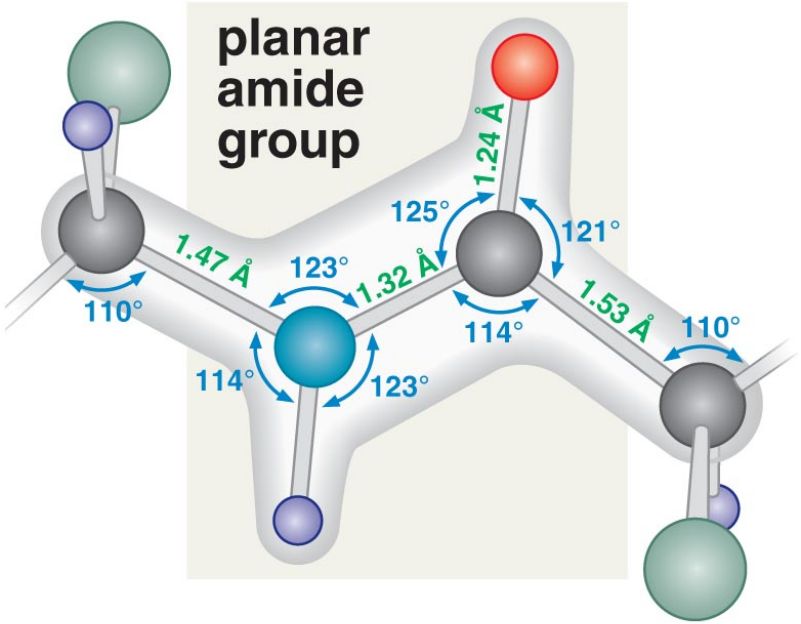
Anatomy of a protein (polypeptide): N terminal, C terminal, and protruding amino acid side chains

Very distinct chemistries of the side chains – many options for folding and intrachain interactions!

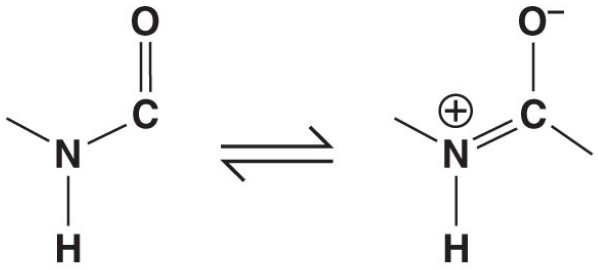
Figure 2-24 *Molecular Biology of the Cell* (© Garland Science 2008)

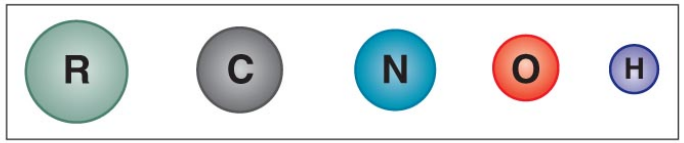
The amide bond itself between each pair of amino acids is planar, stabilized by resonance

“next” amino acid
(toward C terminus)



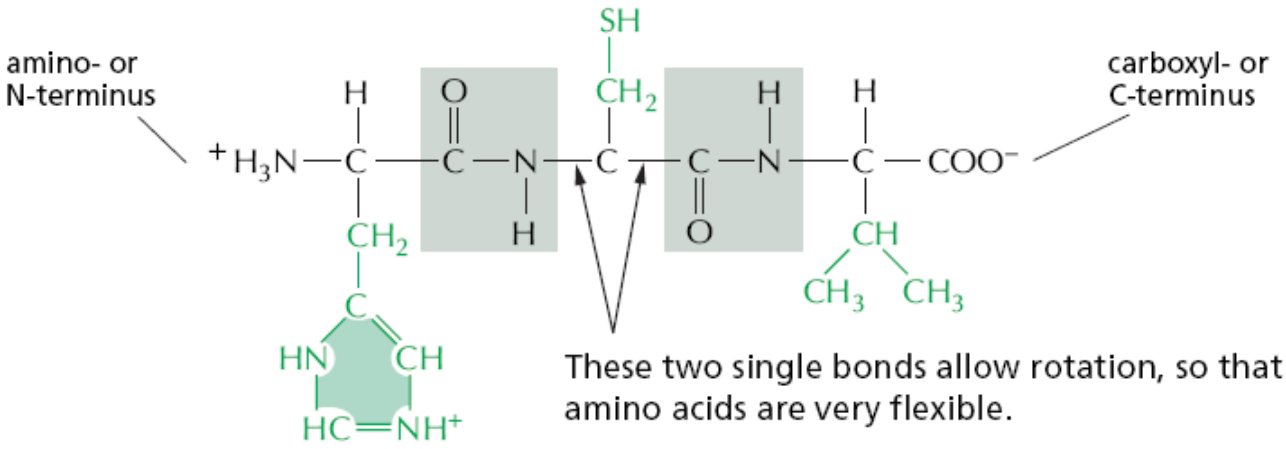
“previous” amino acid
(toward N terminus)

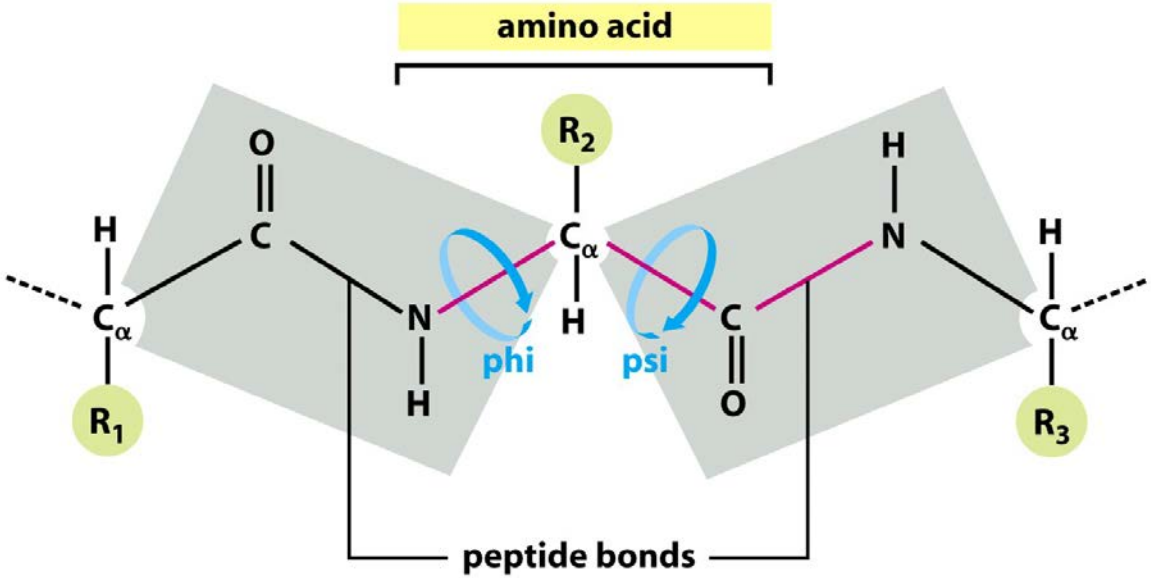




(Pictures like these are by Irving Geis, famous molecular illustrator)

Protein structure – 3D placement of aa1 relative to aa3 – depends on rotation angles at Cα-NH and Cα-C=O bonds, *within* aa2

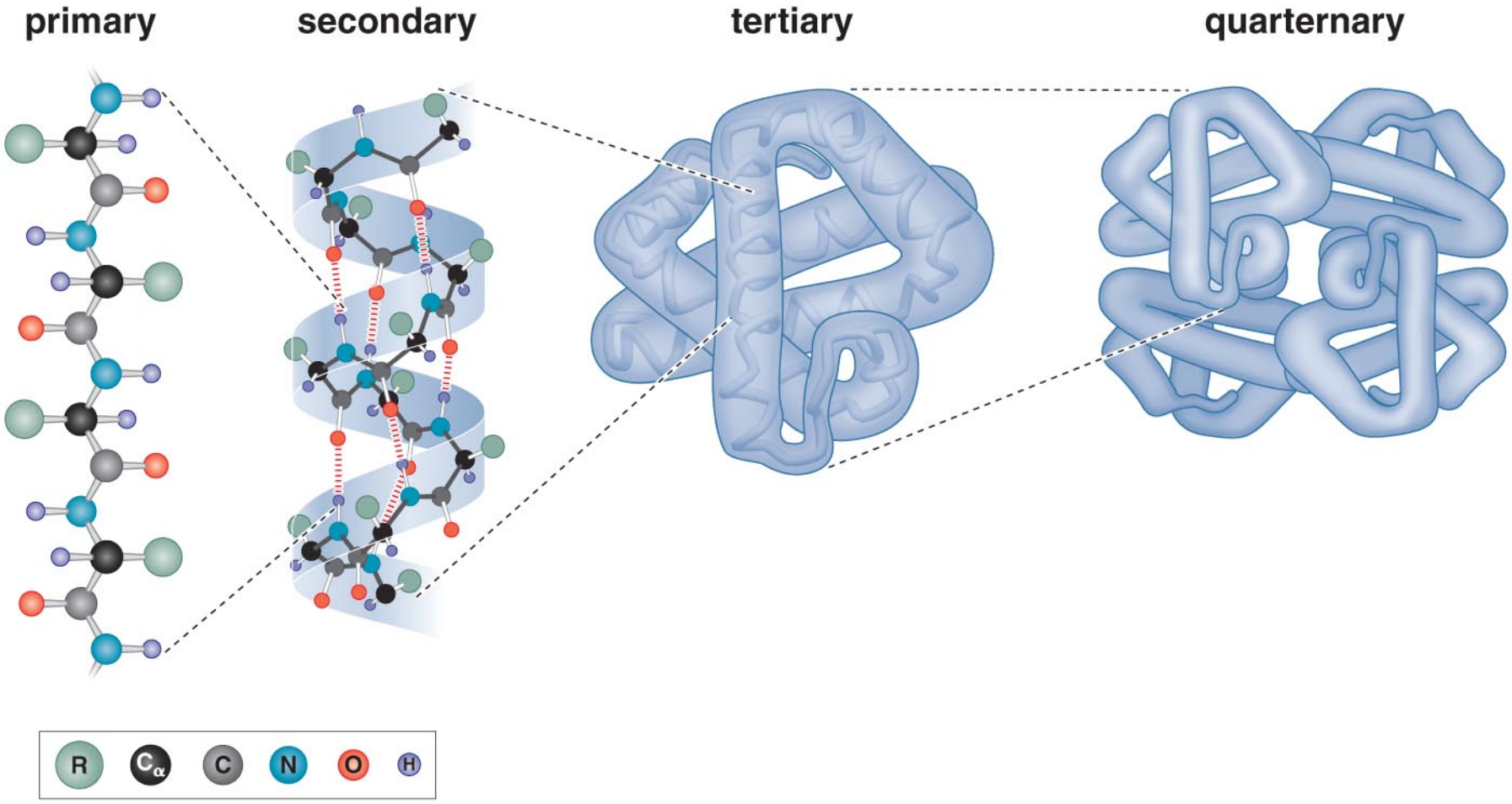


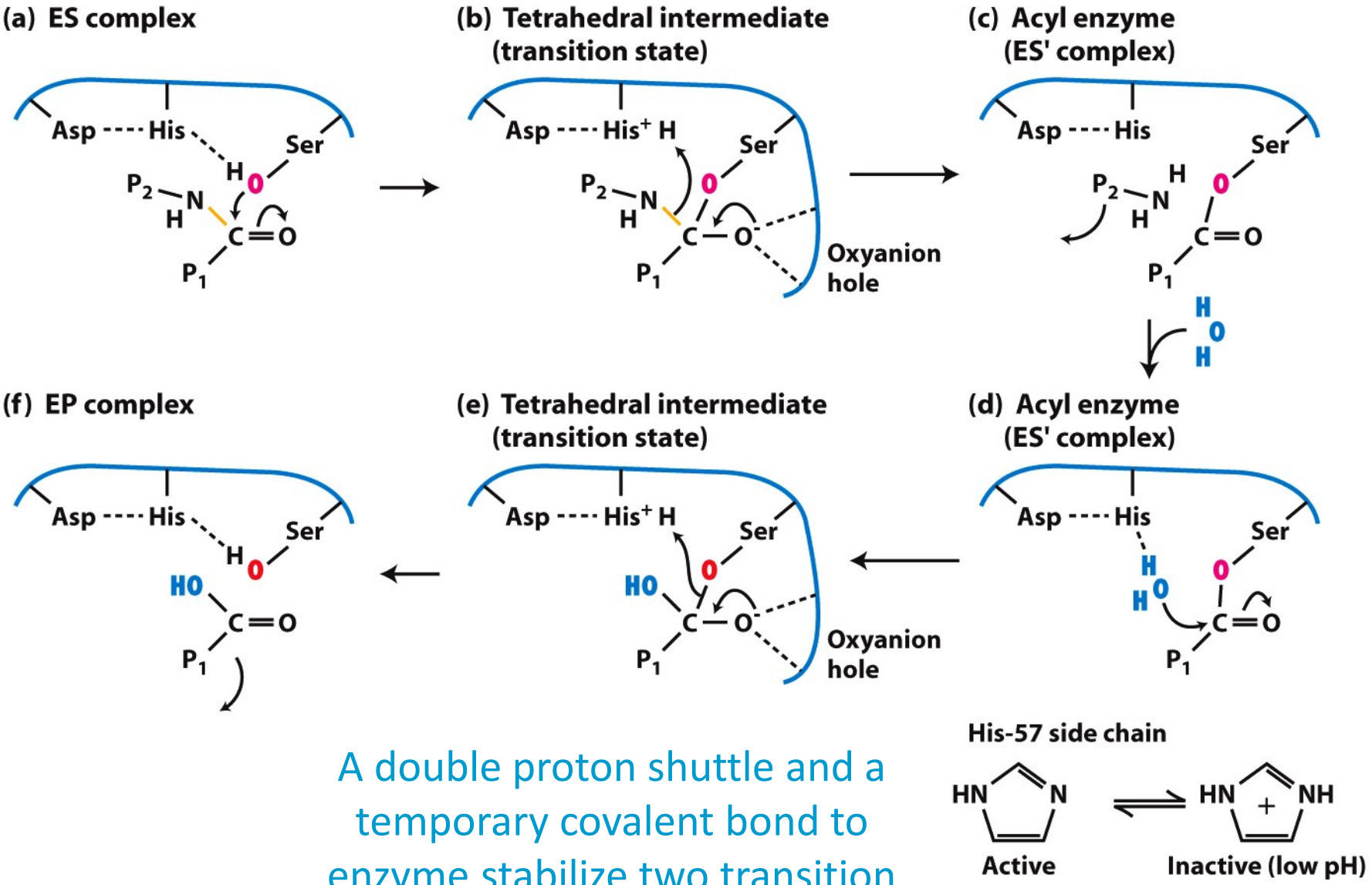


Bonds can rotate to position side chains with specific chemistries along specific faces of a protein…
…thus creating distinct biochemical domains

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

Protein structure at four levels of organization: building domains





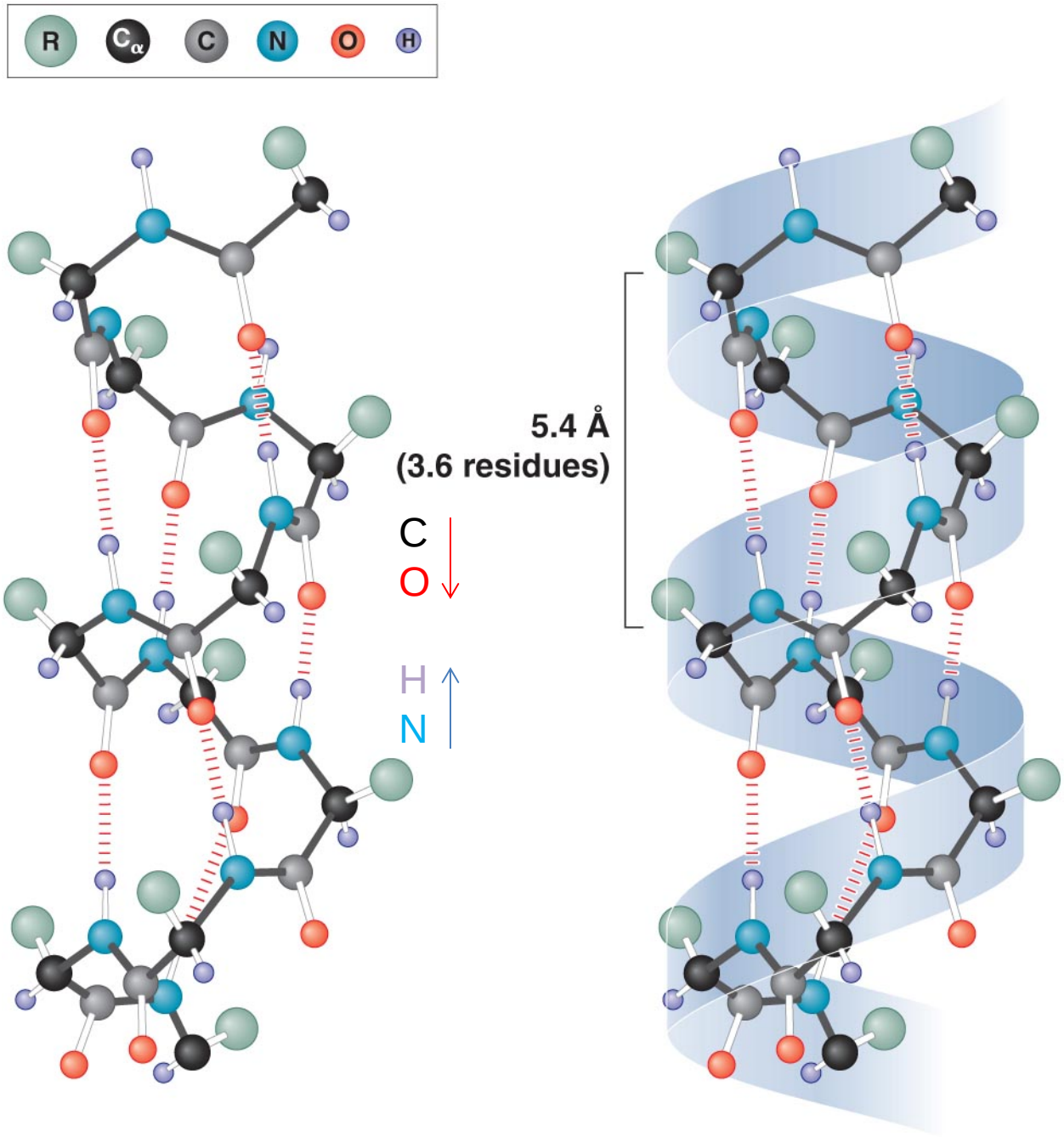
A double proton shuttle and a temporary covalent bond to enzyme stabilize two transition states during chymotrypsin catalysis

Figure 3-26
Molecular Cell Biology, Sixth Edition

Major structure elements: the α-helix

Noncovalent hydrogen bonds between C=O and -N-H stabilize major types of protein structure

C=O interacts with N-H of 4th aa "onward" → C-term



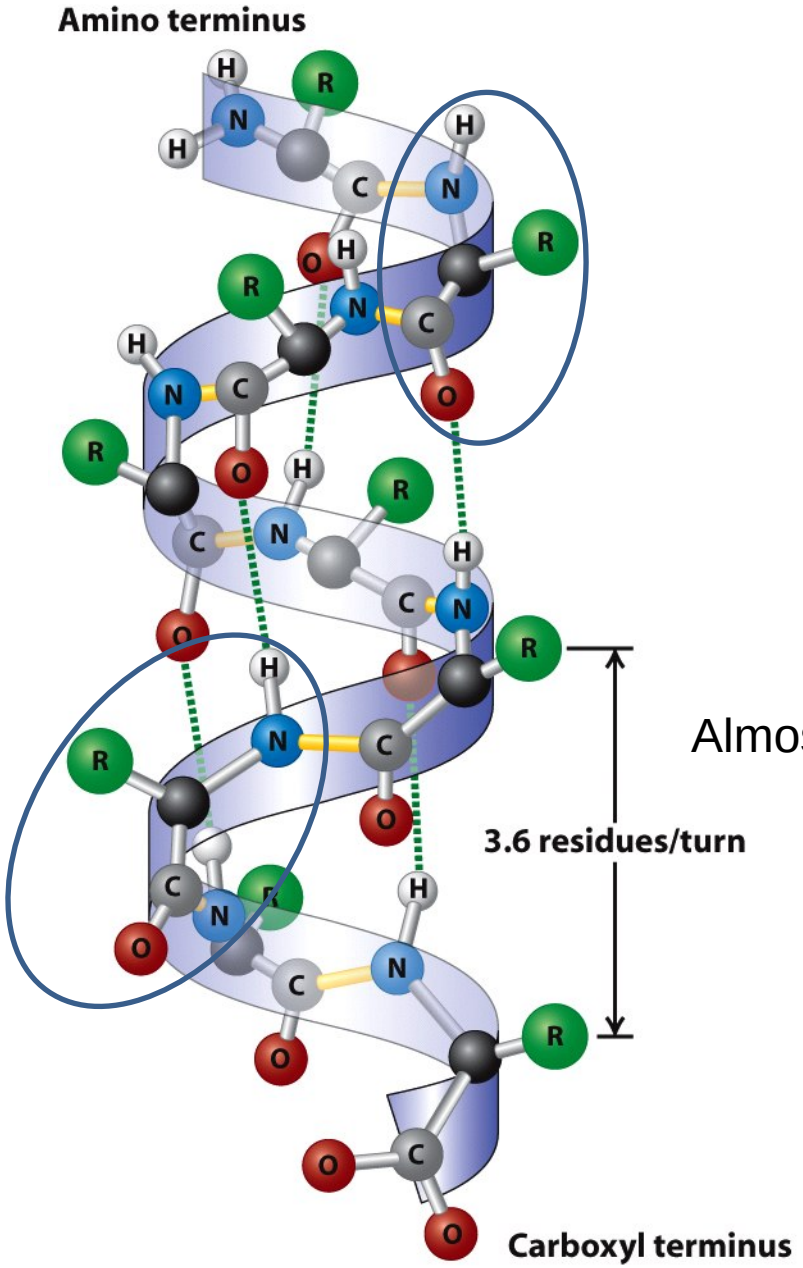


Figure 3-4
Molecular Cell Biology, Sixth Edition

Inter-turn H-bonds occupy peptide backbone polar residues… but side chains bristle in α-helical structure

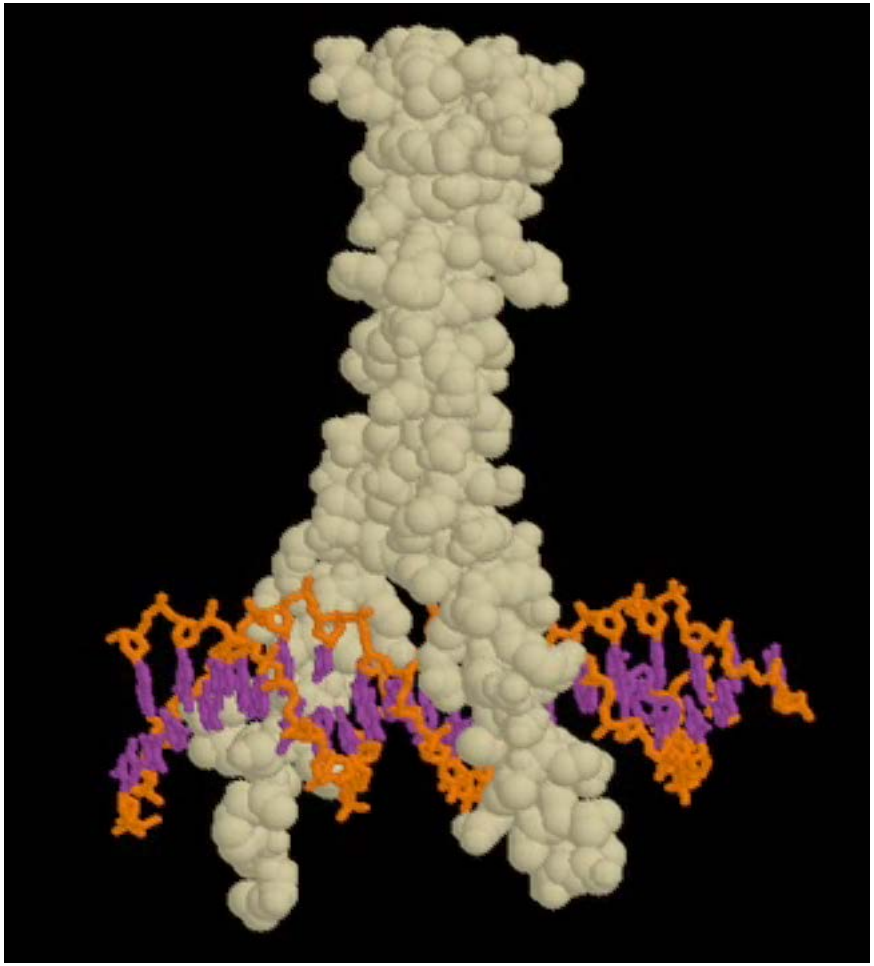
Almost exactly two turns per 7 aas

A very good structure for proteins that must cross a hydrophobic (oily) domain like a membrane, provided that side chains are hydrophobic -- nonpolar

The Leucine zipper motif
a famous coiled coil with strong dimerization power

(side)

(top)



1, 4 repeating positions in heptad are leucines in Leu zipper transcription factors

[movie]

β-pleated sheet: orderly alternation of side chain projections perpendicular to plane of sheet

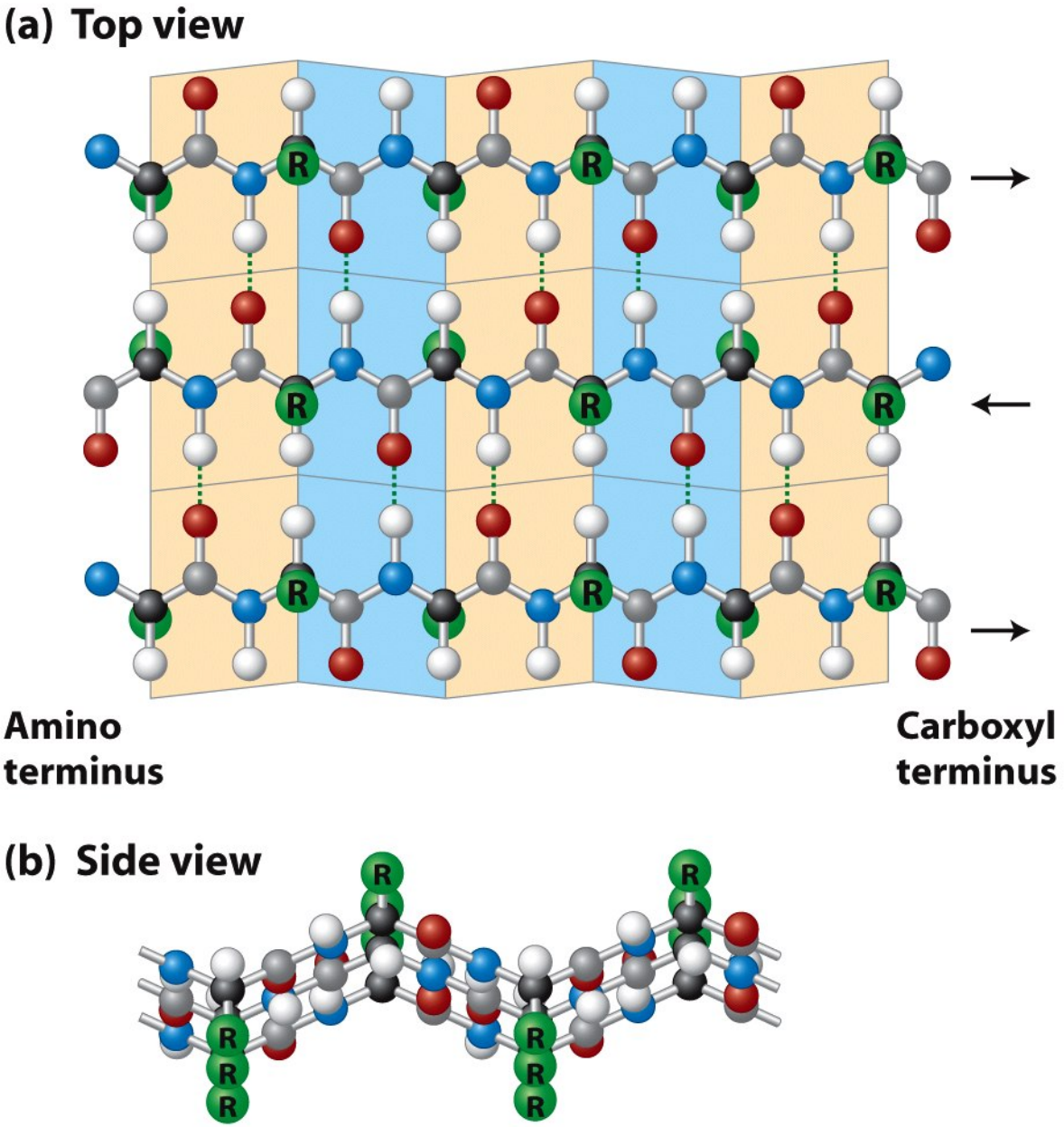
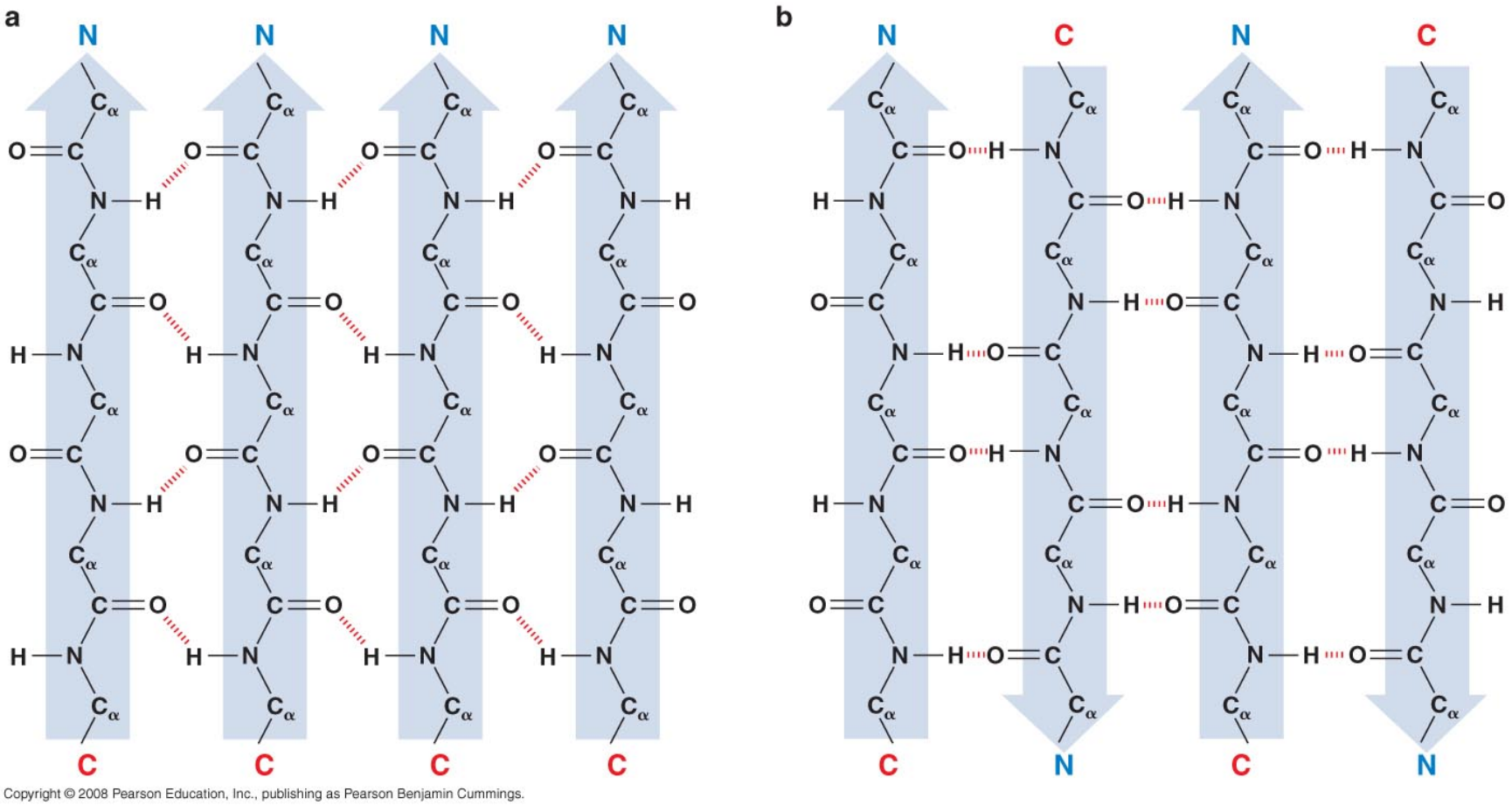


Figure 3-5
Molecular Cell Biology, Sixth Edition

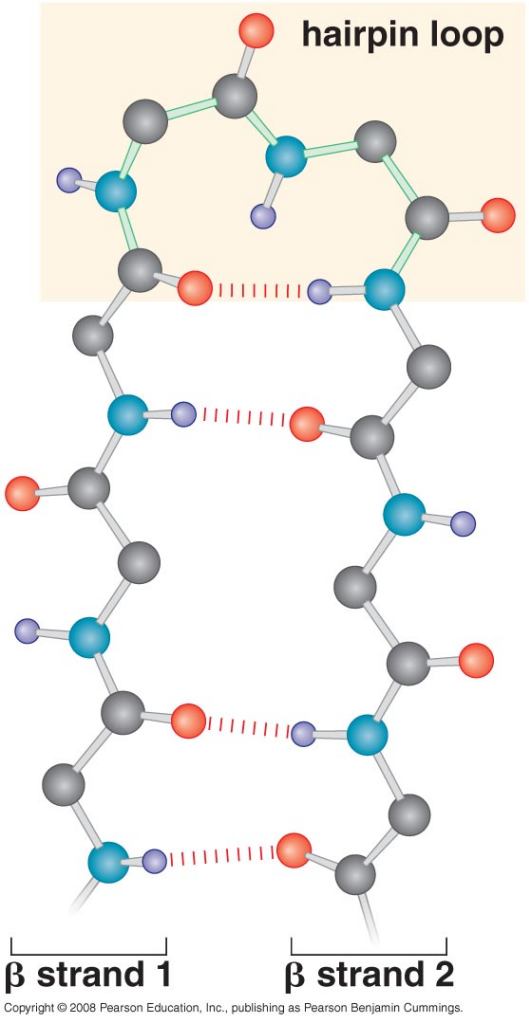
Parallel and antiparallel β-pleated sheets



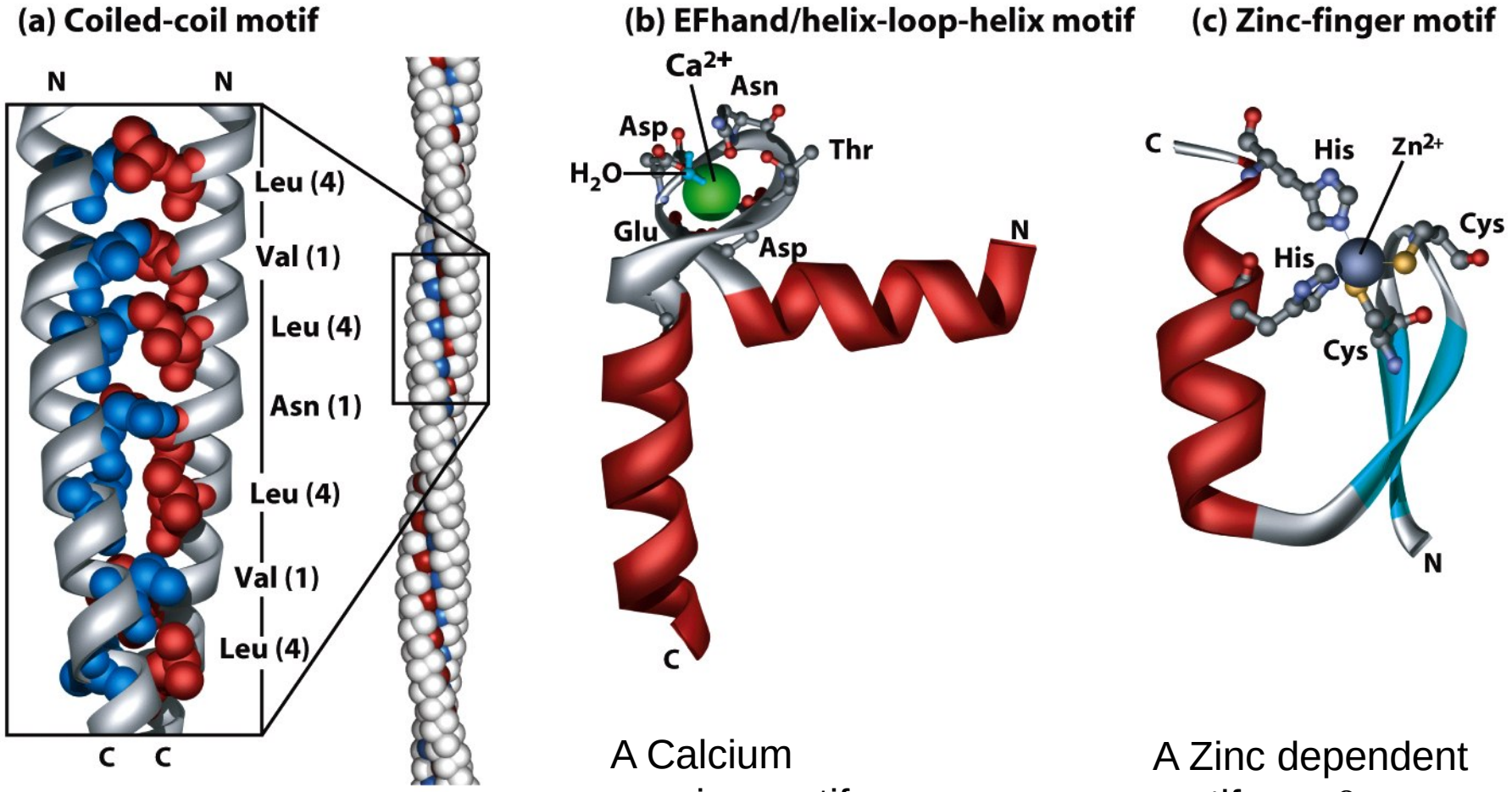
An H-bond strength advantage when N_H…O bonds are straight (alignment of bonding orbitals) – this is antiparallel case

β-strands have just the right dimension to turn corners at ends of strands for antiparallel sheets

Thus many β-pleated sheets are formed by intrachain foldbacks



Secondary structure features assemble into recognizable higher-order motifs

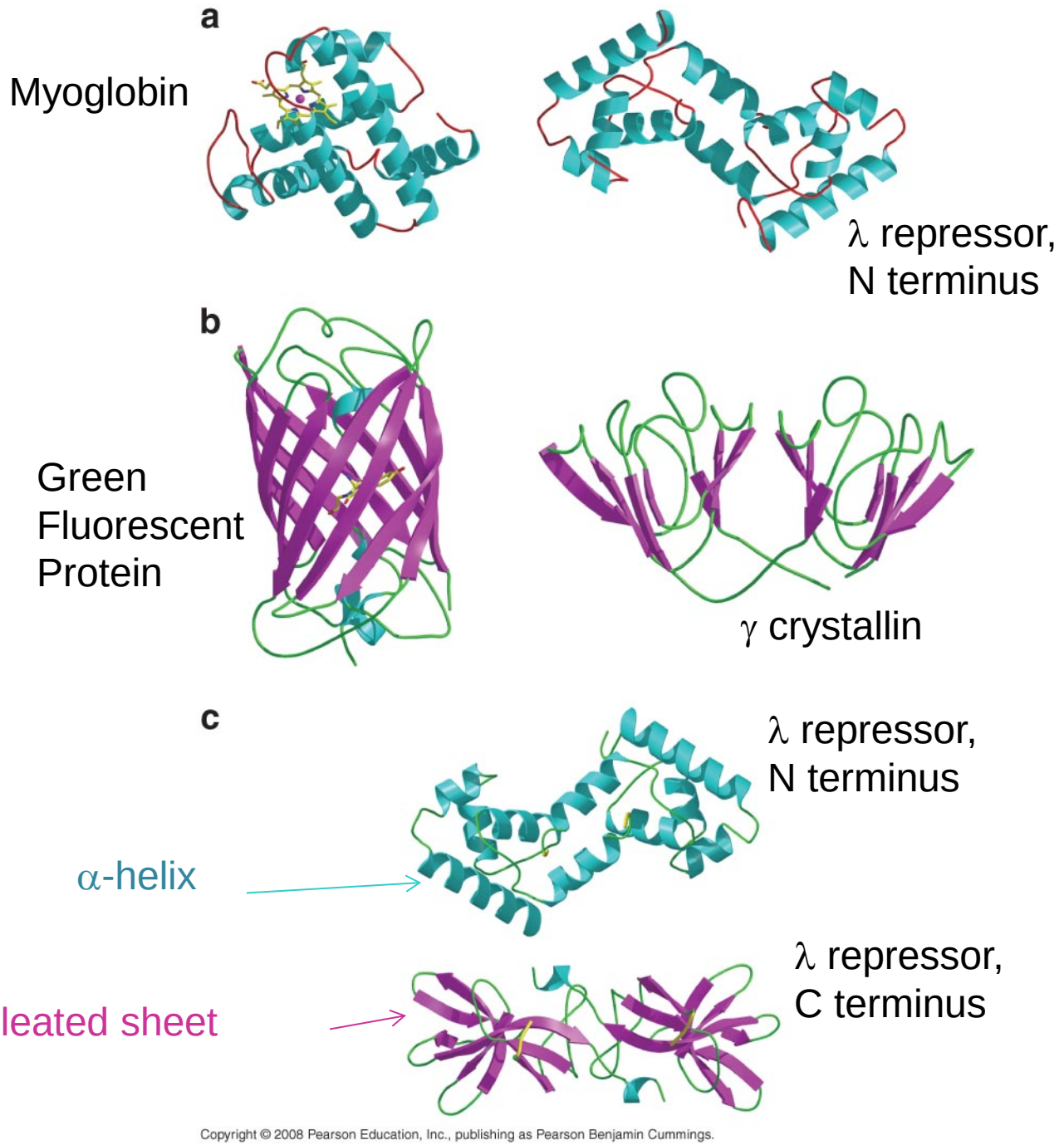


A Calcium sensing motif

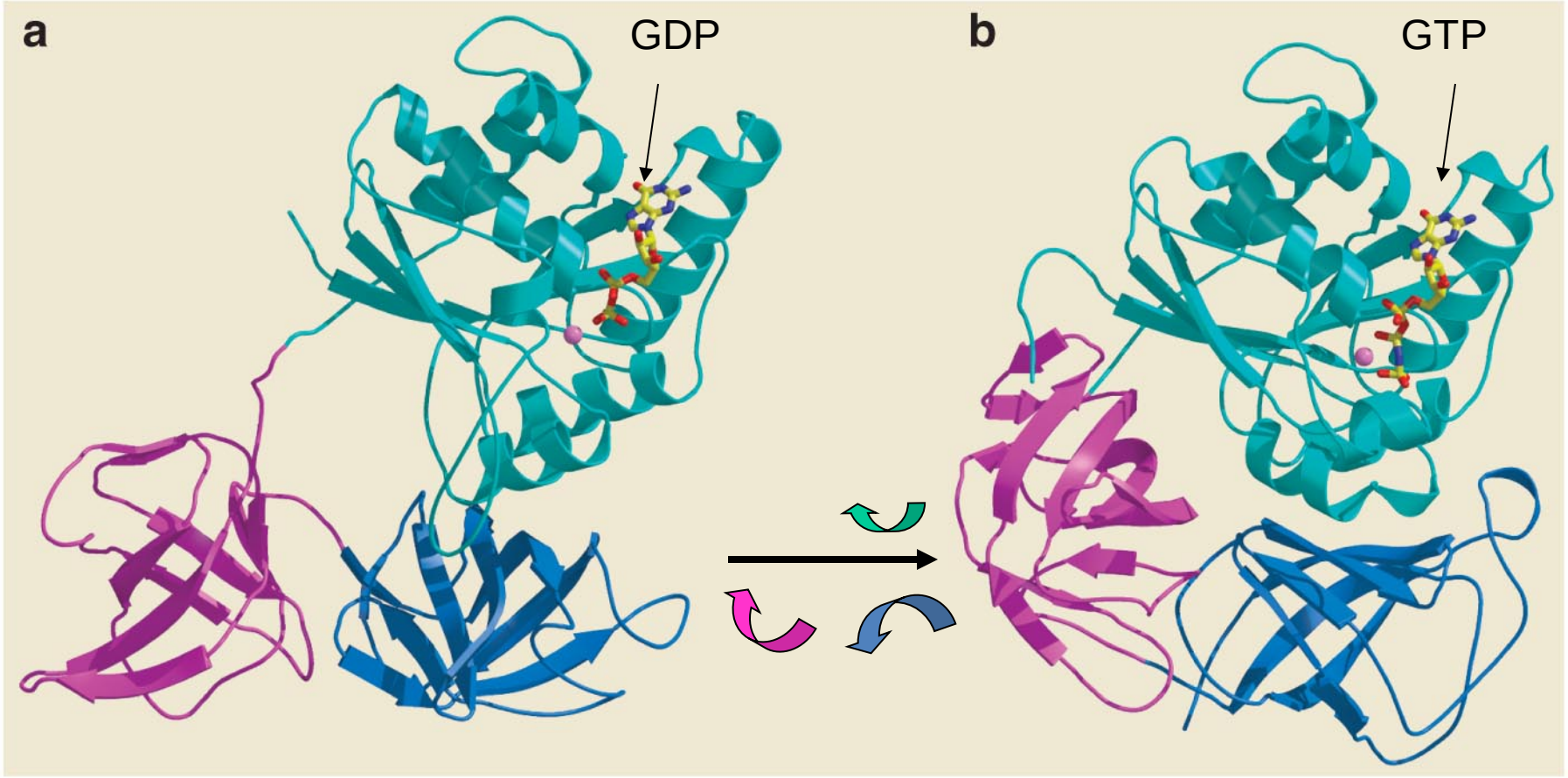
A Zinc dependent motif: $\alpha + \beta$

Figure 3-9
Molecular Cell Biology, Sixth Edition

Proteins or whole protein domains can be dominated by particular secondary structures



Flexibility of different domains in a protein can be the basis of function: Different 3-D protein conformations of EF-Tu factor depend on binding to GTP vs. GDP



Structural impacts of protein folding on function

- Creating a structure of fixed dimensionality to fit in a particular environment
  - Membrane proteins also need to create nonpolar outer surface for membrane-spanning regions
  - Need to create polar core for torus or cylinder with nonpolar exterior if they need to transport charged solutes
- Creating perfect structural complementarity for partners at binding interfaces
- Creating specific alternative structures for switch behavior (EF-Tu & other small GTP binding proteins)
- Positioning key residues in enzyme for substrate binding and for catalysis

A major contribution to inter- and intradomain linkages comes from Cys-Cys disulfide bridges

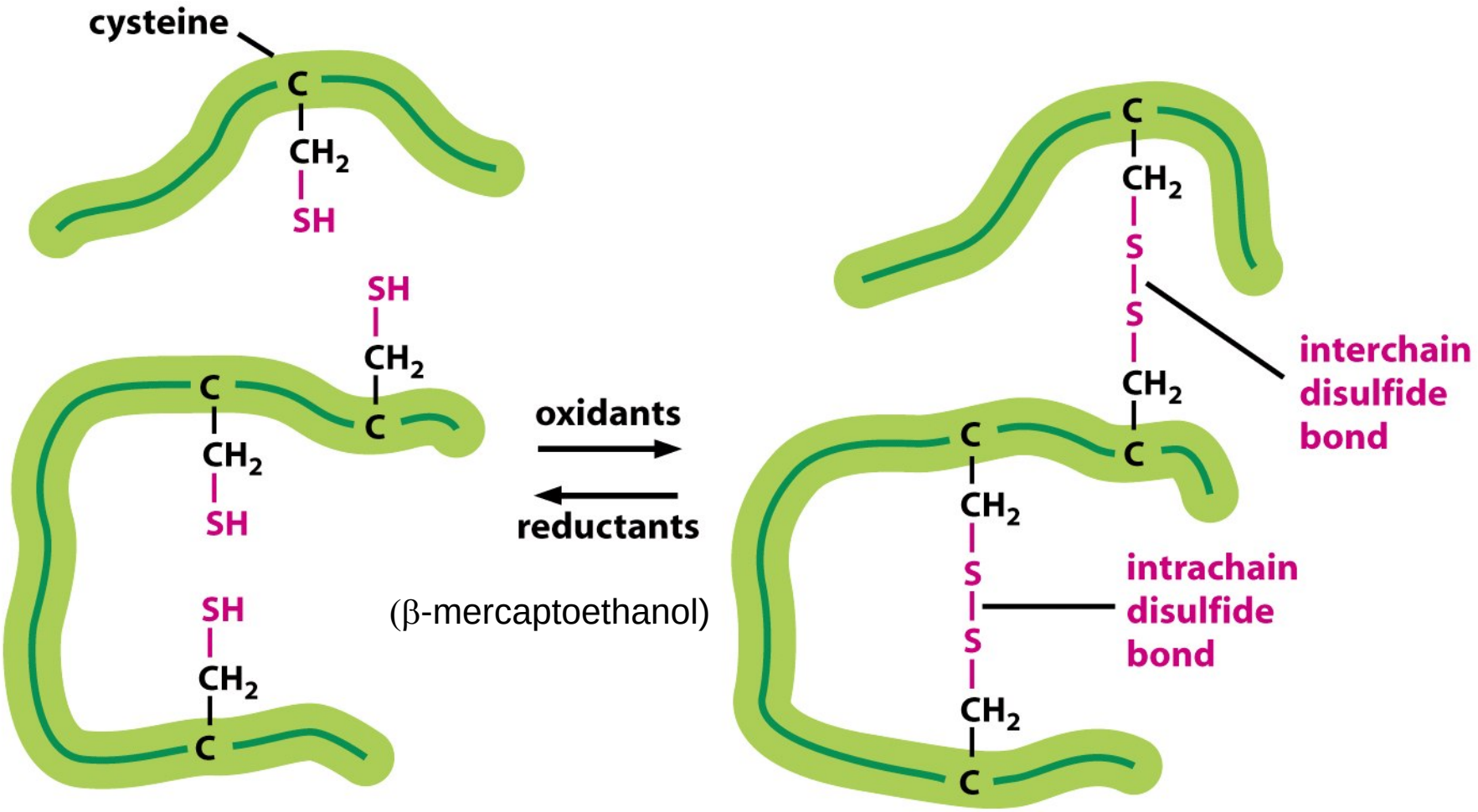


Figure 3-28 *Molecular Biology of the Cell* (© Garland Science 2008)

Sometimes distant amino acids interact in folded structure: creating a binding pocket for a modified nucleotide cofactor: cyclic AMP

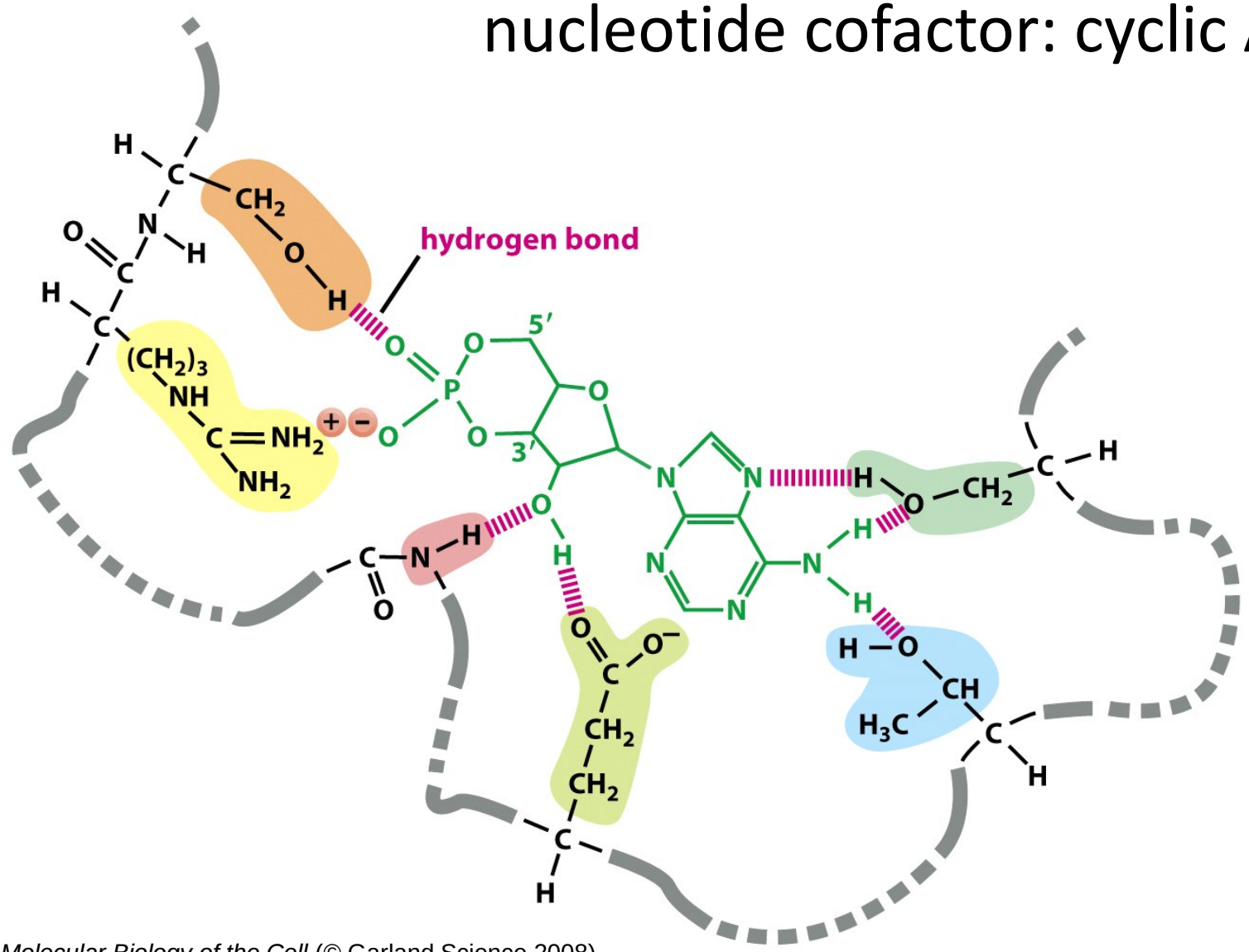


Figure 3-37b *Molecular Biology of the Cell* (© Garland Science 2008)

Example: how protease chymotrypsin holds its substrate in place for cleavage

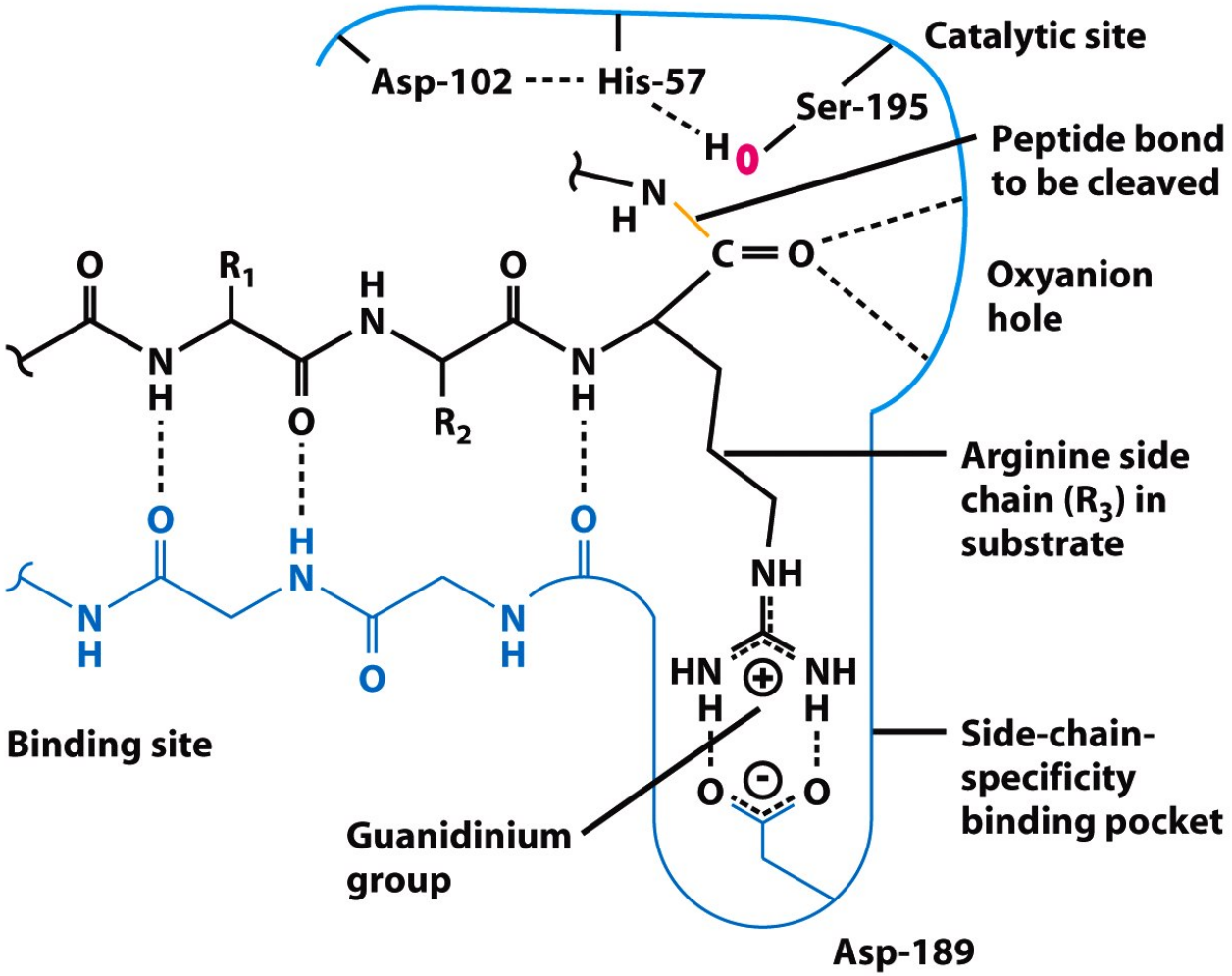
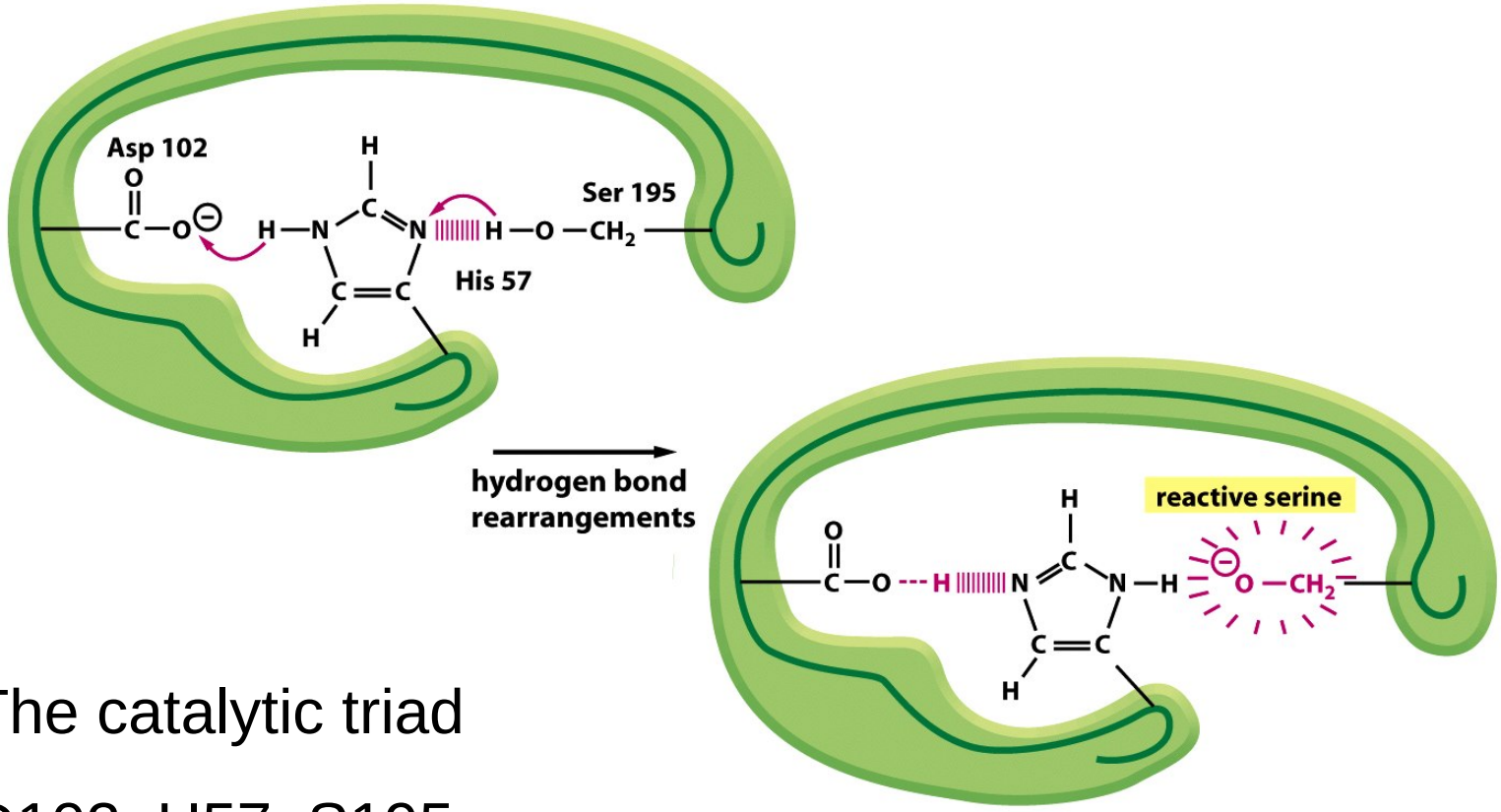


Figure 3-25a
Molecular Cell Biology, Sixth Edition

Activating the active site of chymotrypsin: stripping Ser for action by a proton convoy



The catalytic triad

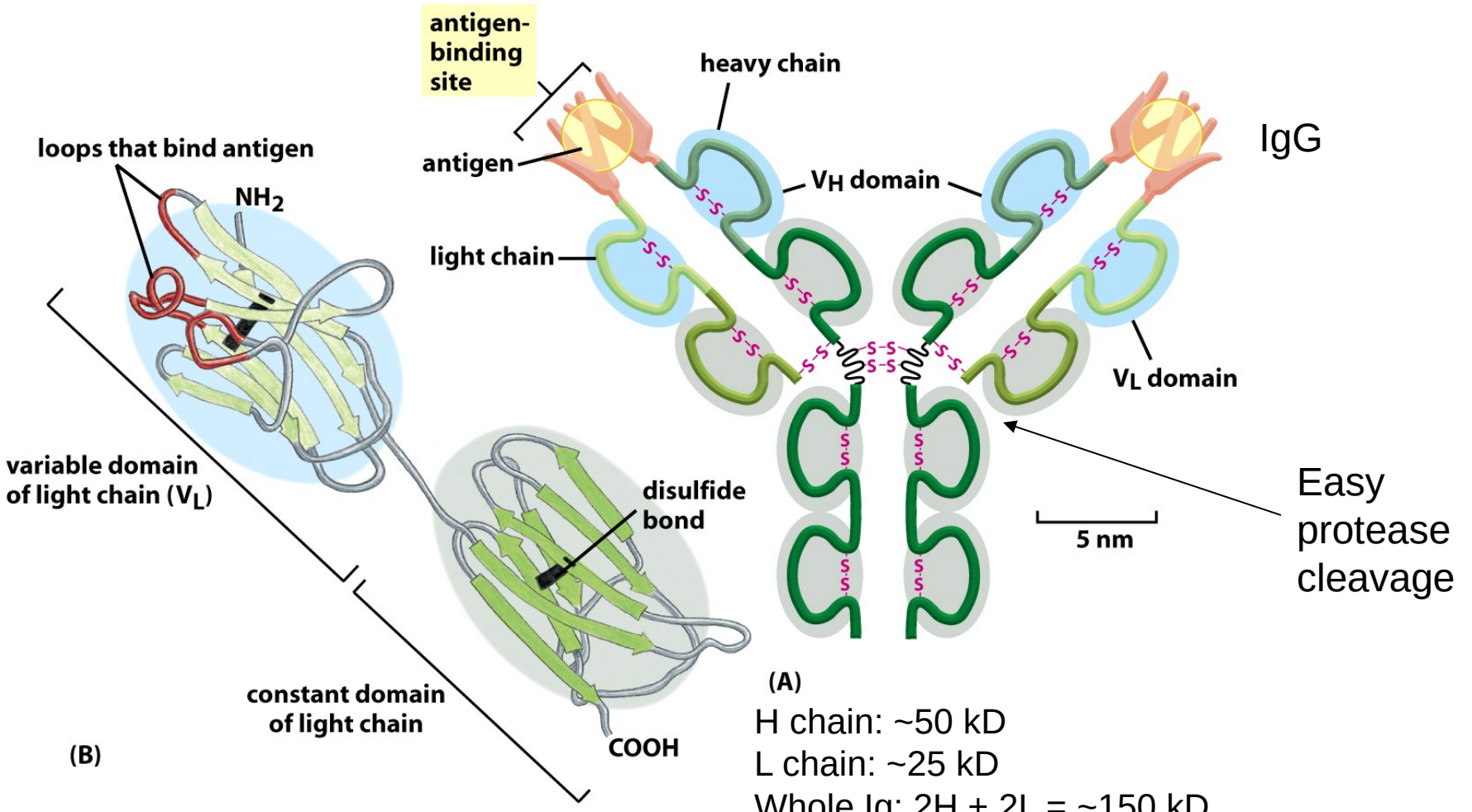
D102: H57: S195

Figure 3-38 *Molecular Biology of the Cell* (© Garland Science 2008)

Initial tools for dissecting tertiary (& quaternary) protein structure in a “new” protein

- Sizing +/- denaturation (multimer vs. individual subunits)
- Sizing +/- disulfide reduction (Cys side chains)
- Reaction with antibody: features can be mapped relative to a particular part of protein that antibody binds
- Limited proteolysis: can allow you to separate distinct, compactly folded domains
- Sequence analysis!!

Chains, domains, and disulfide bonds create structure of immunoglobulin (antibody) molecules



H chain: ~50 kD
L chain: ~25 kD
Whole Ig: 2H + 2L = ~150 kD
+ βME: 2 x 50 kD, 2 x 25 kD; none bind antigen
+ protease: 3 x ~50 kD, some bind antigen
+ both: all 25 kD pieces, none bind antigen

Figure 3-41 *Molecular Biology of the Cell* (© Garland Science 2008)

Nucleic acid manipulation can be key to establishing function of proteins and protein domains

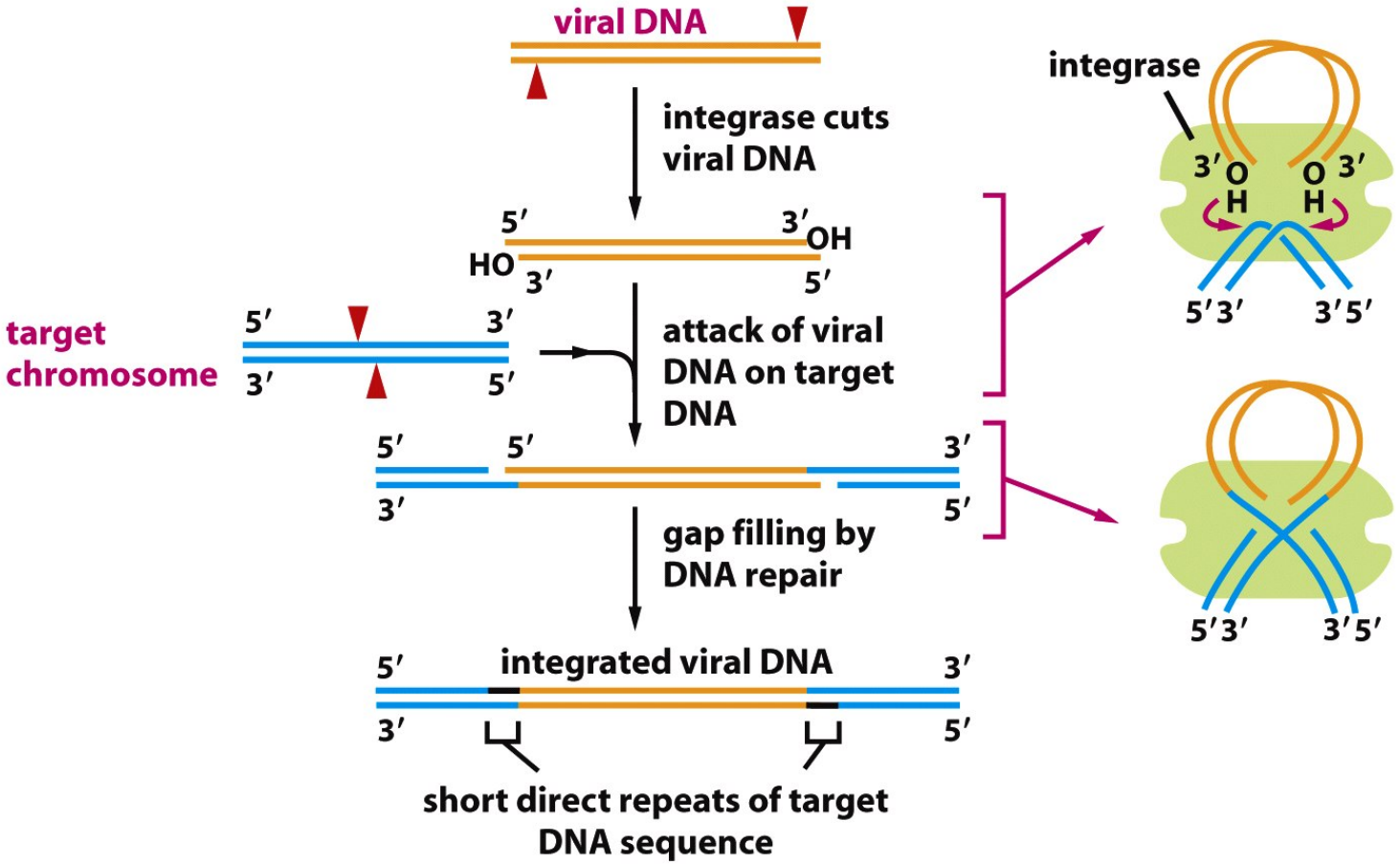
- DNA can be expressed into proteins of interest in vivo or in vitro
  - Transfection into cells in culture (cDNA or genomic)
  - Change in function or protein expression relative to background (hopefully low or zero!)
  - Also available: in vitro, coupled transcription/ translation systems: put in DNA (cDNA cloned in vector with promoter), get out protein
- If you have a functional assay, you can see effect of adding a protein on a system
- With a functional assay, compare effect of adding wildtype vs. protein coded by deleted or mutated cDNA
  => infer key domains, key amino acids!!

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

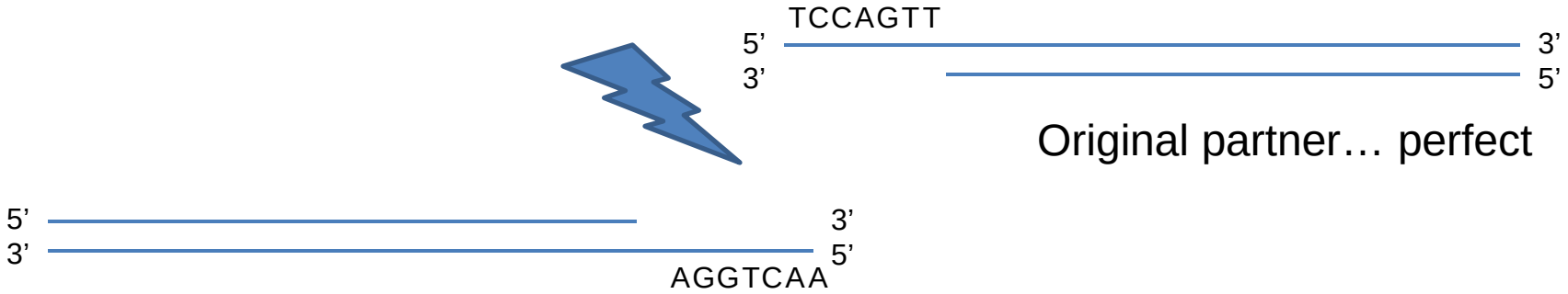
Viruses in prokaryotes and eukaryotes modify host genomes "for a living": they encode their own equivalents of restriction enzymes



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)

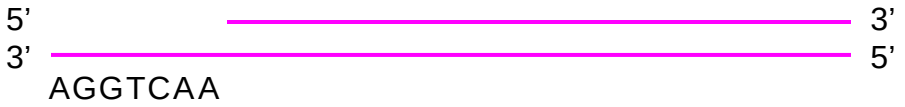
Sequence homology, even over short distance, can enhance DNA break repair



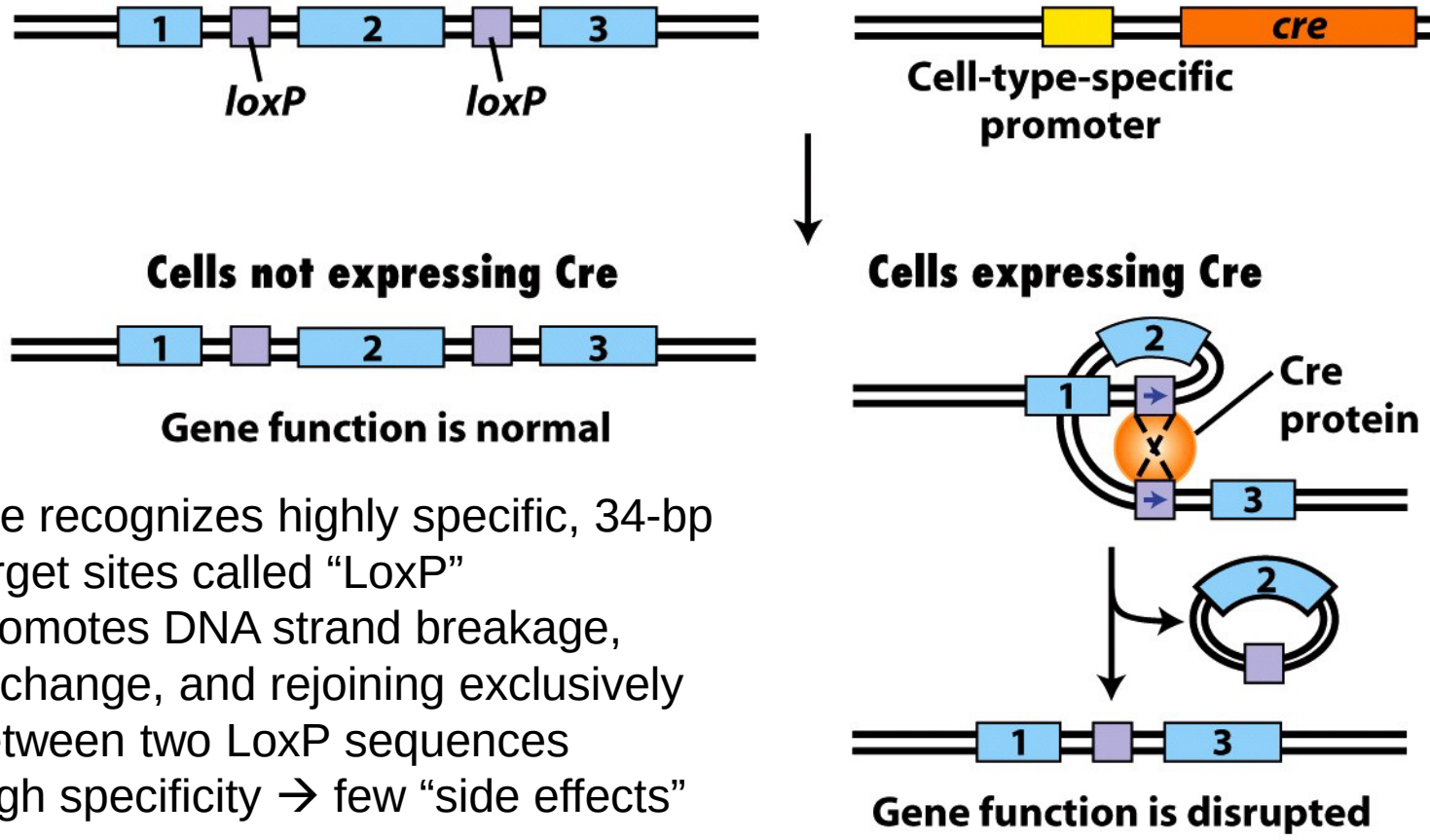
Poor partner for resealing

Imperfect … but adequate!

Poor partner for resealing

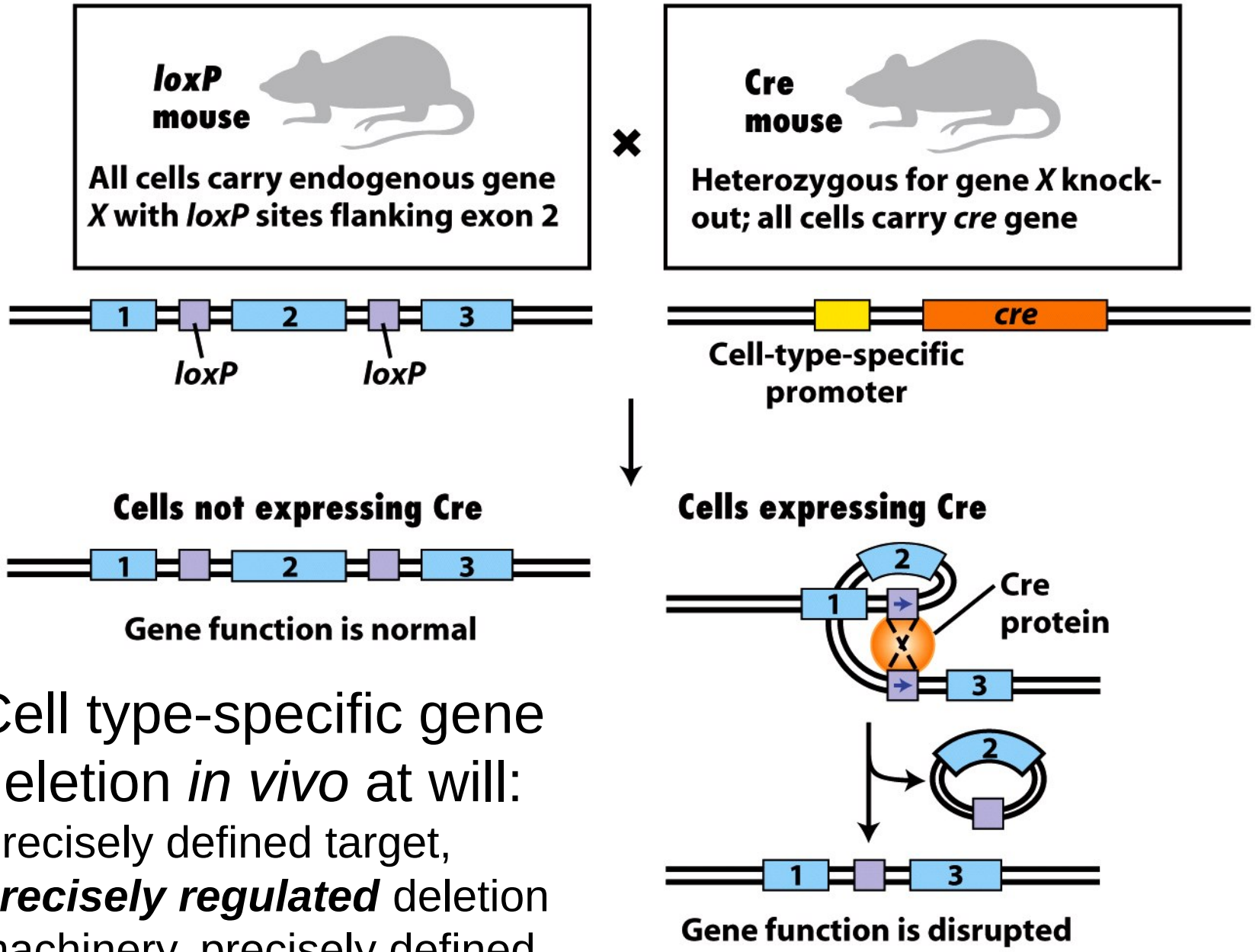


Cre: A particularly useful recombination enzyme – site-specific cutting and rejoining from a single enzyme



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



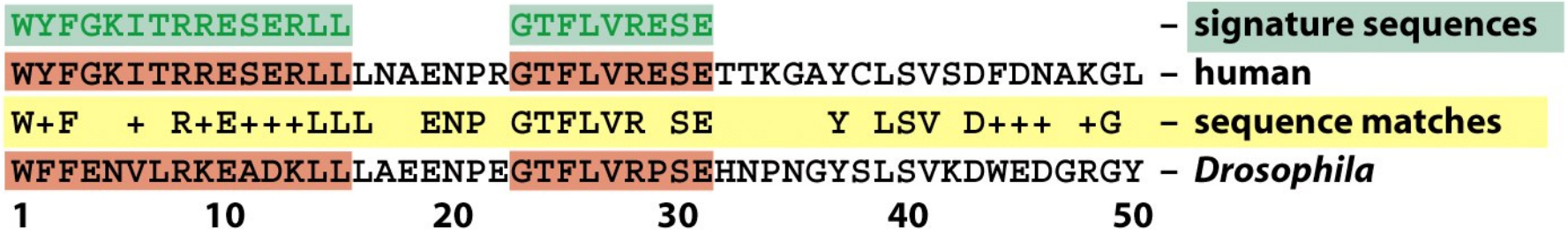
Cell type-specific gene deletion *in vivo* at will:

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

Related functions and structures can be predicted for proteins in distant species or among distant family members just based on sequences of genes that code for them…

… educated guessing even before starting biochemistry!

- Find possible exons predicted from continuous reading frames
- Virtual translation "*in silico*"
- Similarities can reveal matches to other known proteins in database



Human and Drosophila Src: SH2 (protein interaction) domain

Generally substantially the same folding even if only 20-25% sequence identity! (especially if conserved residues are known to be in key points of structure based on well-studied examples)

Figure 3-14 *Molecular Biology of the Cell* (© Garland Science 2008)

Deep evolutionary conservation of protein structure even when amino acid sequence has drifted (homeodomain proteins)

Conservation of ***particular*** amino acids that are needed for structural function in particular domains is more important than ***overall*** conservation

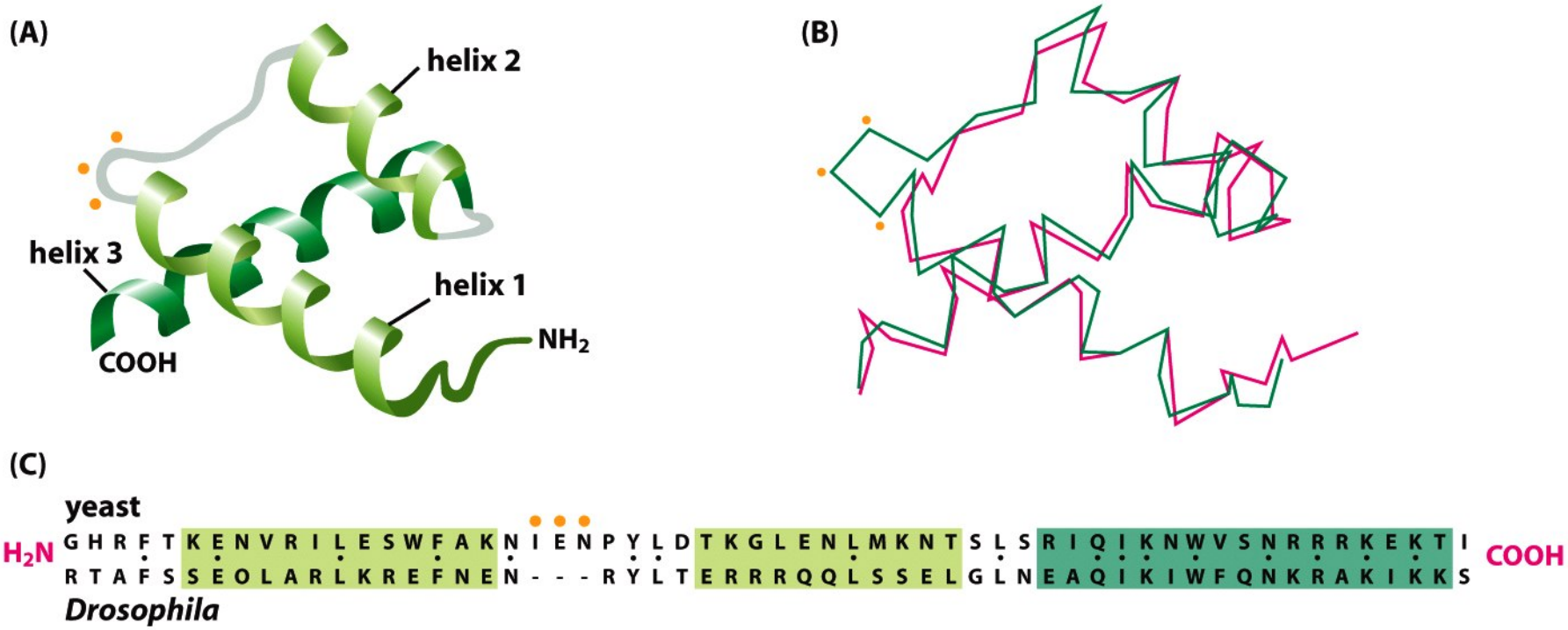
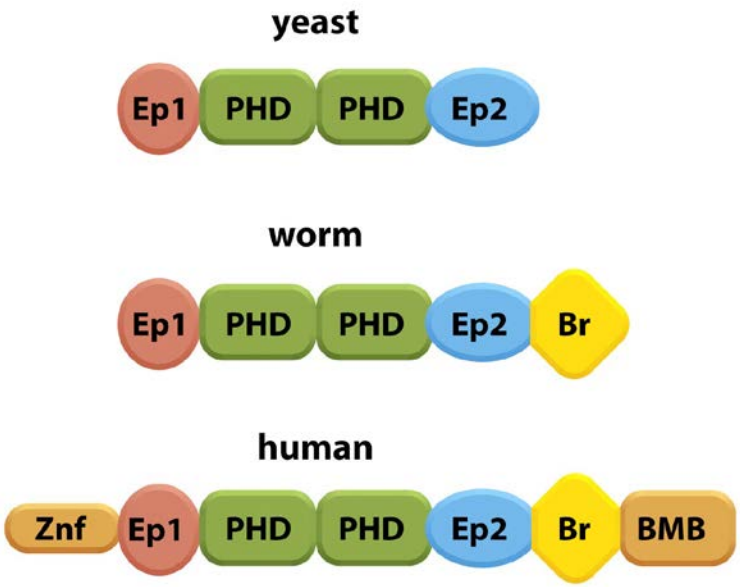
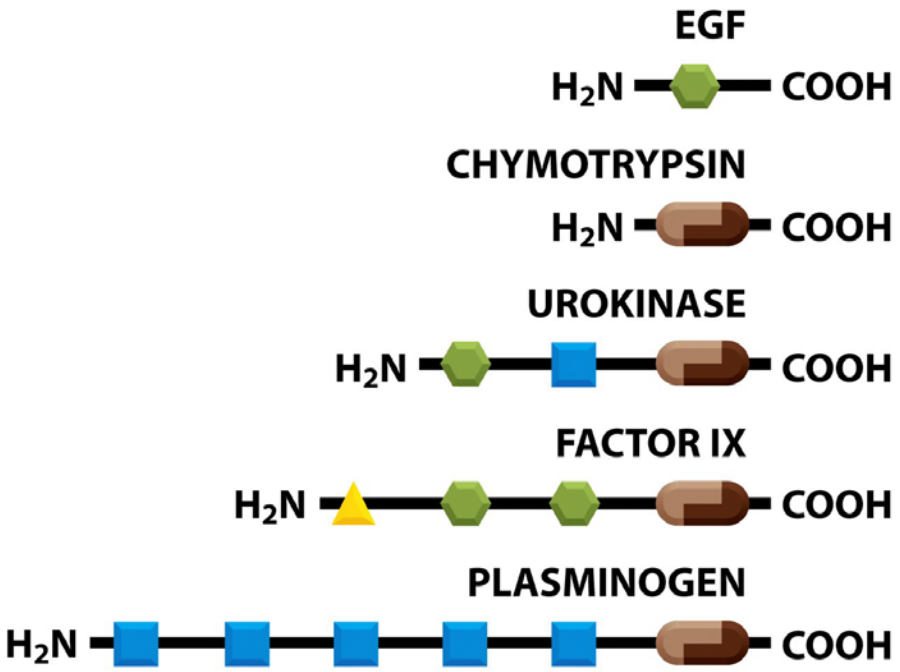


Figure 3-13 *Molecular Biology of the Cell* (© Garland Science 2008)

Domain based modules are the foundation of protein evolution: match, multiply, and mix





Modularity is easy to encode in genome when domains are encoded by discrete exons… introns give lots of room for copying,cutting & pasting

Figure 3-19 *Molecular Biology of the Cell* (© Garland Science 2008)