

# Protein Domains

**Το domain είναι μια πολυπεπτιδική αλυσίδα (ή μέρος αυτής) που μπορεί να αναδιπλωθεί ανεξάρτητα σε μια σταθερή συμπαγή τριτοταγή δομή ή να αναδιπλωθεί**

**Το module είναι μια εξαιρετικά συντηρημένη αλληλουχία που παρατηρείται σε διαφορετικά πλαίσια σε πρωτεΐνες πολλαπλών domain.**

# Domains in proteins

**TABLE 2.1 Percentage of single and multiple-domain proteins in different genomes**

| Genome group | Number of domains in protein |           |          |           |          |          |
|--------------|------------------------------|-----------|----------|-----------|----------|----------|
|              | Exactly 1                    | Exactly 2 | $\geq 2$ | Exactly 3 | $\geq 3$ | $\geq 4$ |
| Archaea      | 36                           | 9         | 43       | 2         | 9        | 2        |
| Bacteria     | 35                           | 10        | 42       | 2         | 10       | 2        |
| Yeast        | 22                           | 5         | 57       | 1         | 5        | 3        |
| Metazoa      | 23                           | 4         | 52       | 1         | 4        | 7        |

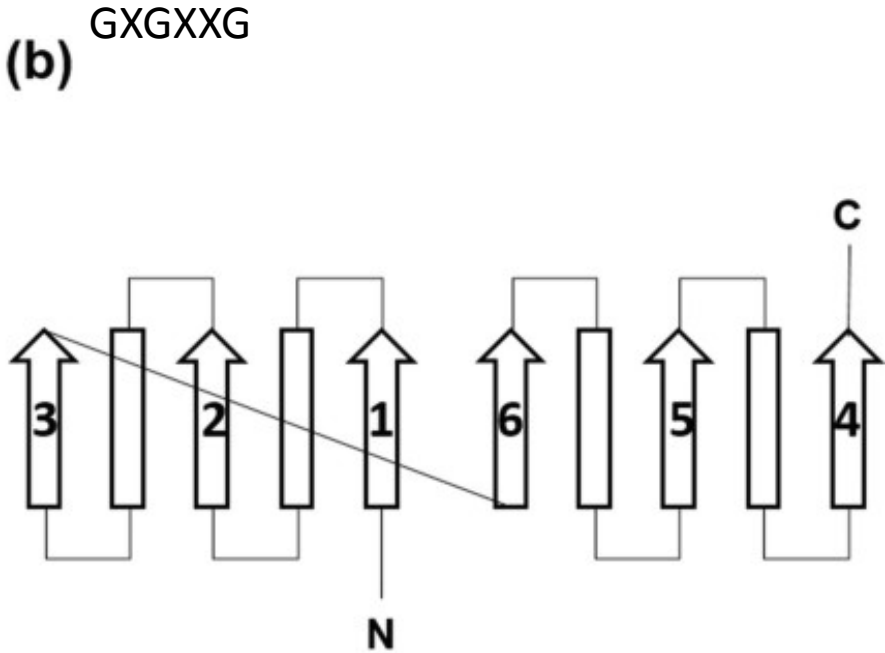
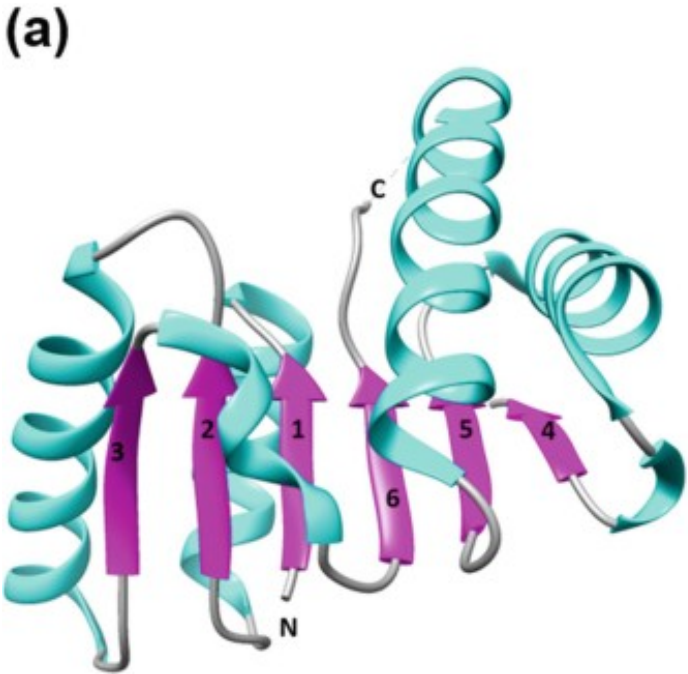
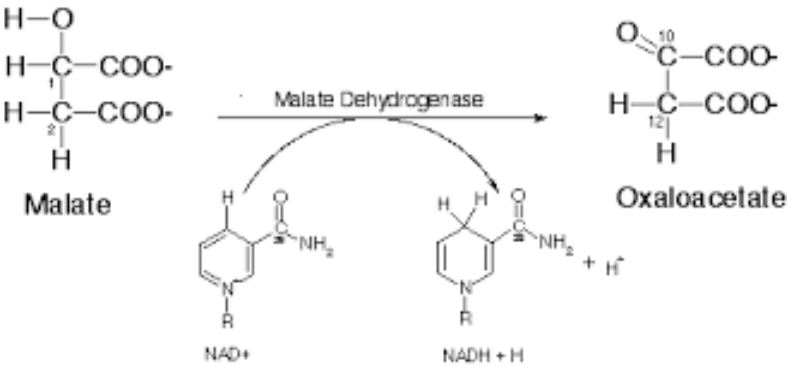
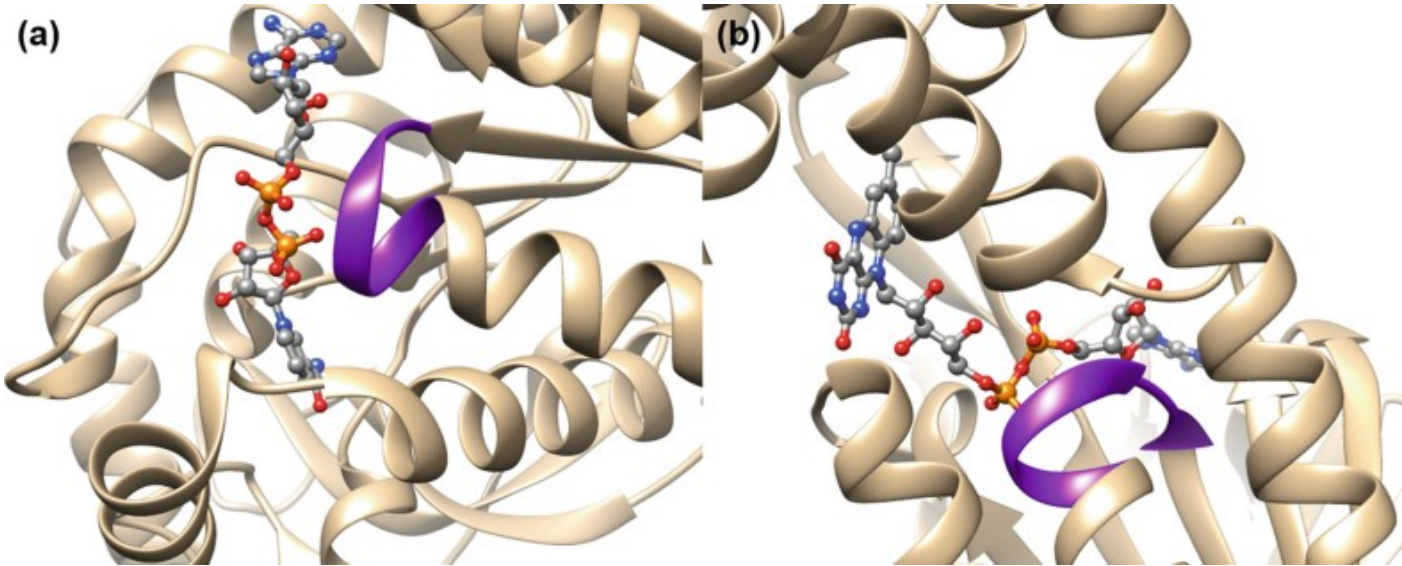
These data are derived from fitting the sequences to domains as defined in SCOP. “Exactly 1” means that the whole sequence fits precisely to a single domain, whereas “ $\geq 2$ ” means that only part of the sequence fits to a single domain, implying that there must be more than one domain present. Similar arguments apply to higher numbers. Thus, the percentage in the column “ $\geq 3$ ”, for example, is a minimum, because some of the proteins in the column “ $\geq 2$ ” will in fact have three or more domains. (Data from G. Apic, J. Gough and S.A. Teichman, *J. Mol. Biol.* 310: 311–325, 2001. With permission from Elsevier.)

# Functions of Domains

**Rossmann fold**

Domains often have clearly identifiable functions.

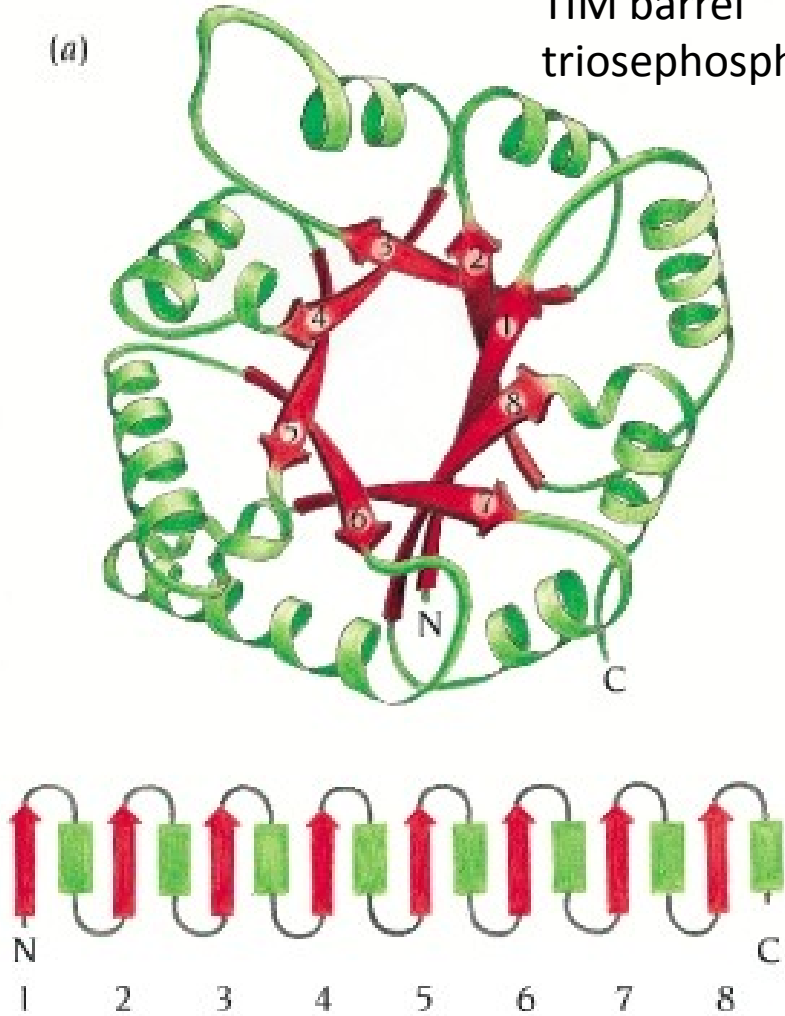
malate dehydrogenase



# Functions of Domains

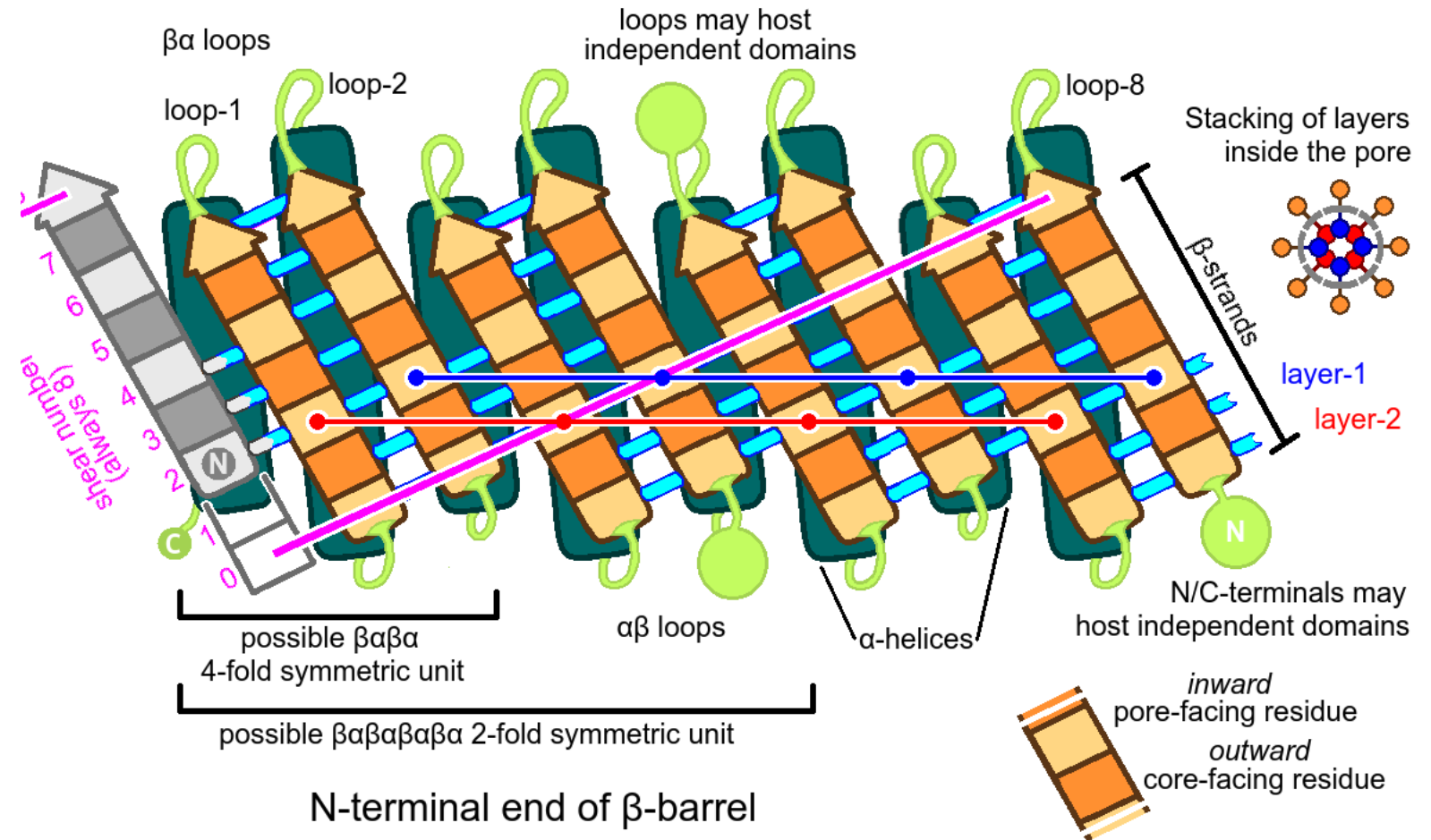
b barrel formed by a series of b-a-b **structure motifs**

(a) TIM barrel  
triosephosphate isomerase



C-terminal end of  $\beta$ -barrel  
contains active site (catalytic) residues

7  $\beta$ -strand residues are shown here  
actual  $\beta$ -strand lengths vary  
amongst TIM barrels



N-terminal end of  $\beta$ -barrel

# Functions of Domains

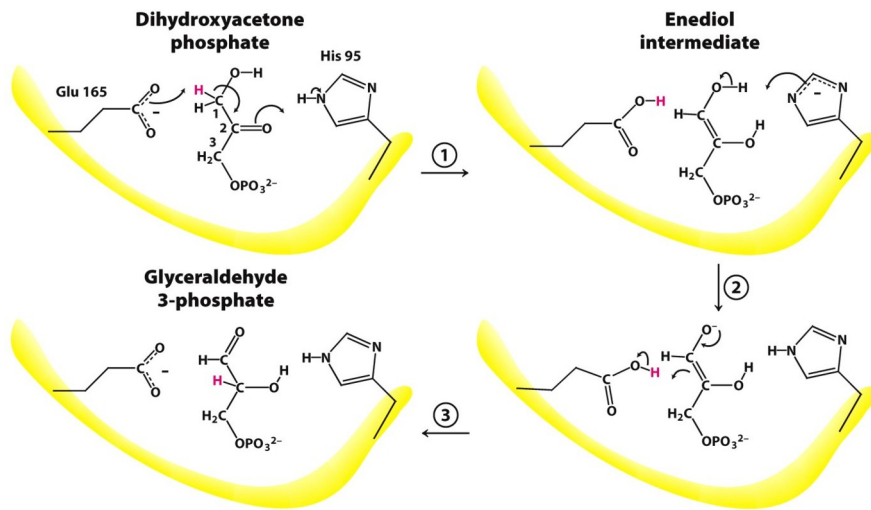


Figure 16.5  
Biochemistry, Seventh Edition  
© 2012 W. H. Freeman and Company

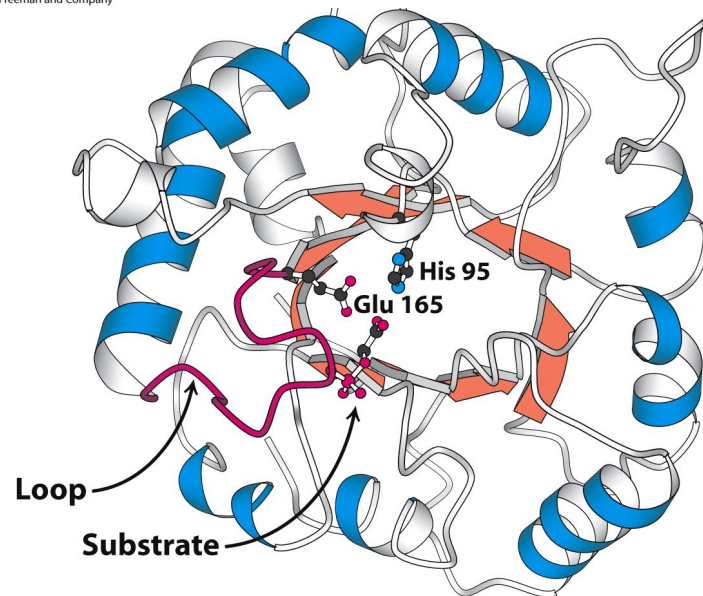
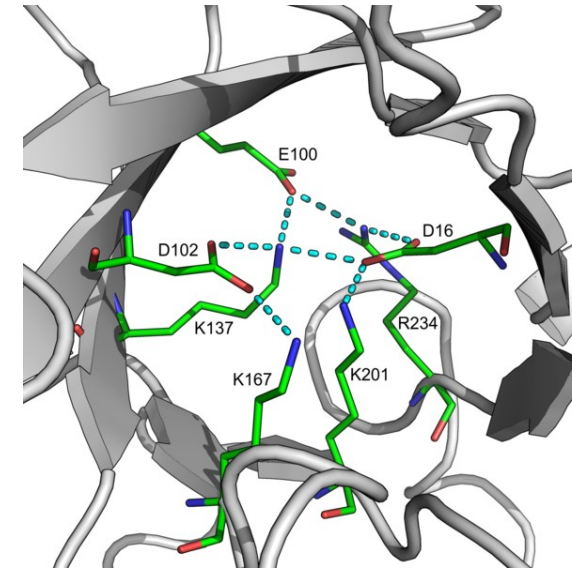
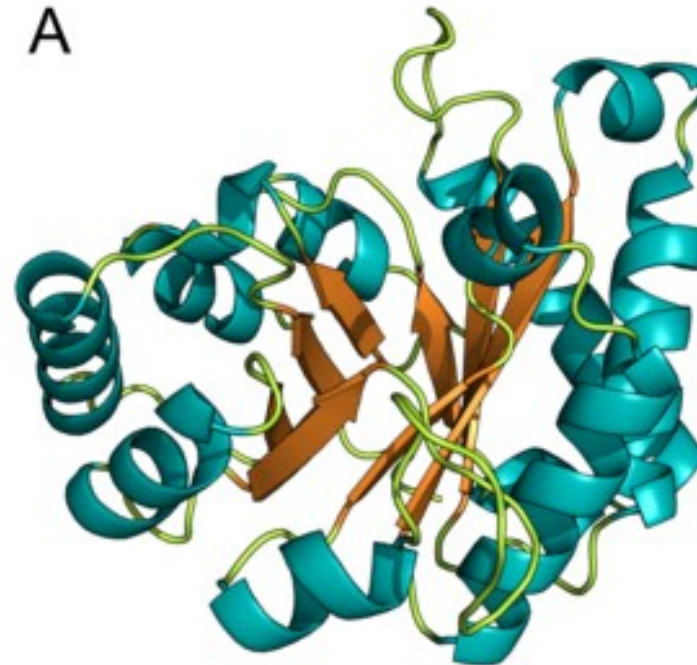


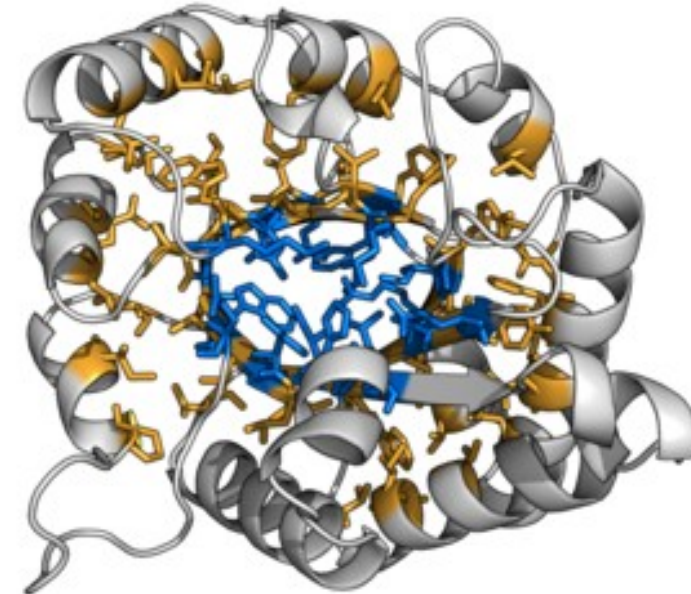
Figure 16.4  
Biochemistry, Seventh Edition  
© 2012 W. H. Freeman and Company



A

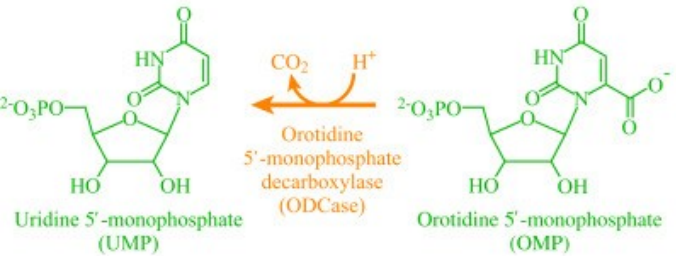


B

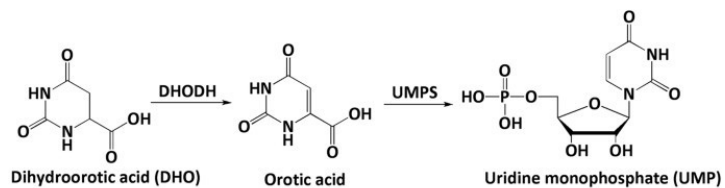


# Structural inserts

*Bacillus subtilis* Orotidine 5'-monophosphate decarboxylase

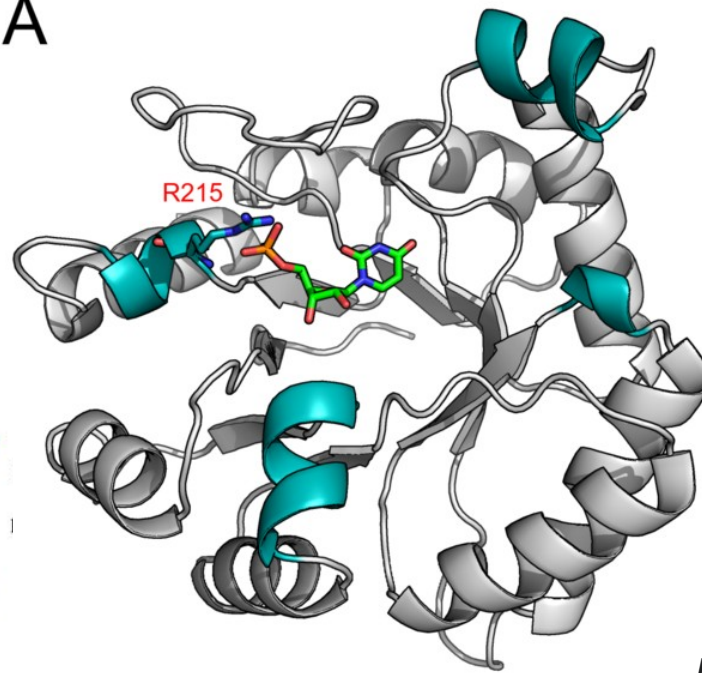


*Lactococcus lactis* dihydroorotate dehydrogenase

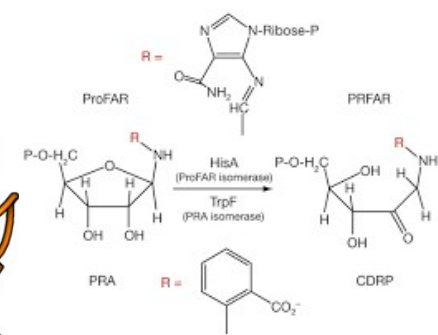
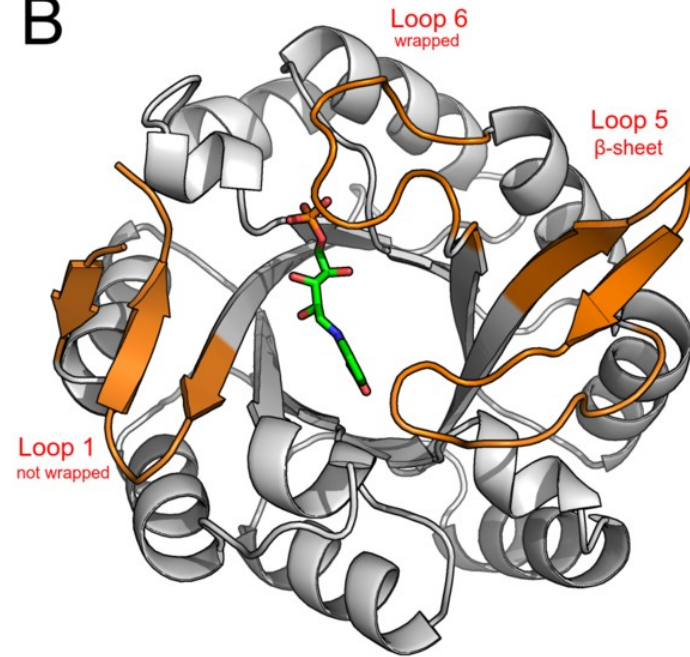


DHODH: dihydroorotate dehydrogenase  
UMPS: uridine monophosphate synthase

A

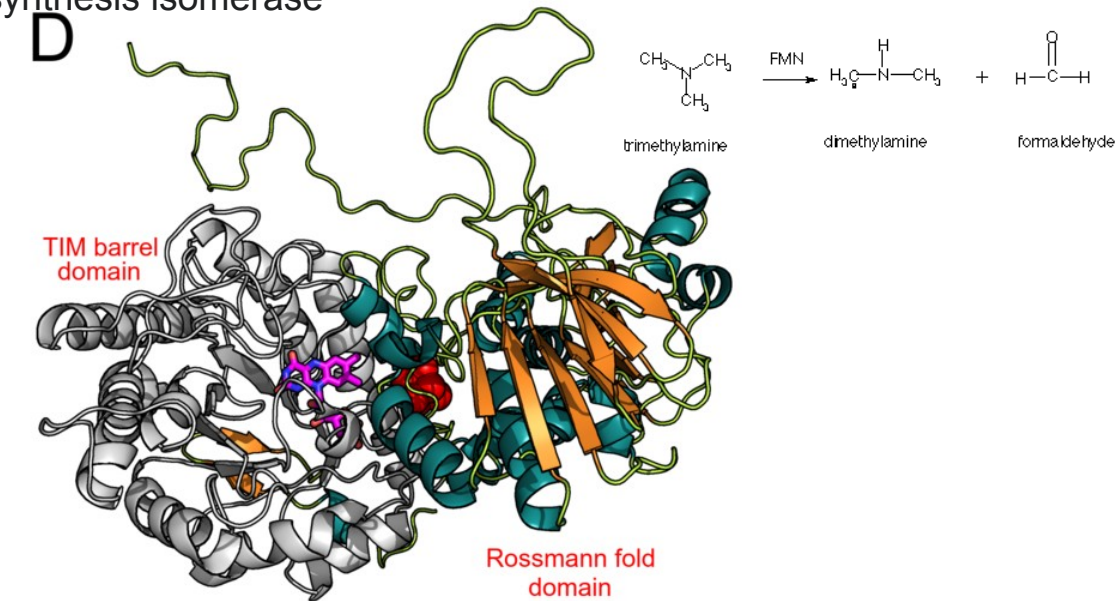


B



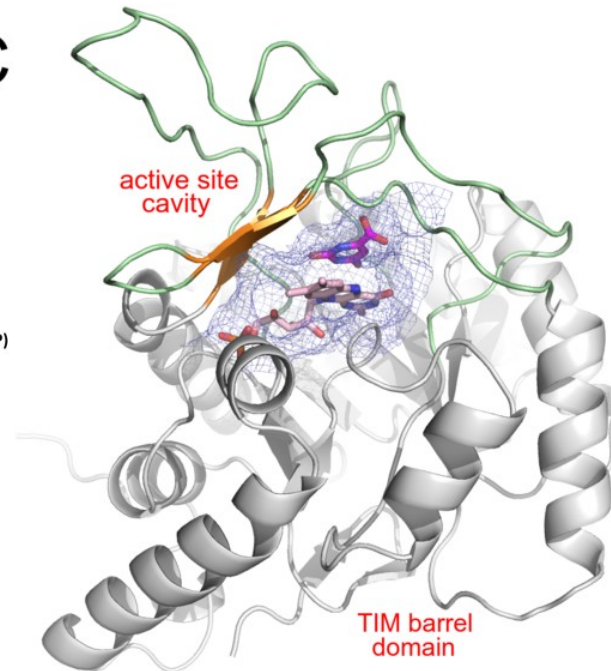
*Mycobacterium tuberculosis* bifunctional histidine/tryptophan biosynthesis isomerase

D

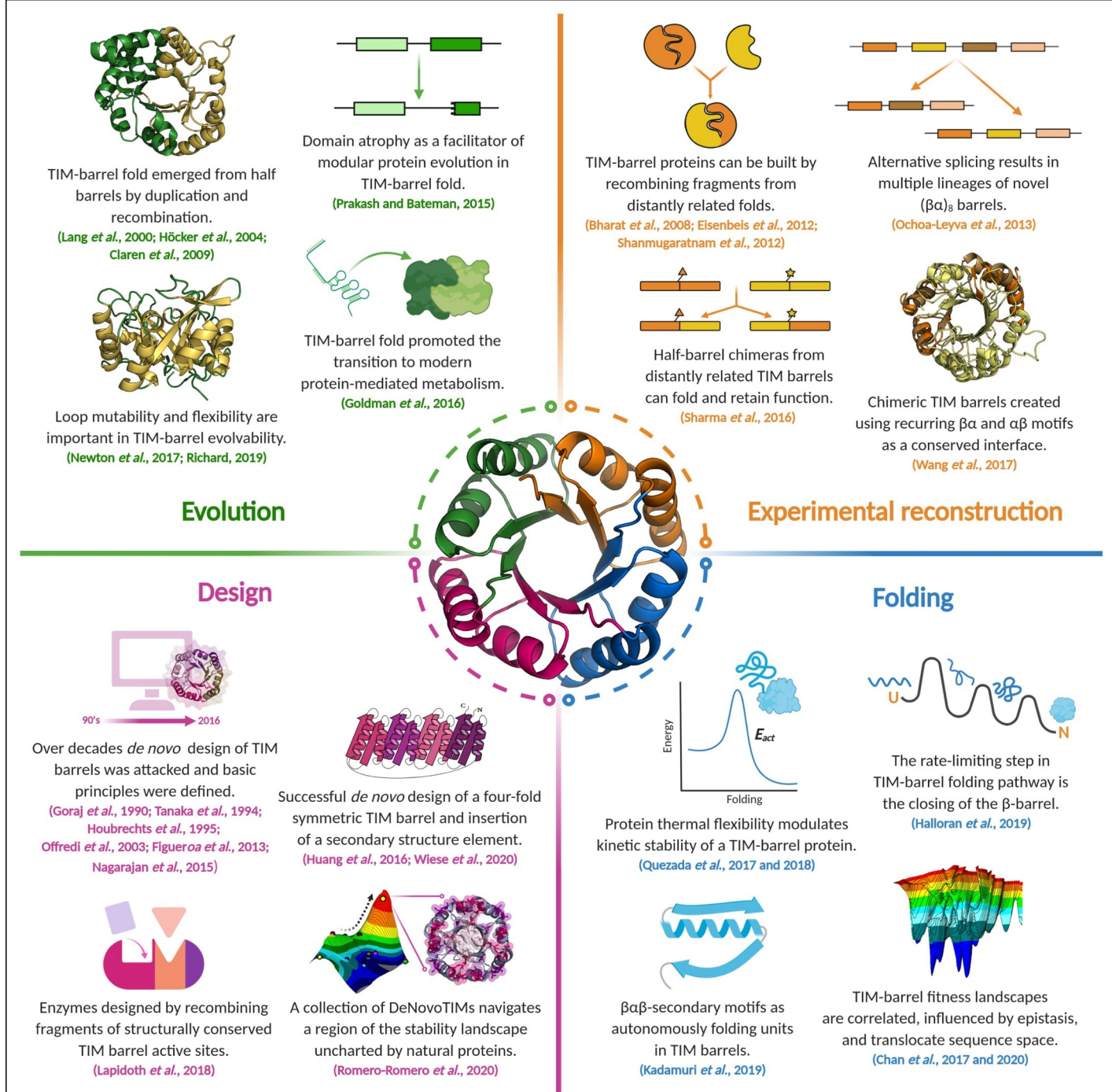


*Methylophilus methylotrophus* trimethylamine dehydrogenase

C



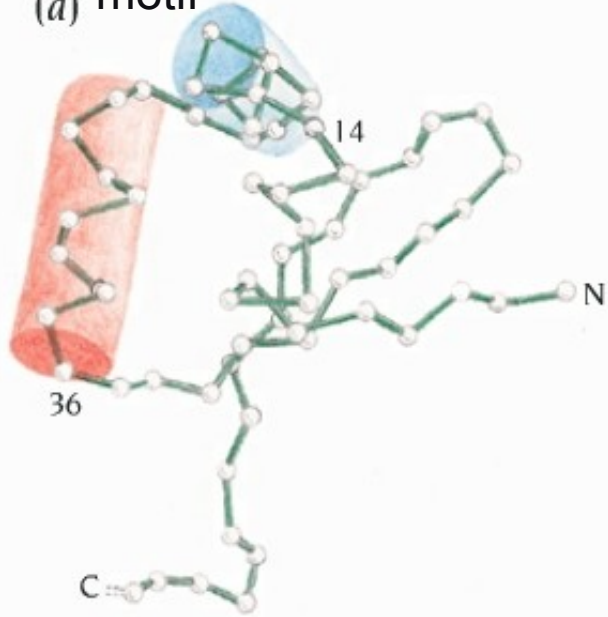
# Evolution, folding, and design



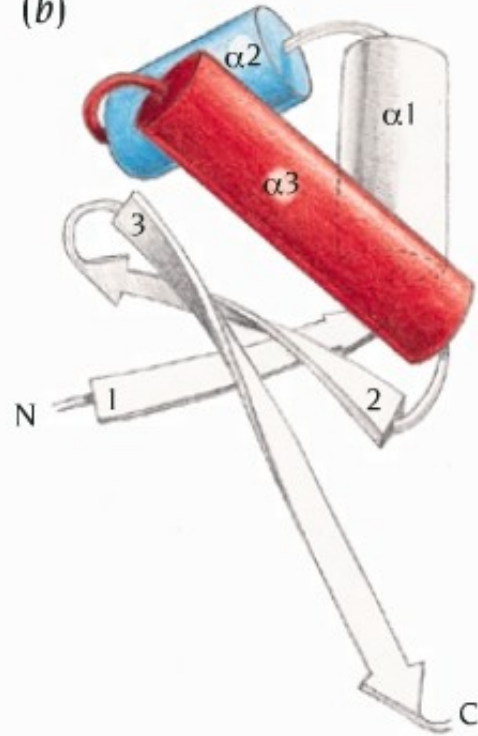
# Functions of Domains

“helix–turn–helix” (HTH) motif

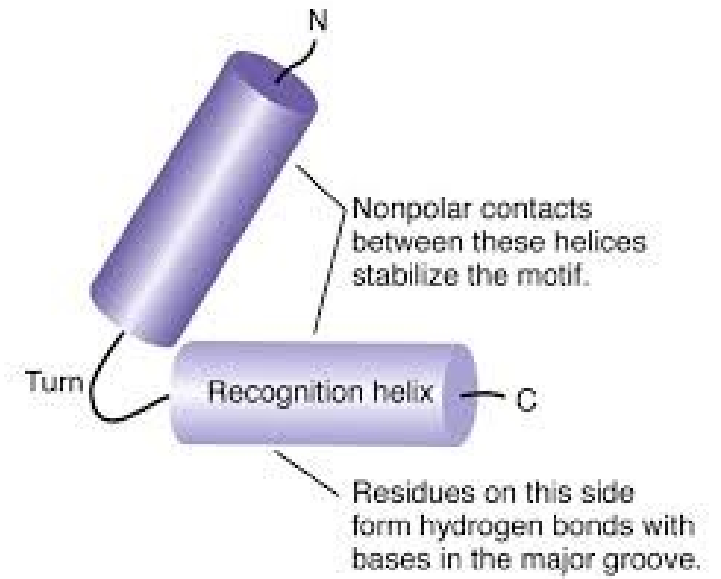
(a) motif



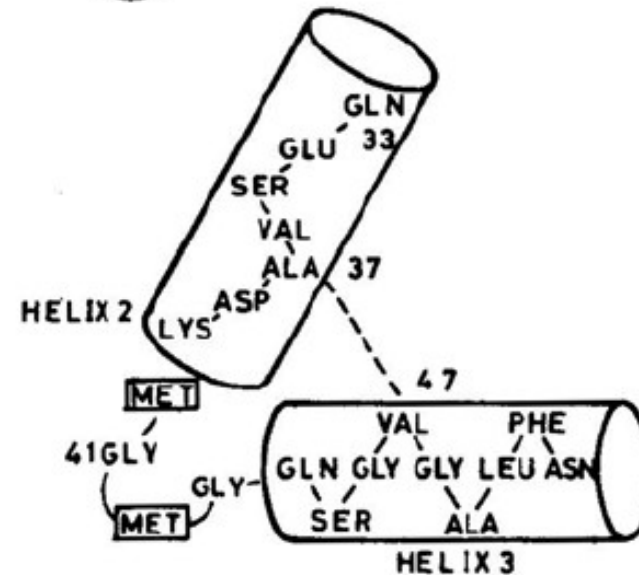
(b)



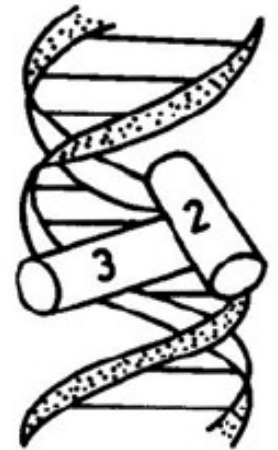
© 1999 GARLAND PUBLISHING INC.  
A member of the Taylor & Francis Group



(A)

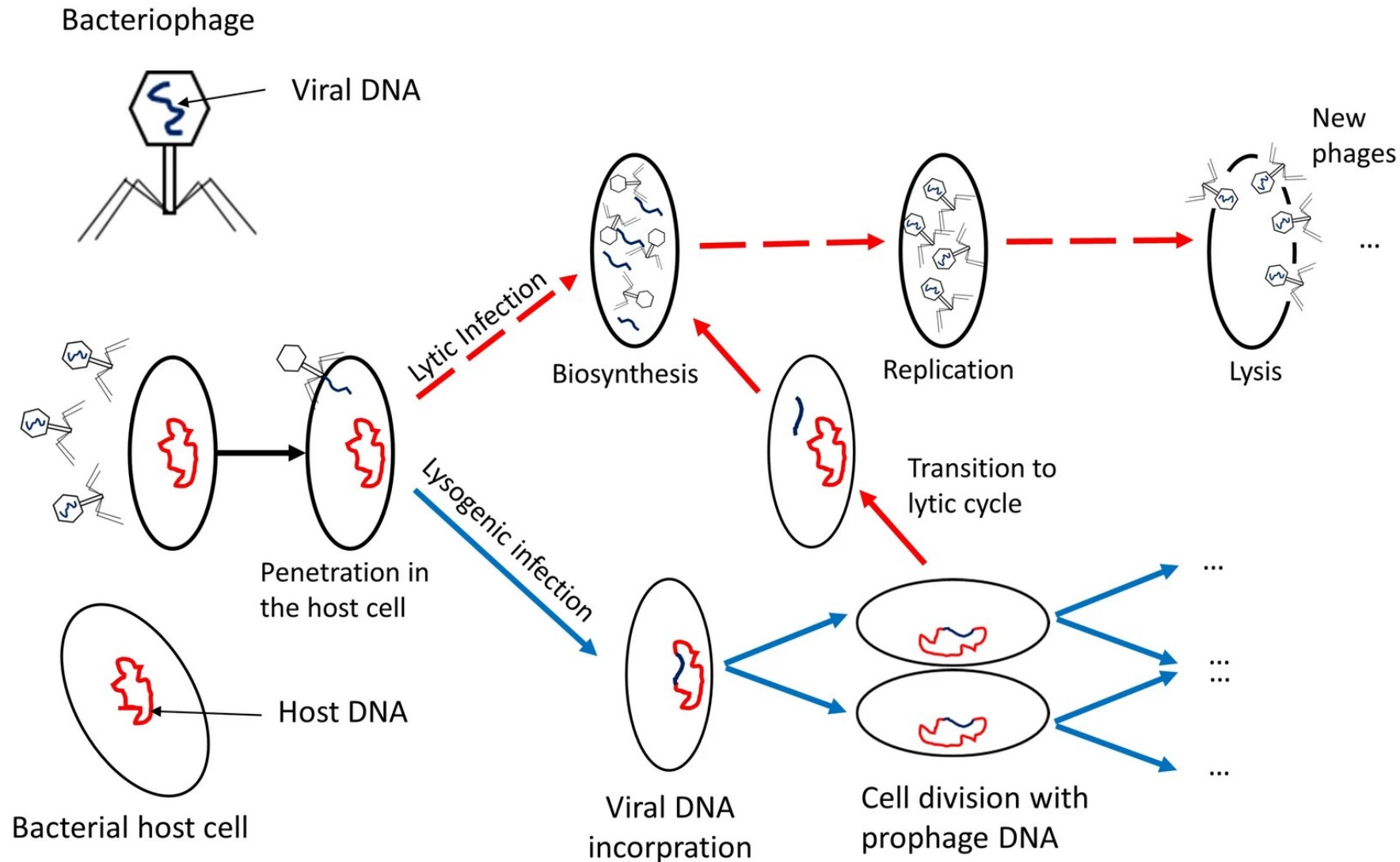
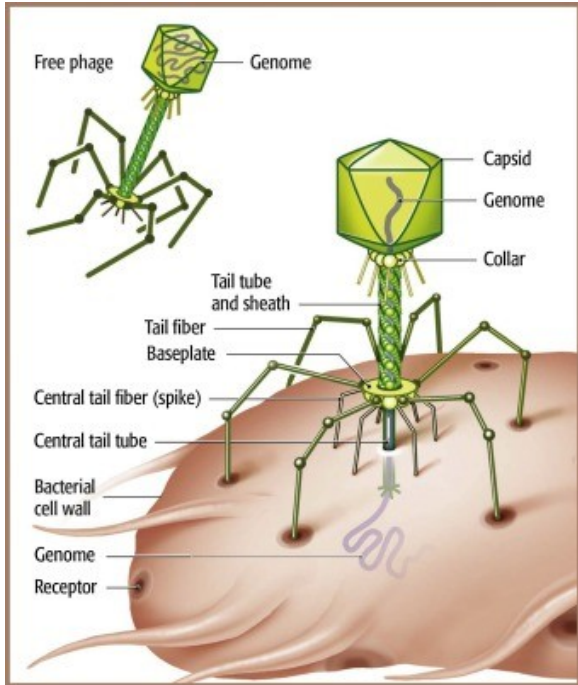


(B)

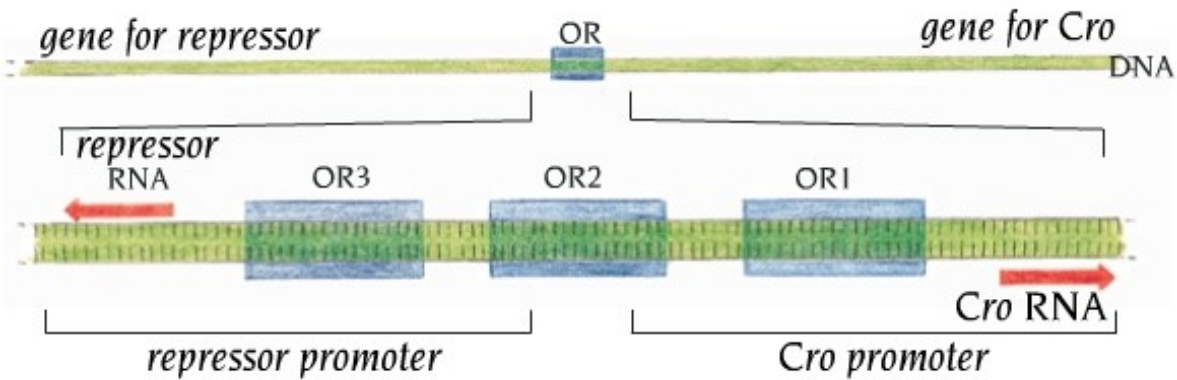


# Functions of Domains

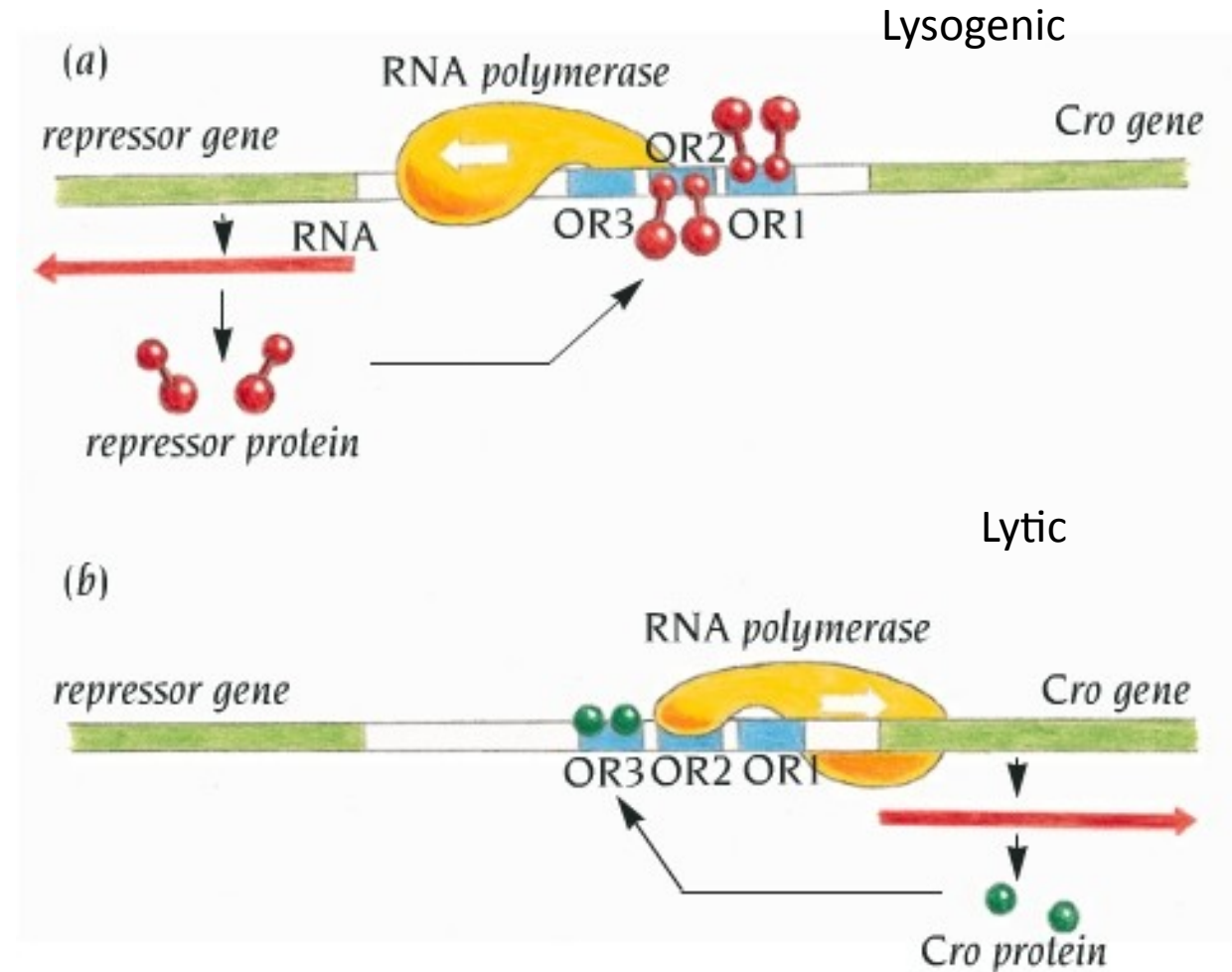
“helix–turn–helix” (HTH)  
motif



# Functions of Domains



© 1999 GARLAND PUBLISHING INC.  
A member of the Taylor & Francis Group



© 1999 GARLAND PUBLISHING INC.  
A member of the Taylor & Francis Group

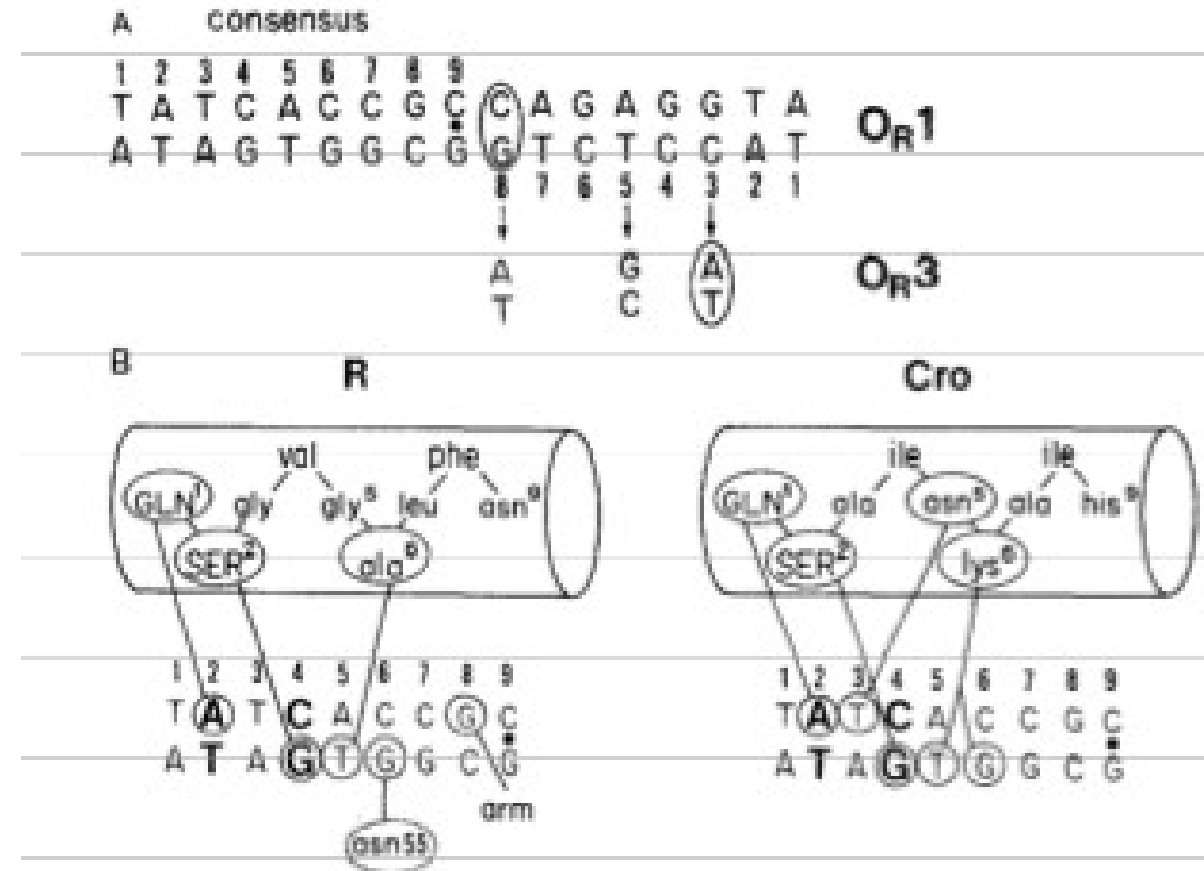
# Functions of Domains

**Table 8.1** The nucleotide sequences of the three protein-binding regions OR1, OR2, and OR3 of the operator of bacteriophage lambda

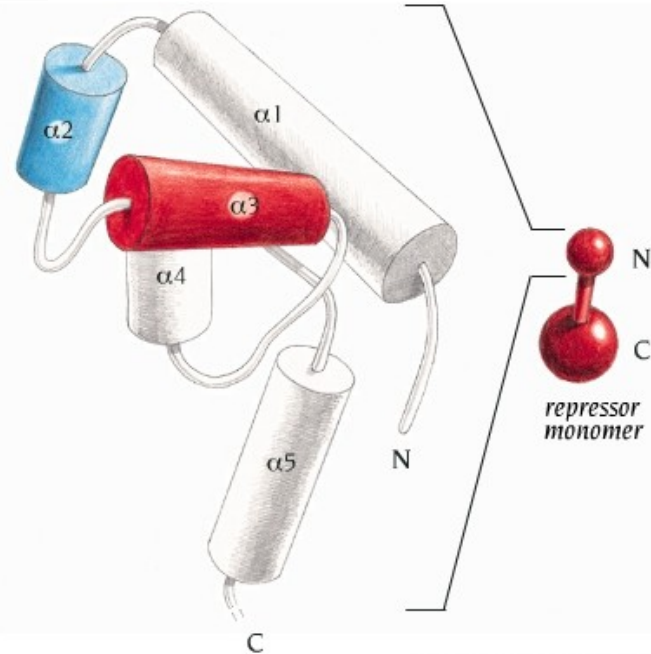
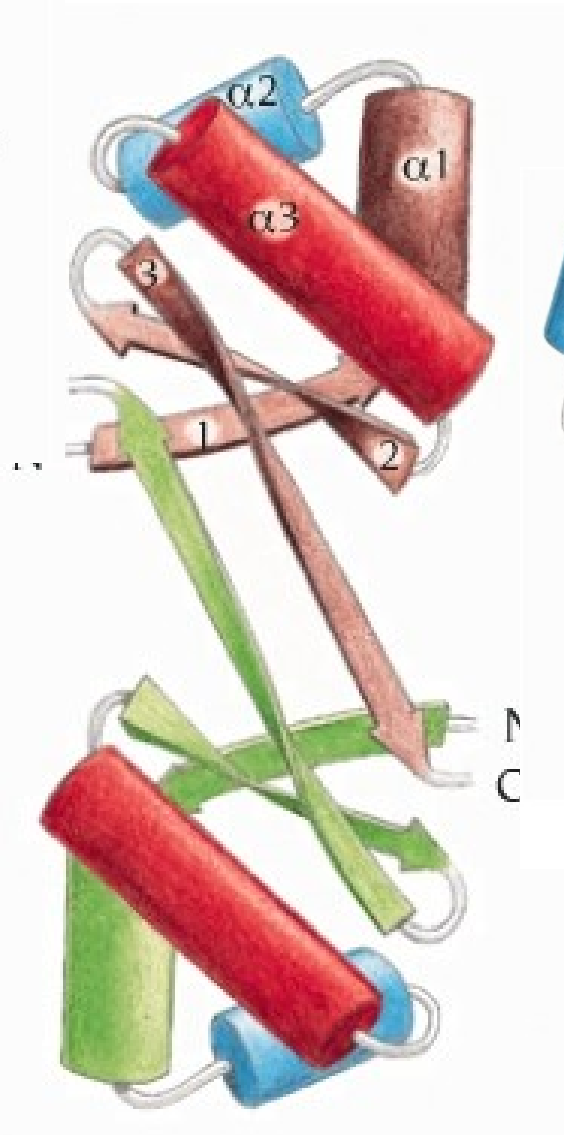
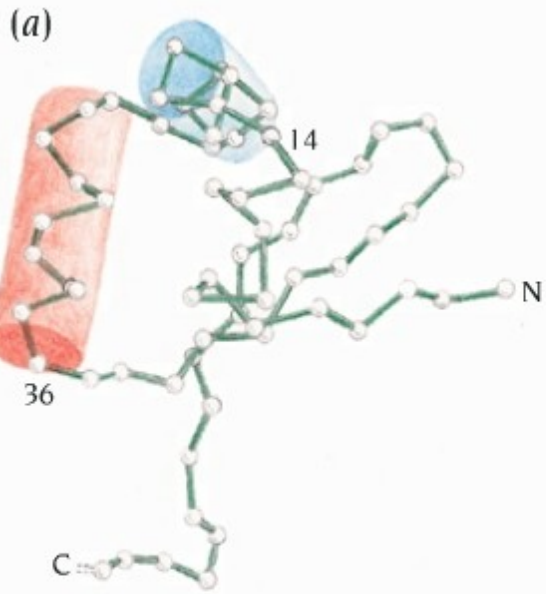
|     |    | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |    |
|-----|----|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|
| OR1 | 5' | T | A | T | C | A | C | C | G | C | C  | A  | G  | T  | G  | G  | T  | A  | 3' |
|     | 3' | A | T | A | G | T | G | G | C | G | G  | T  | C  | A  | C  | C  | A  | T  | 5' |
| OR2 | 5' | T | A | A | C | A | C | C | G | T | G  | C  | G  | T  | G  | T  | T  | G  | 3' |
|     | 3' | A | T | T | G | T | G | G | C | A | C  | G  | C  | A  | C  | A  | A  | C  | 5' |
| OR3 | 5' | T | A | T | C | A | C | C | G | C | A  | A  | G  | G  | G  | A  | T  | A  | 3' |
|     | 3' | A | T | A | G | T | G | G | C | G | T  | T  | C  | C  | C  | T  | A  | T  | 5' |

Palindromic base pairs that are most frequent at the two ends are green, and the pseudo-twofold symmetry axis is indicated by a red dot.

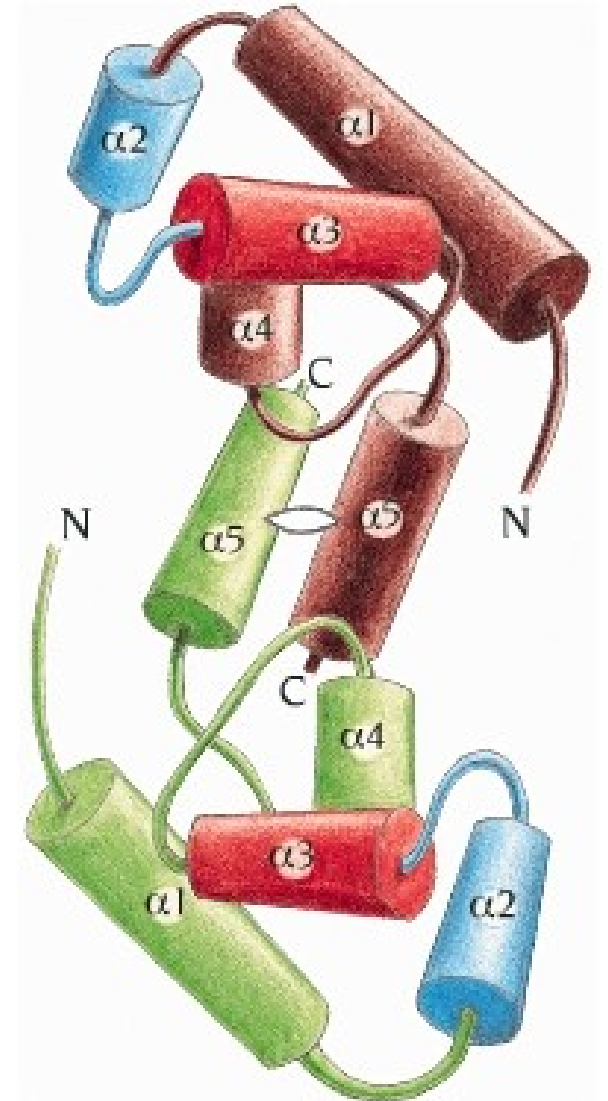
©1999 GARLAND PUBLISHING INC.  
A member of the Taylor & Francis Group



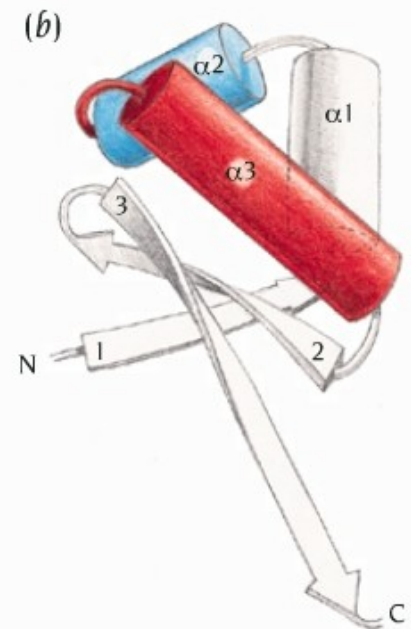
# Cro and repressor proteins



©1999 GARLAND PUBLISHING INC.  
A member of the Taylor & Francis Group



©1999 GARLAND PUBLISHING INC.  
A member of the Taylor & Francis Group

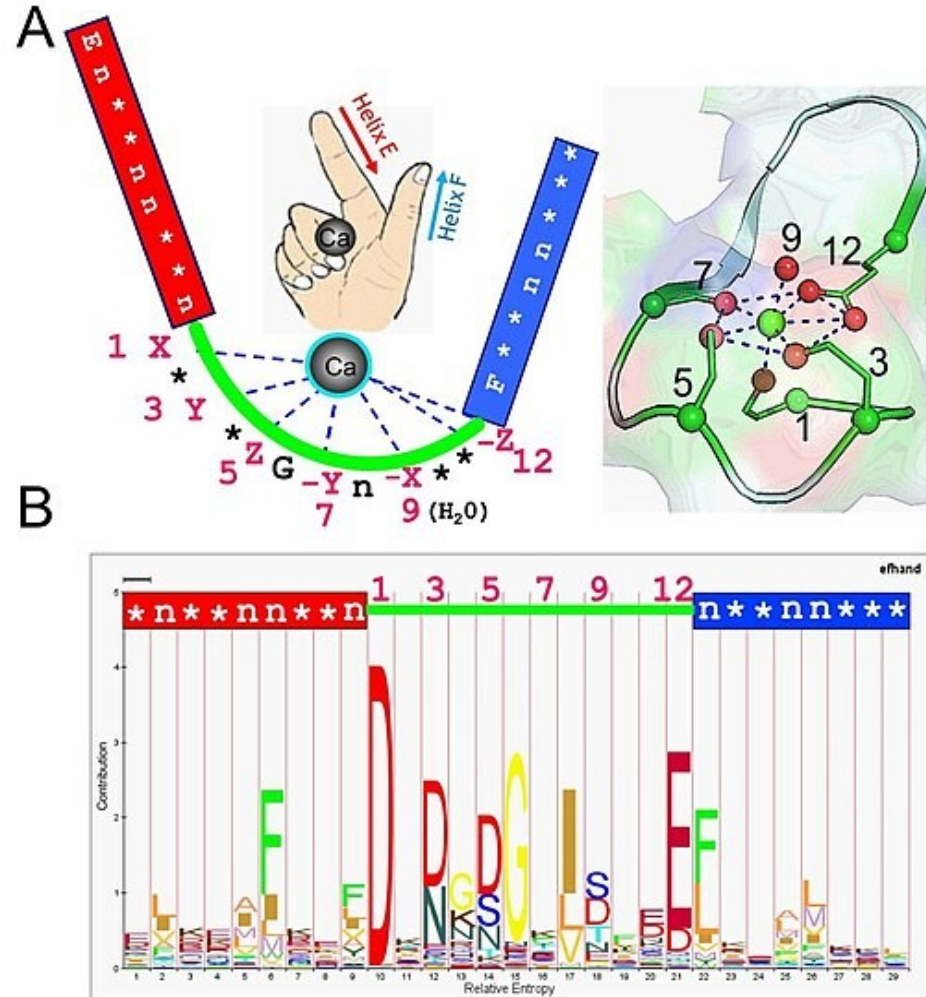


©1999 GARLAND PUBLISHING INC.  
A member of the Taylor & Francis Group

©1999 GARLAND PUBLISHING INC.  
A member of the Taylor & Francis Group

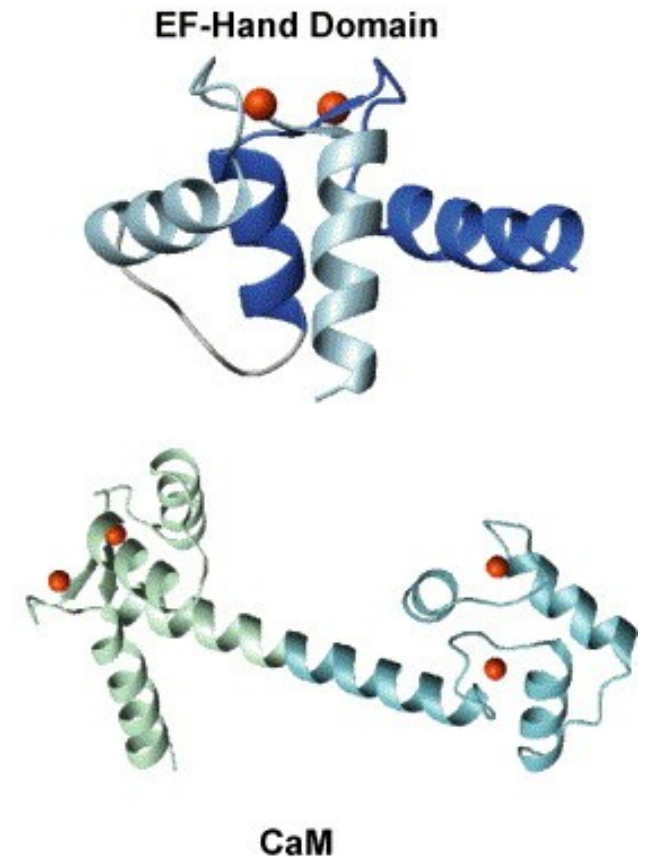
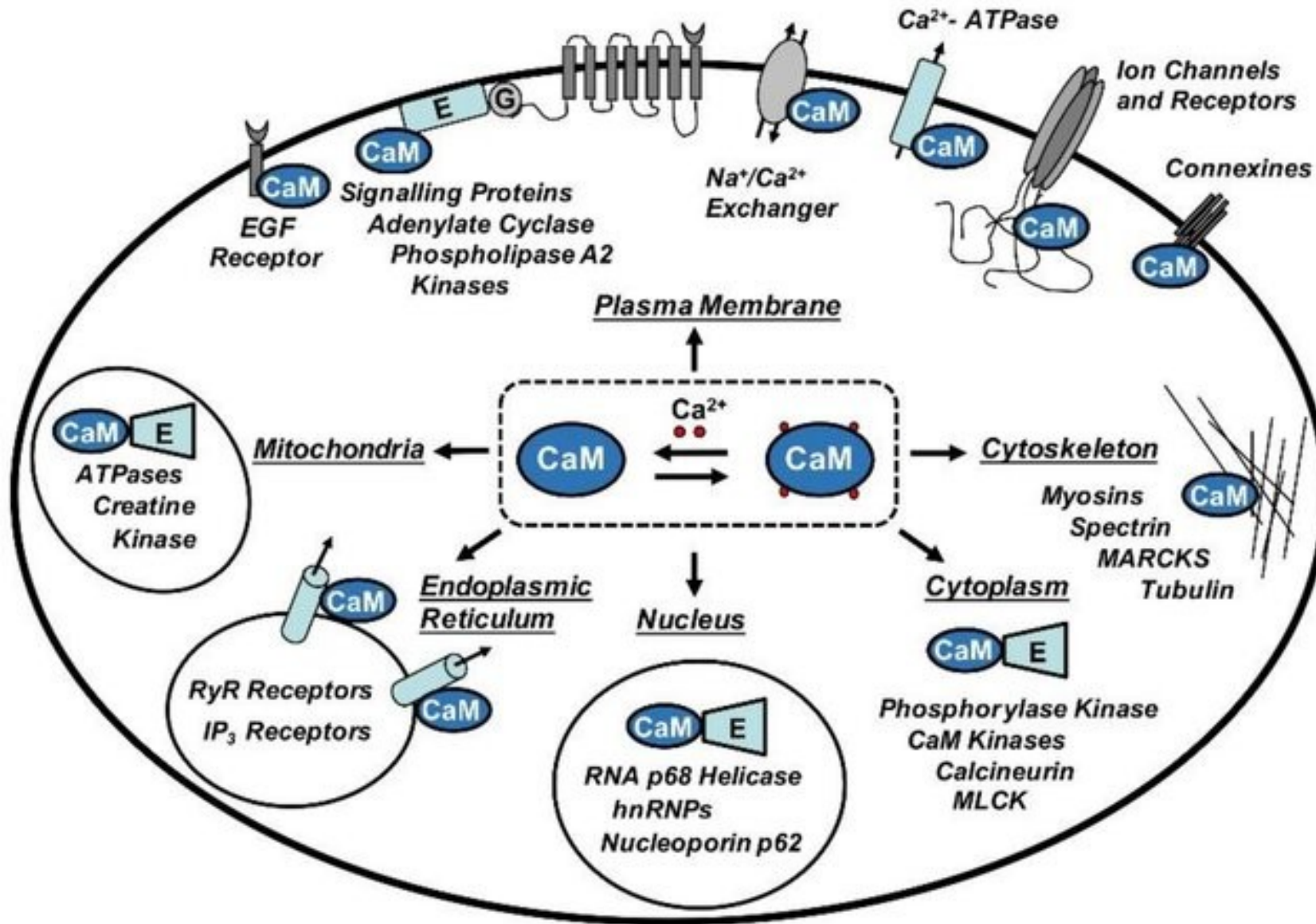
# Functions of Domains

## EF hand

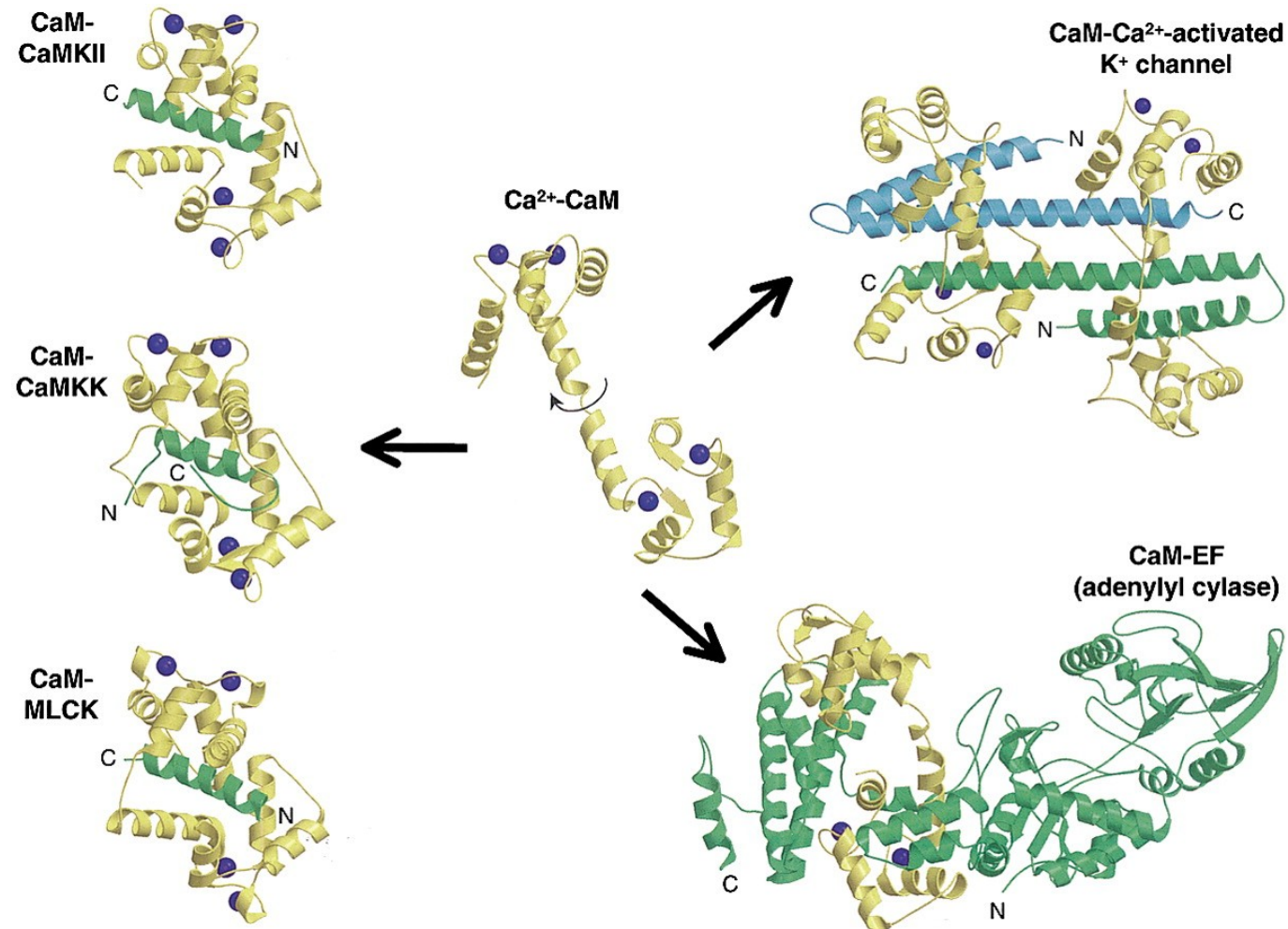


D Ala E Glu (-) K Lys R Arg (+) F Phe L Leu

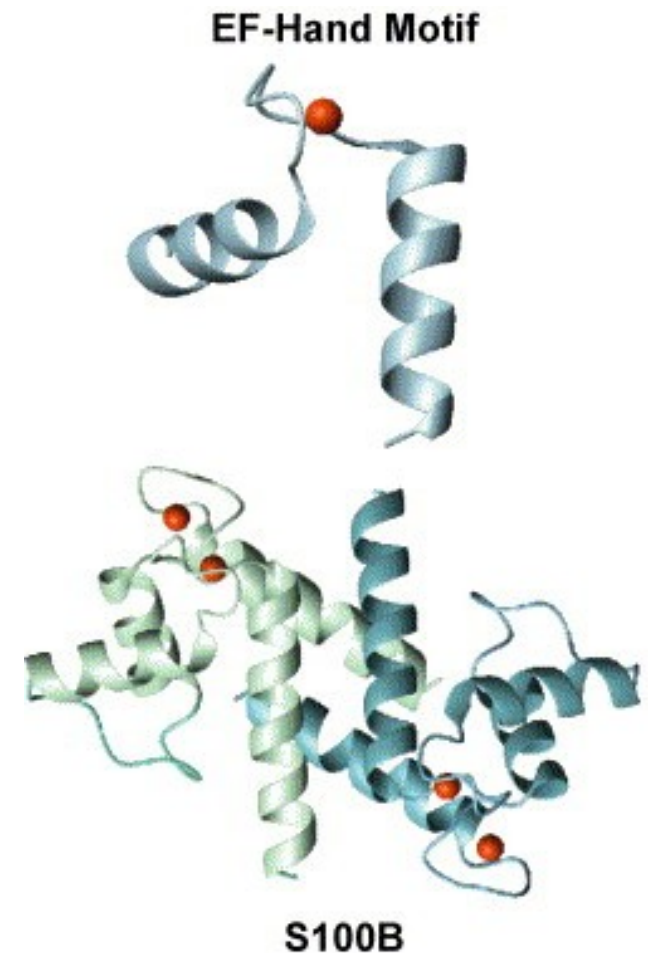
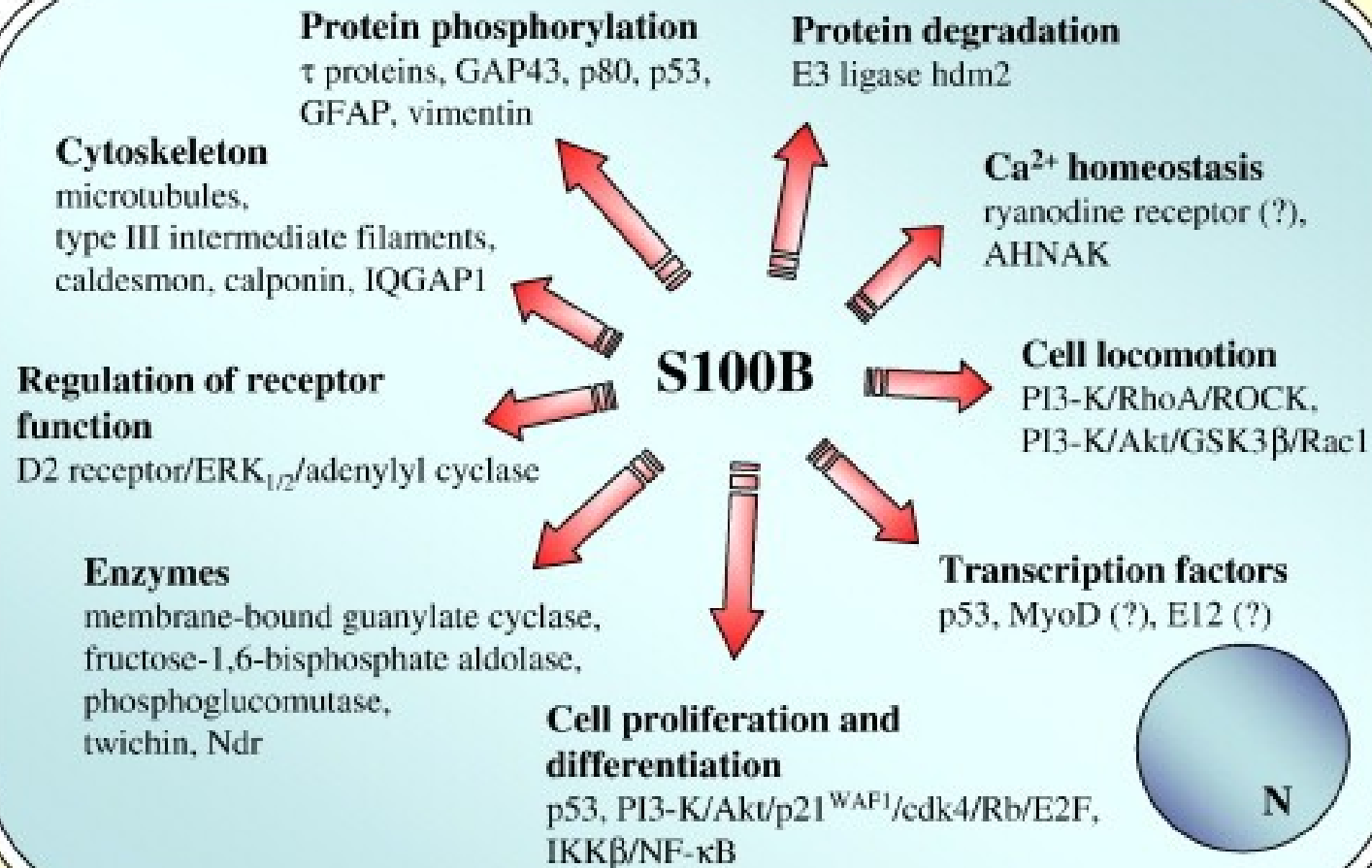
# The central role played by calmodulin (CaM)



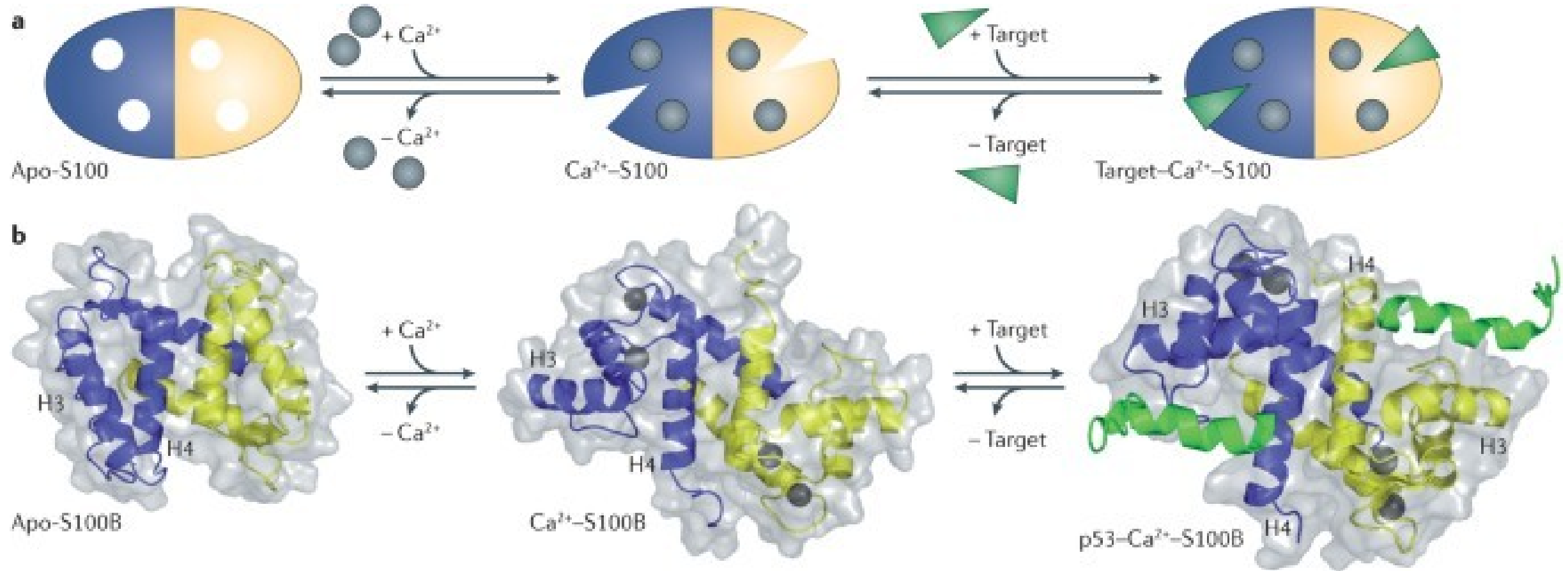
# Action of calmodulin



# S100B protein



# S100B protein action



# Adenylate kinase

three domains:  
a central domain flanked by two smaller  
nucleotide-binding regions

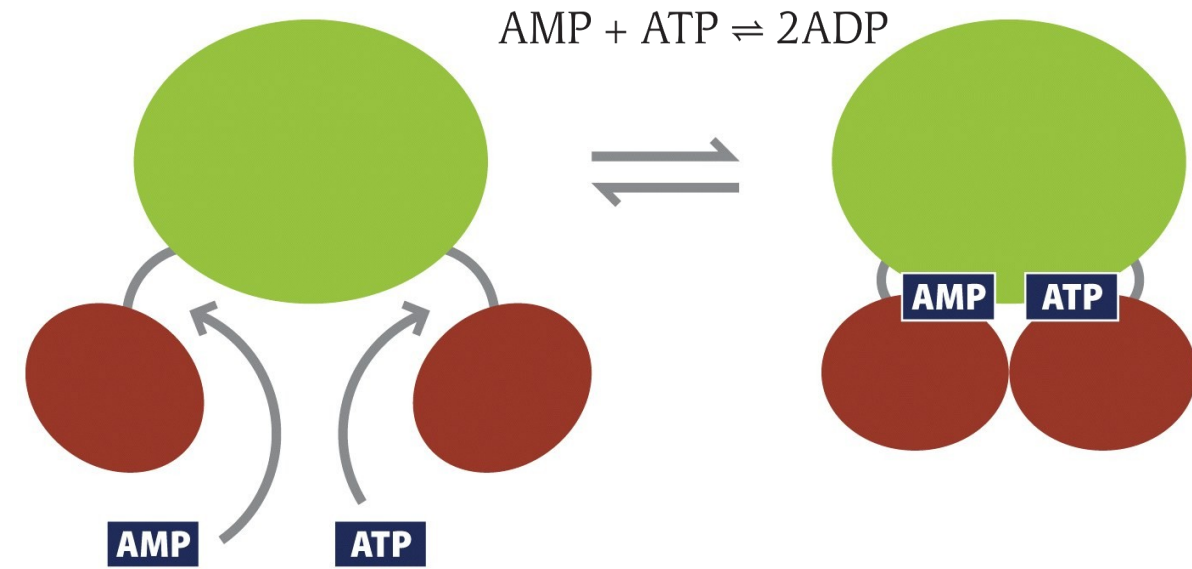
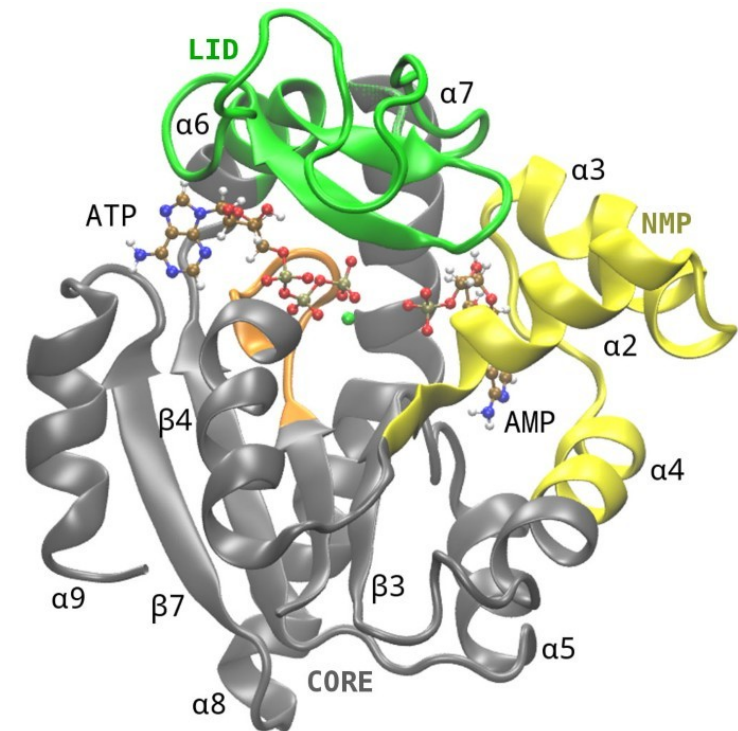
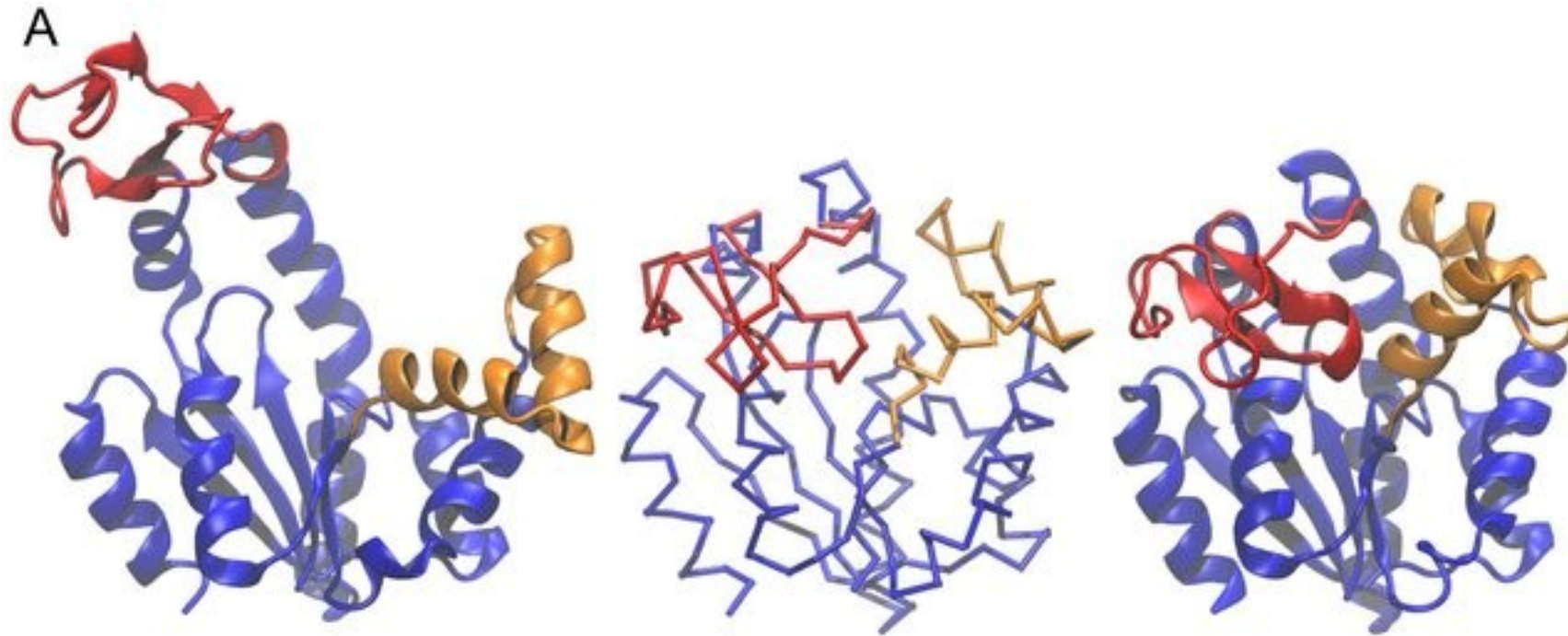
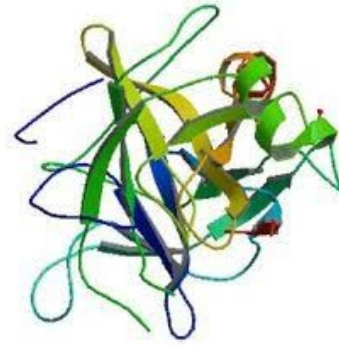
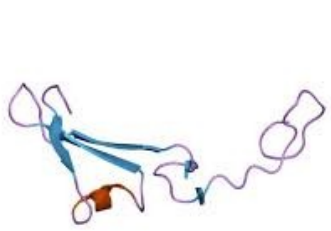


Figure 2.8 How Proteins Work (©2012 Garland Science)



# Modules are transposable domains



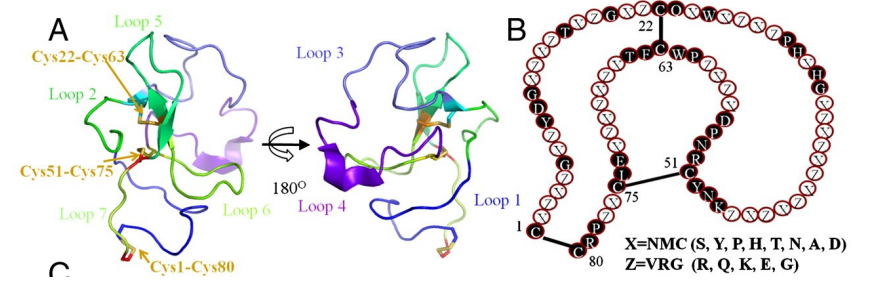
EGF

CHYMOTRYPSIN

UROKINASE

FACTOR IX

PLASMINOGEN



Domains that are homologous to the epidermal growth factor, EGF, which is a small polypeptide chain of 53 amino acids.

Serine proteinase domains that are homologous to chymotrypsin, which has about 245 amino acids arranged in two domains.

Kringle domains, which have a characteristic pattern of three internal disulfide bridges within a region of about 85 amino acid residues.

Calcium-binding domain (see Figure 2.13).

# Tissue plasminogen activator



# Multidomain proteins are produced by exon shuffling

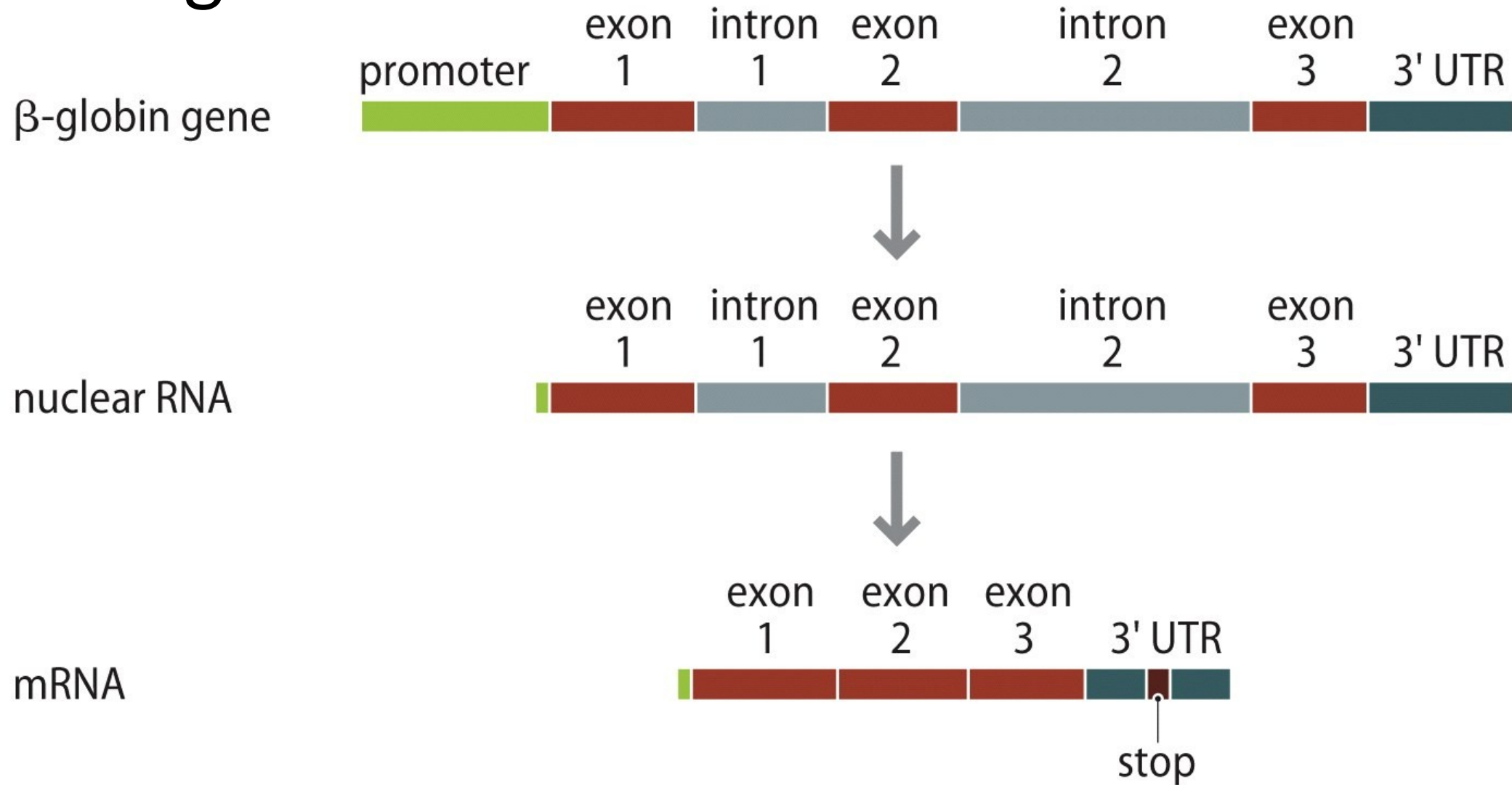
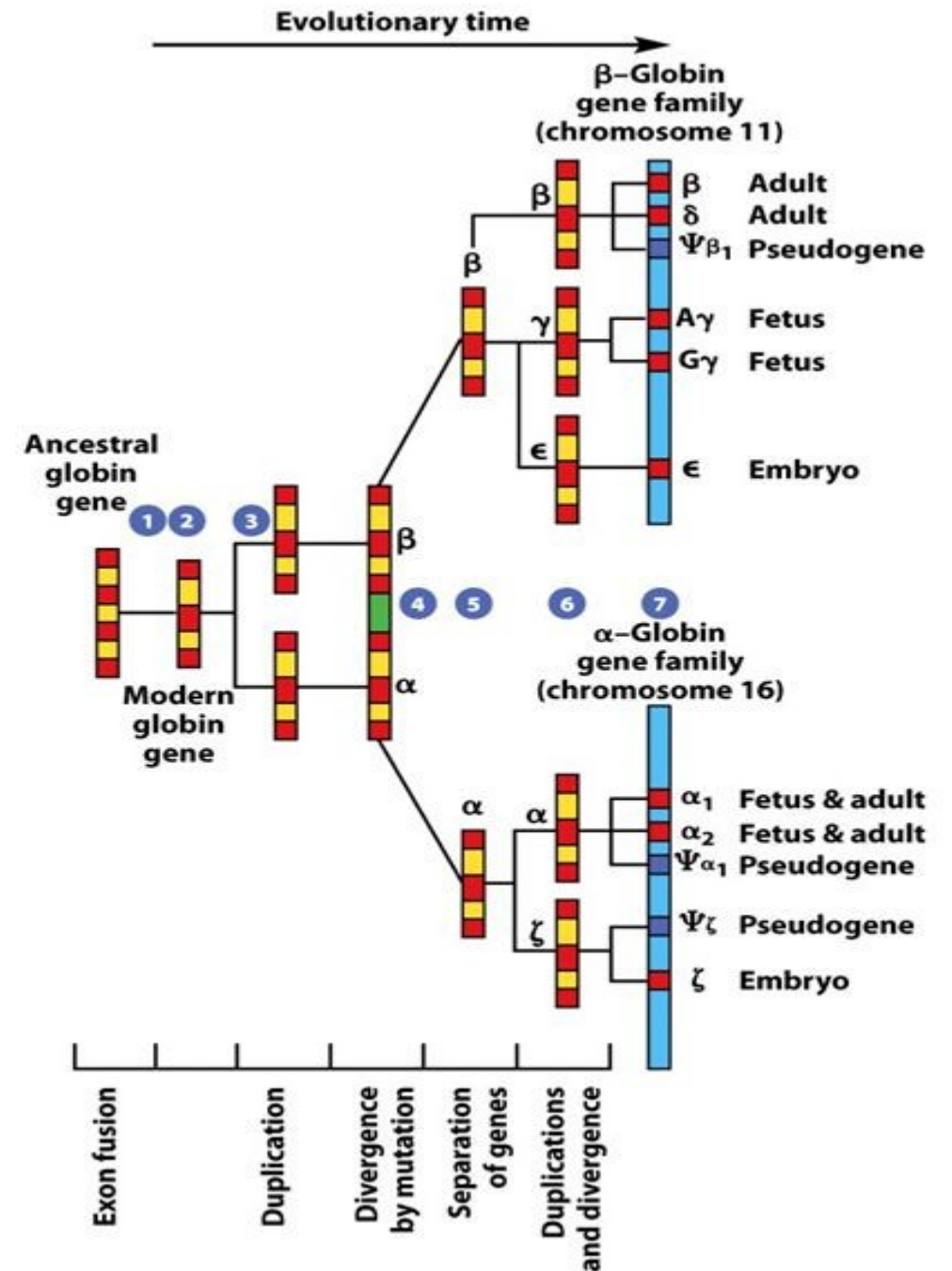


Figure 2.14 How Proteins Work (©2012 Garland Science)

# A pathway for the evolution of globin genes



# Hypothetical evolution

S, signal peptide  
P, trypsin-like protease  
G, EGF  
K, kringle;  
F, finger;  
FN2, fibronectin type 2.

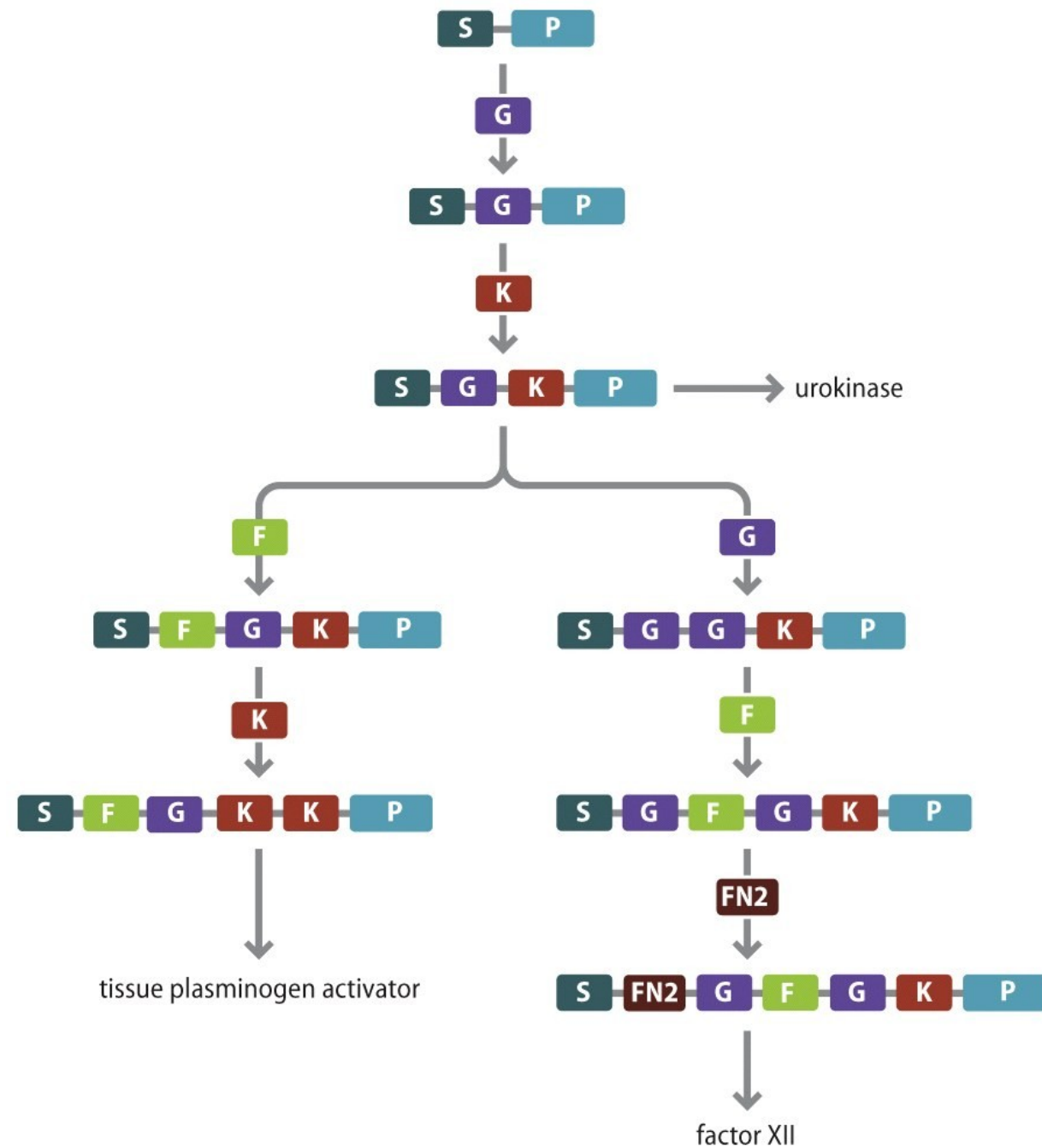


Figure 2.15 How Proteins Work (©2012 Garland Science)

# Complexity

*Mycoplasma genitalium*, a parasite that lives inside the genitourinary tract.

521 genes

482 are coding regions

Only 382 of these are essential (determined by single-gene knockouts).

Bacterium has about 3000 genes.

A single-celled eukaryote such as yeast has 5000 genes

Humans have approximately 21,000 genes

# Cell surface receptors

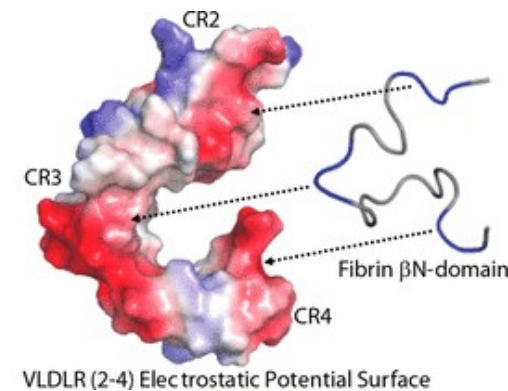
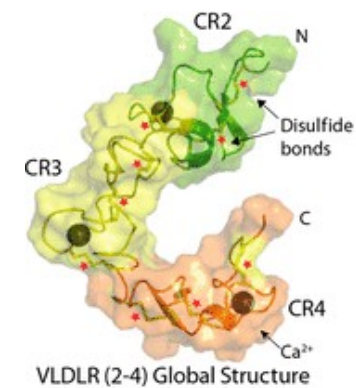
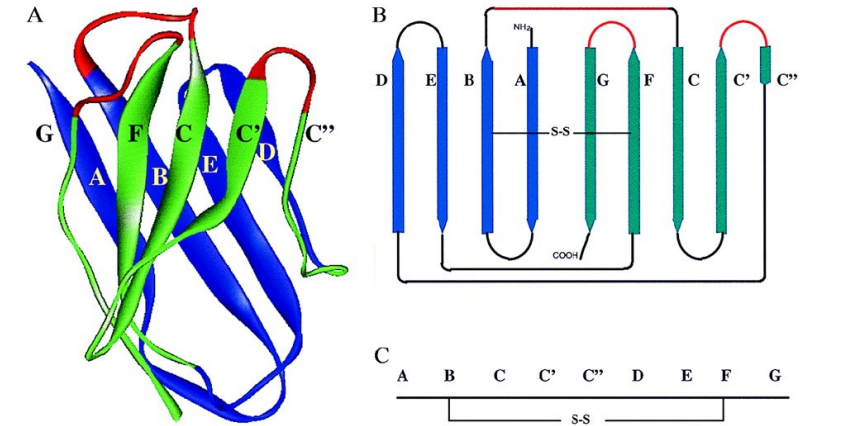
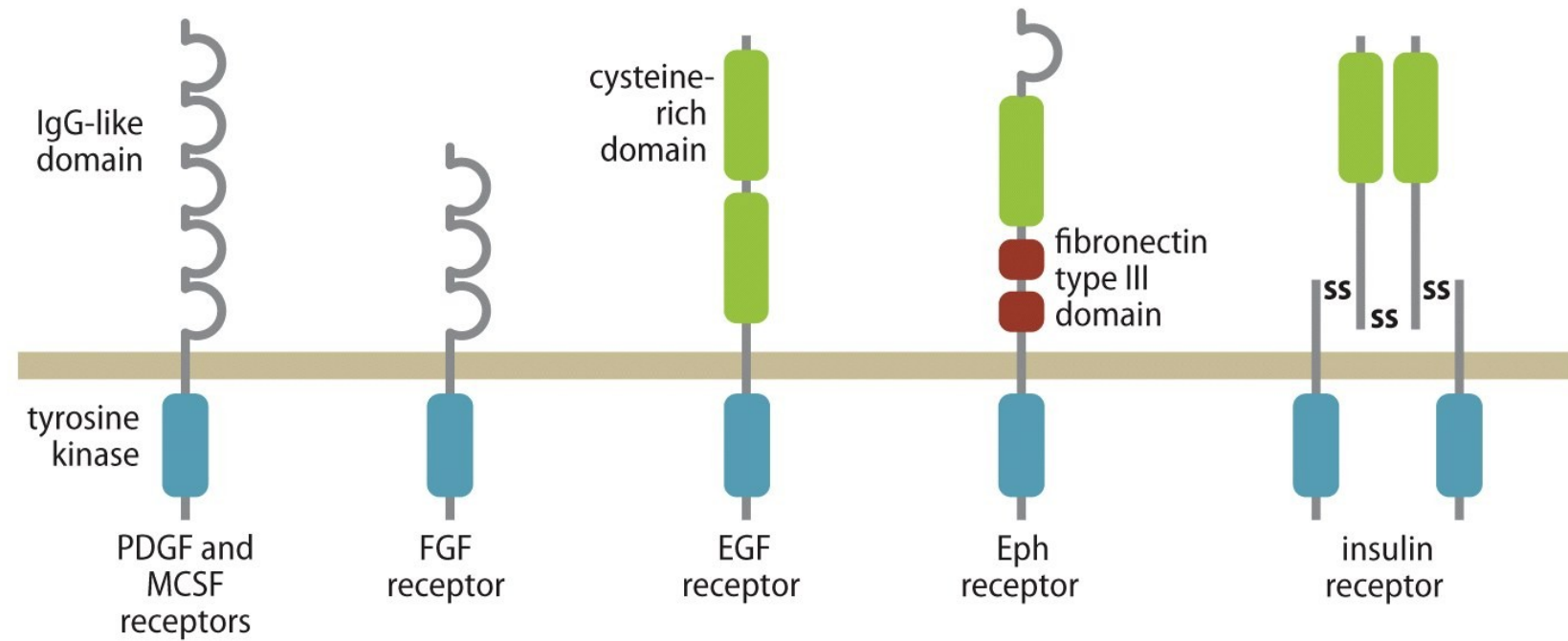
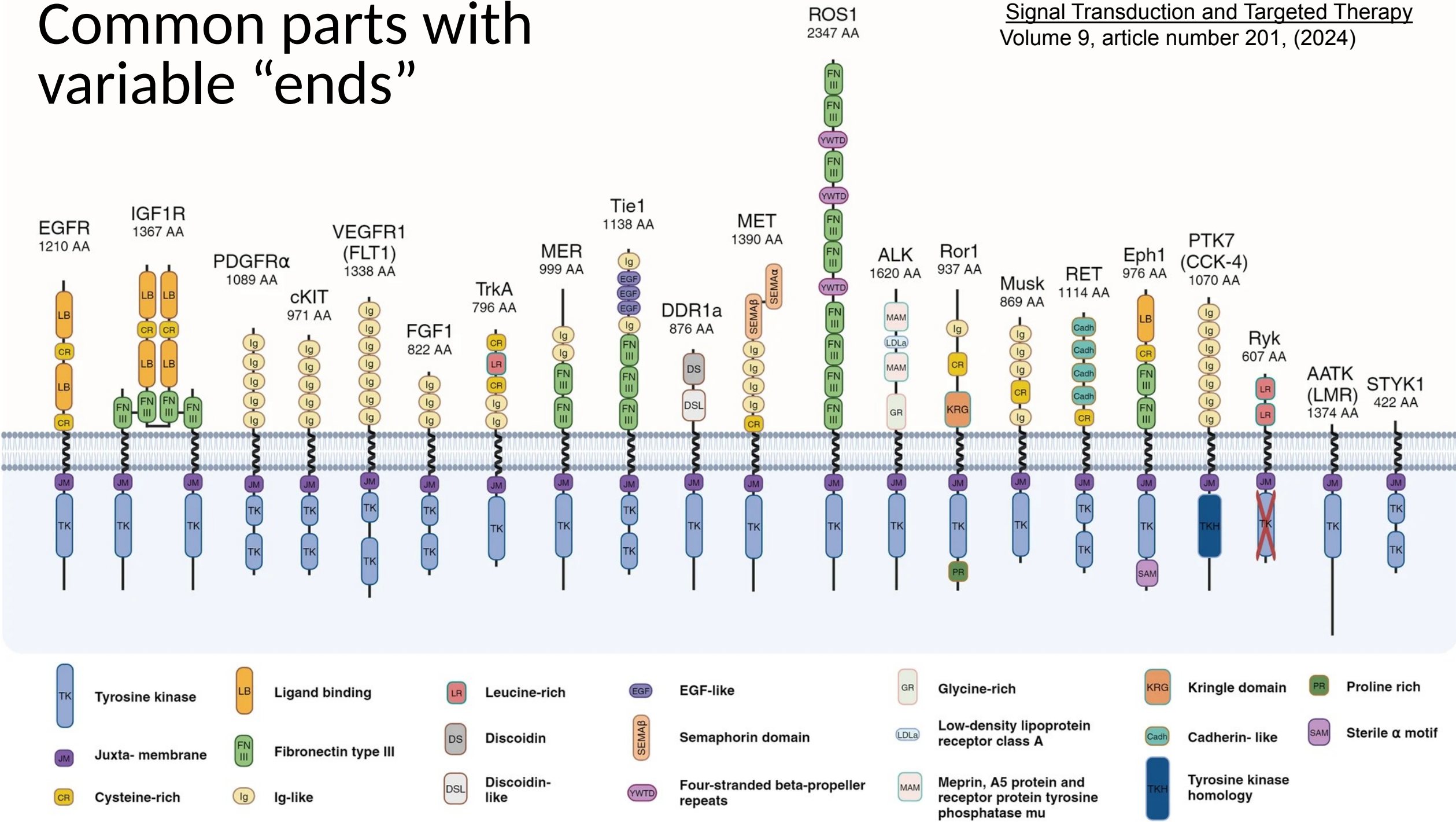
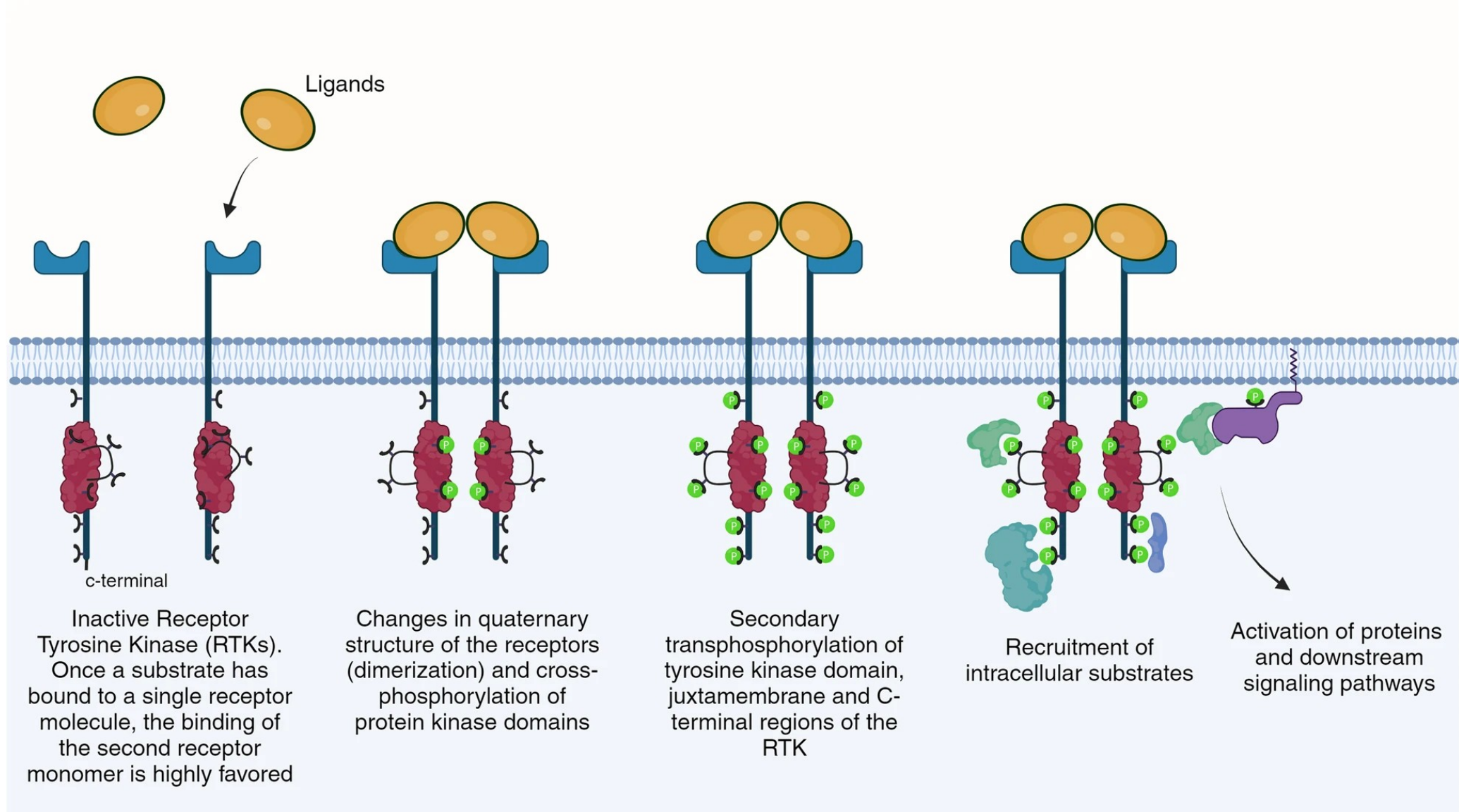


Figure 2.11 How Proteins Work (©2012 Garland Science)

Signal Transduction and Targeted Therapy  
Volume 9, article number 201, (2024)



# Activation and Intracellular Signaling Mechanisms of Receptor Tyrosine Kinase



# Adaptor protein

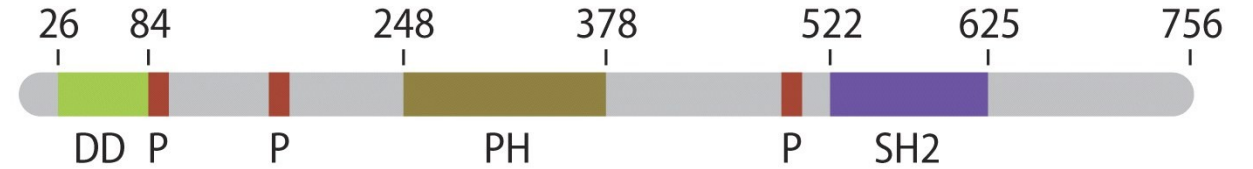
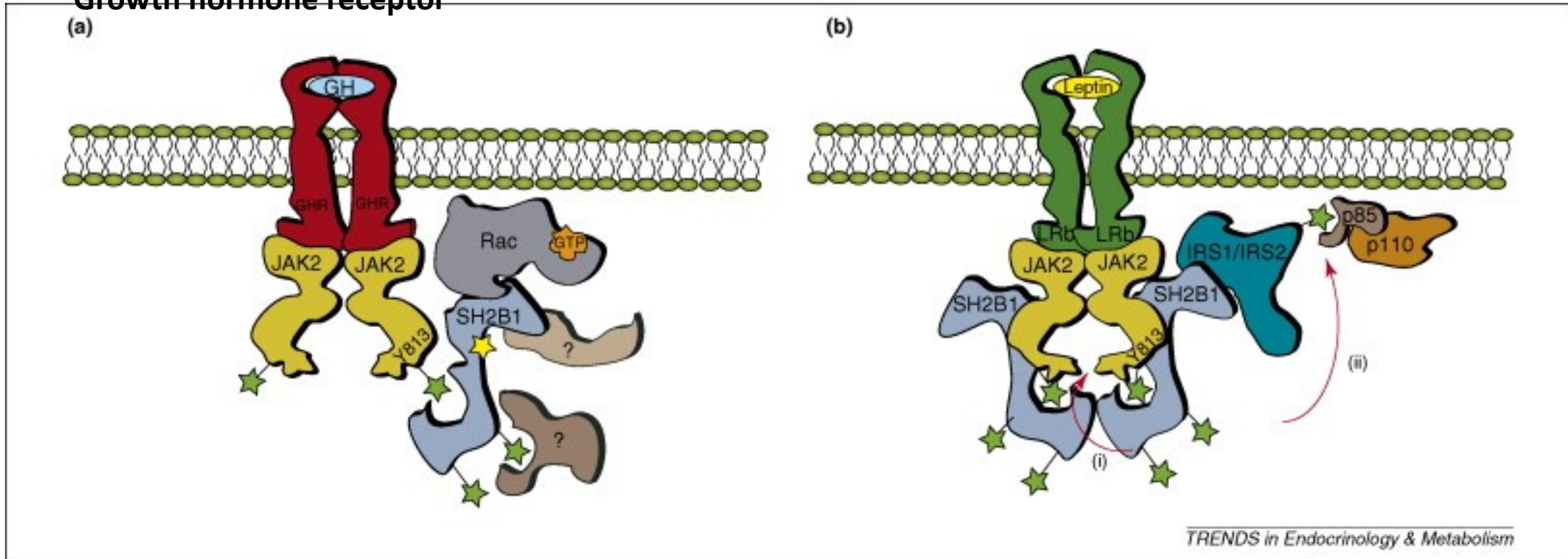


Figure 2.19 How Proteins Work (©2012 Garland Science)

DD, dimerization domain; P, proline-rich sequence;  
PH, pleckstrin homology domain; SH2, Src homology domain 2

## Growth hormone receptor



# Increase binding specificity

SH2 domain

positively red  
negatively blue.  
phosphotyrosine (phosphorus in orange)

| TABLE 2.2 Consensus recognition sequences for SH2 domains |          |       |       |       |
|-----------------------------------------------------------|----------|-------|-------|-------|
| Domain                                                    | Sequence |       |       |       |
| Abl                                                       | pY       | E/T/M | N/E/D | P/V/L |
| Crk                                                       | pY       | D/K/N | H/F/R | P/V/L |
| Fes                                                       | pY       | E     | X     | V/I   |
| Fgr                                                       | pY       | E/Y/D | E/N/D | I/V   |
| Fyn                                                       | pY       | E/T   | E/D/Q | I/V/M |
| Grb2                                                      | pY       | I/V   | N     | I/L/V |
| Lck                                                       | pY       | E/T/Q | E/D   | I/V/M |
| Nck                                                       | pY       | D     | E     | P/D/V |
| Rasa-N                                                    | pY       | I/L/V | X     | Φ     |
| Rasa-C                                                    | pY       | X     | X     | P     |
| SH3BP2                                                    | pY       | E/M/V | N/V/I | X     |
| SHB                                                       | pY       | T/V/I | X     | L     |
| Src                                                       | pY       | E/D/T | E/N/Y | I/M/L |
| STAT1                                                     | pY       | D/E   | P/R   | R/P/Q |
| STAT3                                                     | pY       | X     | X     | Q     |
| Syk-C                                                     | pY       | Q/T/E | E/Q   | L/I   |
| Tns                                                       | pY       | E     | N     | F/I/V |
| Vav1                                                      | pY       | M/L/E | E     | P     |

X represents any amino acid and Φ any hydrophobic amino acid. (Data from B.A. Liu, K. Jablonowski, M. Raina, M. Arcé, T. Pawson and P.D. Nash, *Mol. Cell* 22: 851-868, 2006. With permission from Elsevier.)

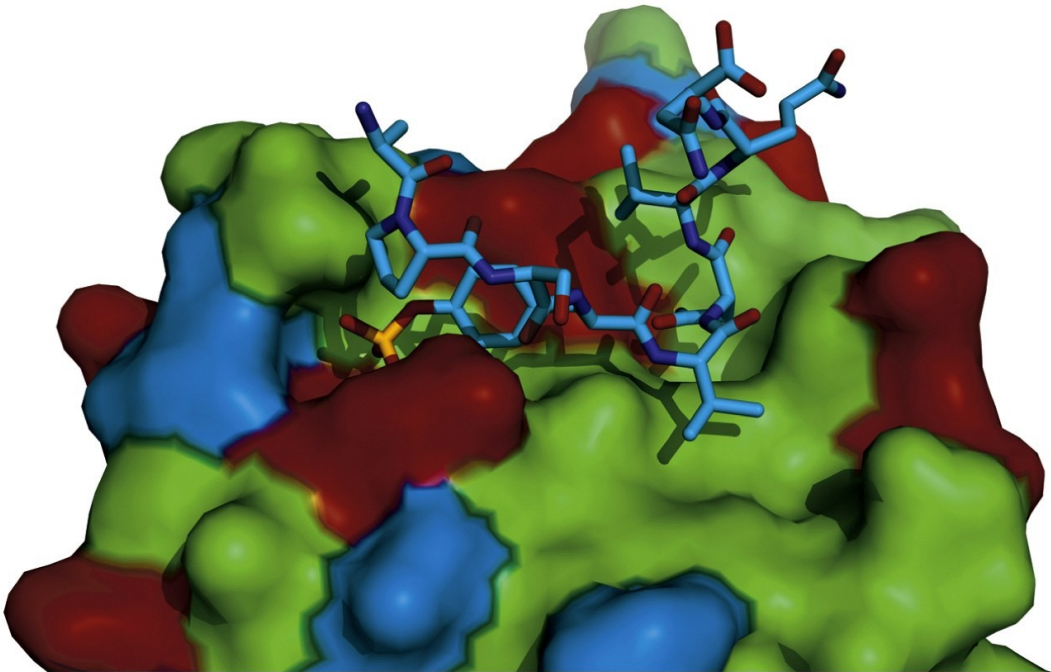
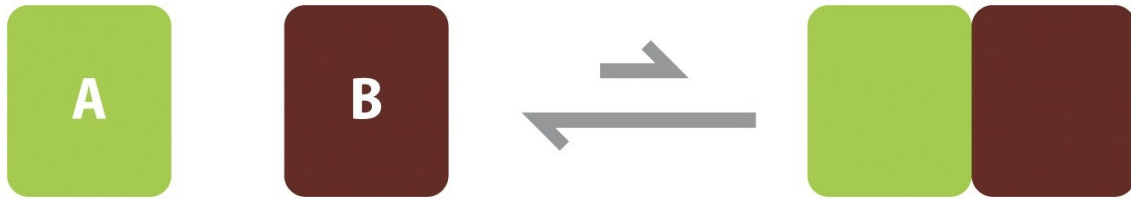
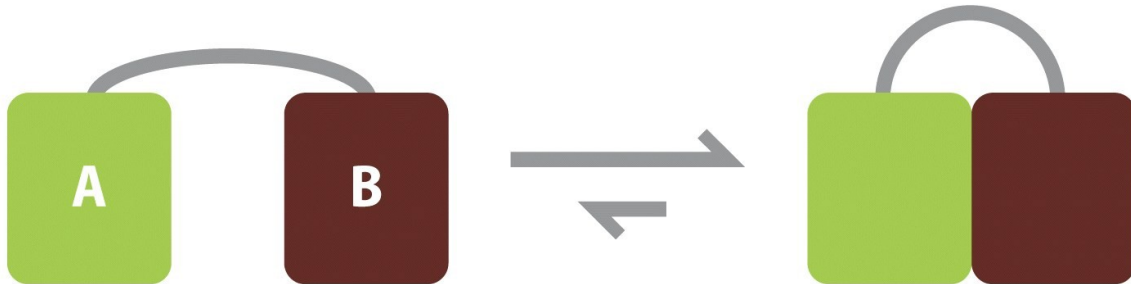


Figure 2.18 How Proteins Work (©2012 Garland Science)

# Interaction of two domains



Η αποτελεσματική συγκέντρωση  $[A/B]$  είναι η συγκέντρωση του μη συνδεδεμένου (ελεύθερου) A που θα χρειαζόταν για να επιτευχθεί η ίδια ποσότητα σύνδεσης με το B όπως στην περίπτωση της ενδομοριακής σύνδεσης A – B.



Η  $[A/B]$  μπορεί να φτάσει τα  $10^5$  M  
υπάρχει τεράστια αύξηση της συγγένειας σύνδεσης δύο μορίων εάν συνδεθούν μεταξύ τους με τον κατάλληλο τρόπο.

# Intramolecular binding

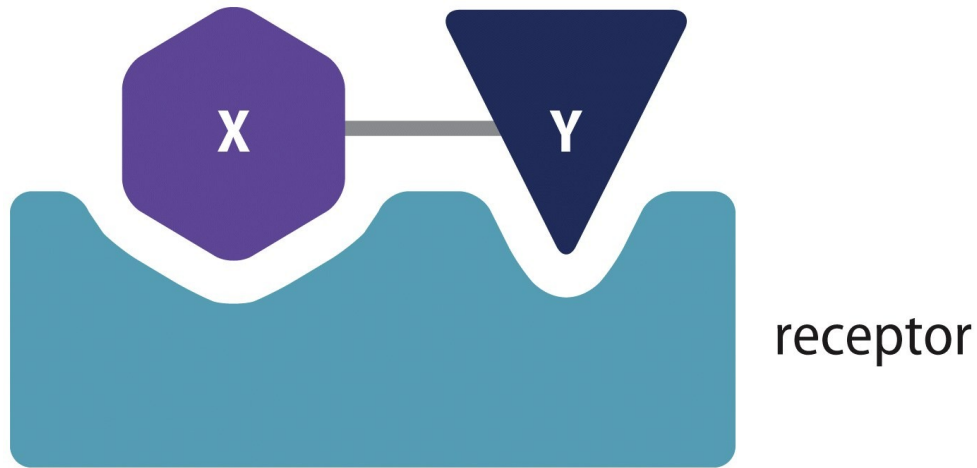


Figure 2.28 How Proteins Work (©2012 Garland Science)

η σύνδεση του μεγαλύτερου προσδέτη X-Y είναι πολύ πιο ευνοϊκή από το άθροισμα των μερών του, κυρίως επειδή ο δυσμενής όρος  $\Delta G_{t+r}$  είναι περίπου ο ίδιος ανεξάρτητα από το μέγεθος του προσδέτη, ενώ οι ευνοϊκές αλληλεπιδράσεις σύνδεσης συνεχίζουν να αυξάνονται καθώς ο προσδέτης γίνεται μεγαλύτερος.

$$\Delta G = \sum \Delta G_i + \Delta G_{t+r} + \Delta G_{\text{conf}}$$

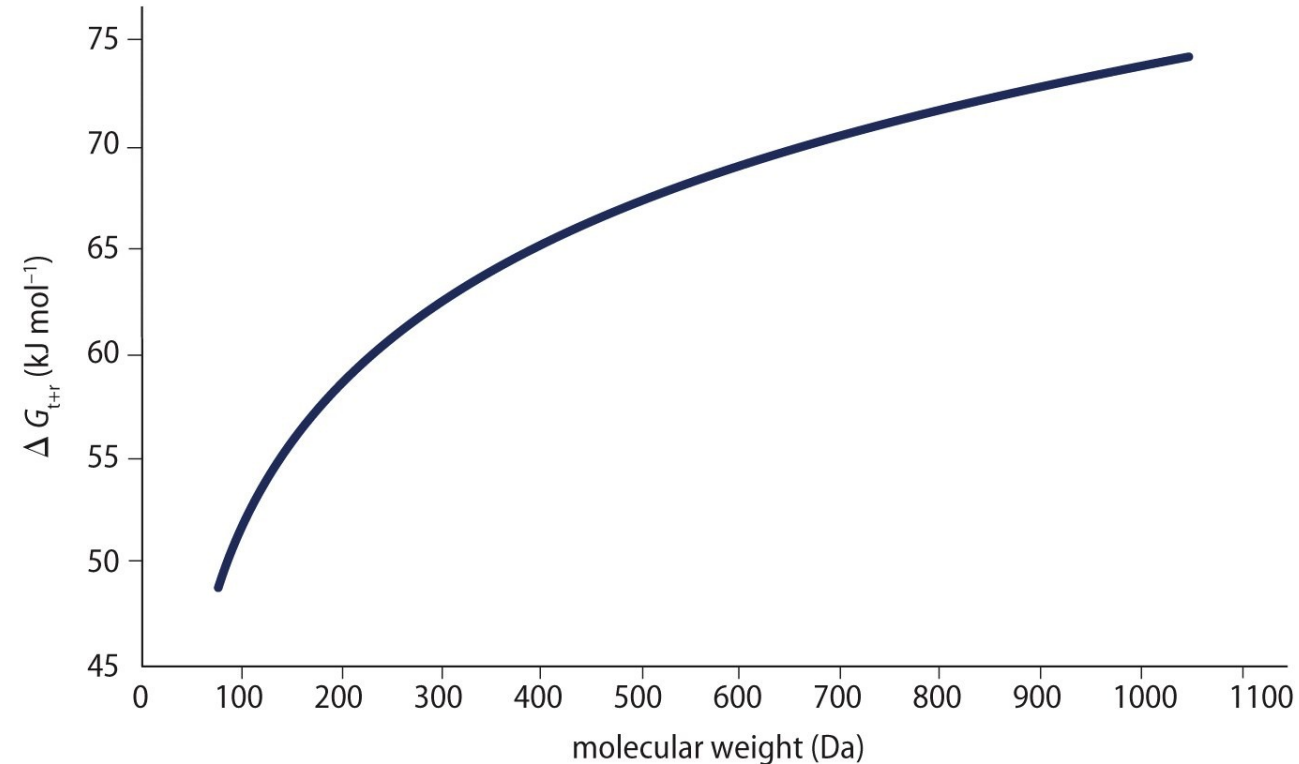


Figure 2.29 How Proteins Work (©2012 Garland Science)

# Intramolecular domain

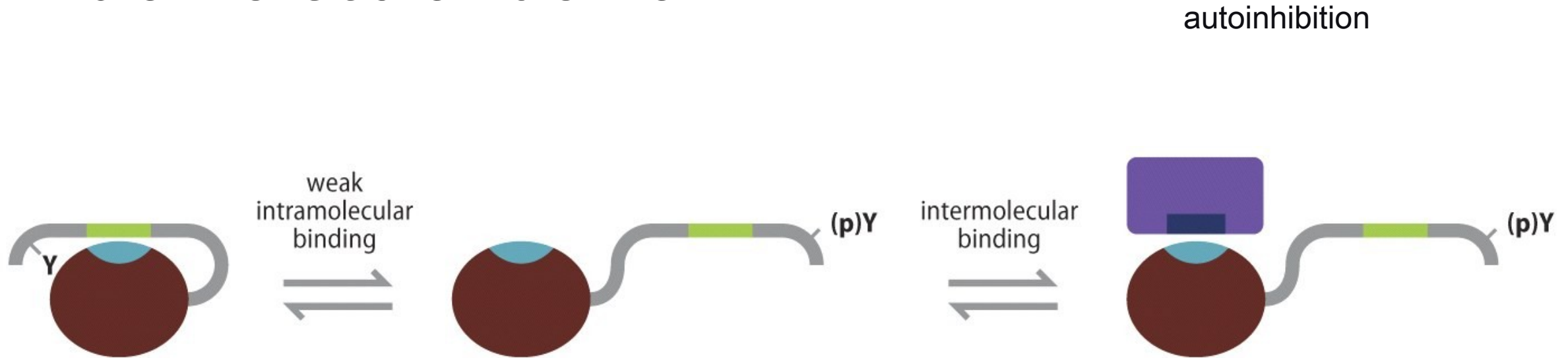
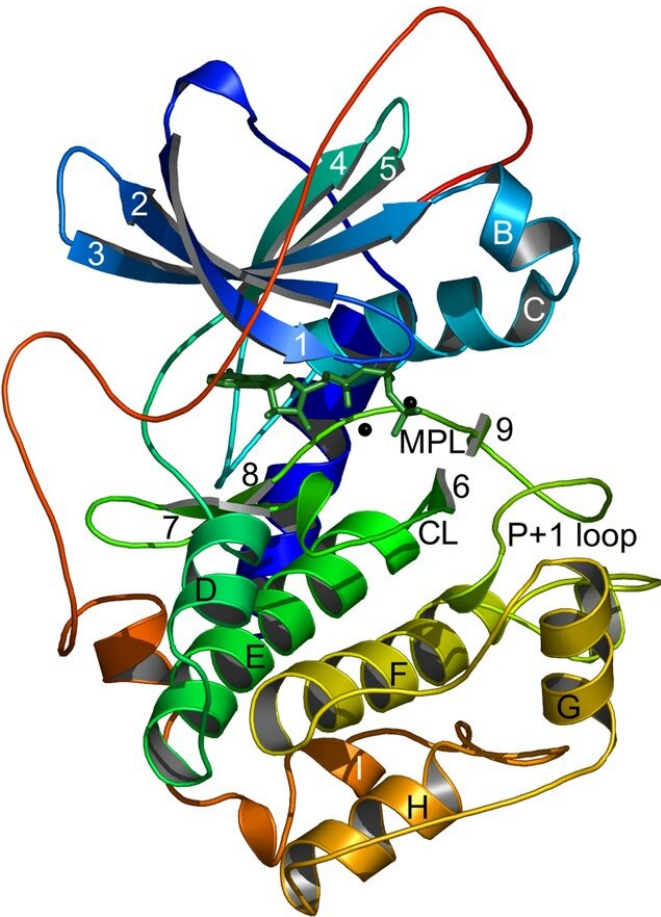


Figure 2.22 How Proteins Work (©2012 Garland Science)

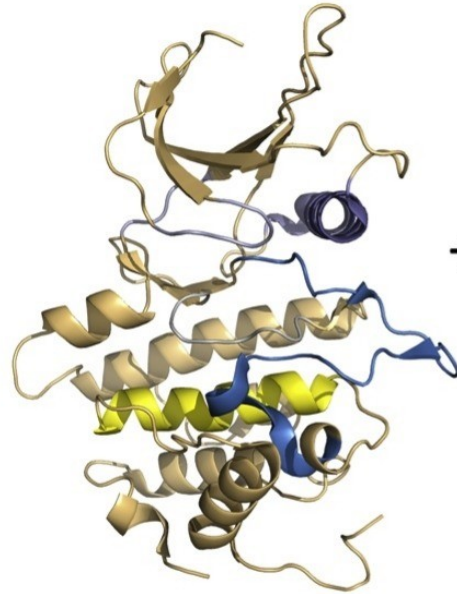
the regulation of the calcium-binding protein calmodulin  
the inhibition of a tandem PDZ domain in X11/Mnt  
the down-regulation of talin, which links the transmembrane integrin cell adhesion molecules to the actin cytoskeleton Autoinhibition is particularly common in signaling

# protein kinase A (PKA)

protein kinase A (PKA)

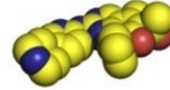


Activated protein kinase

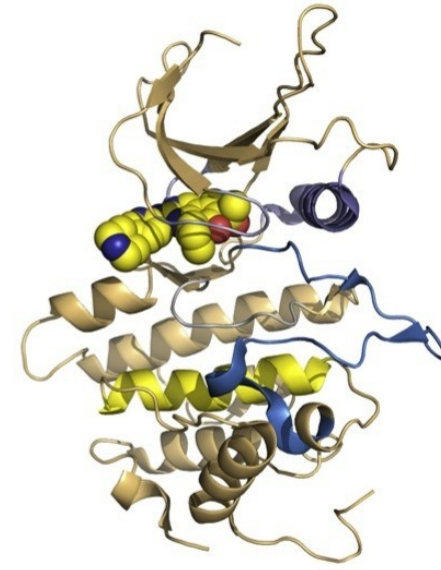


Pathology

+

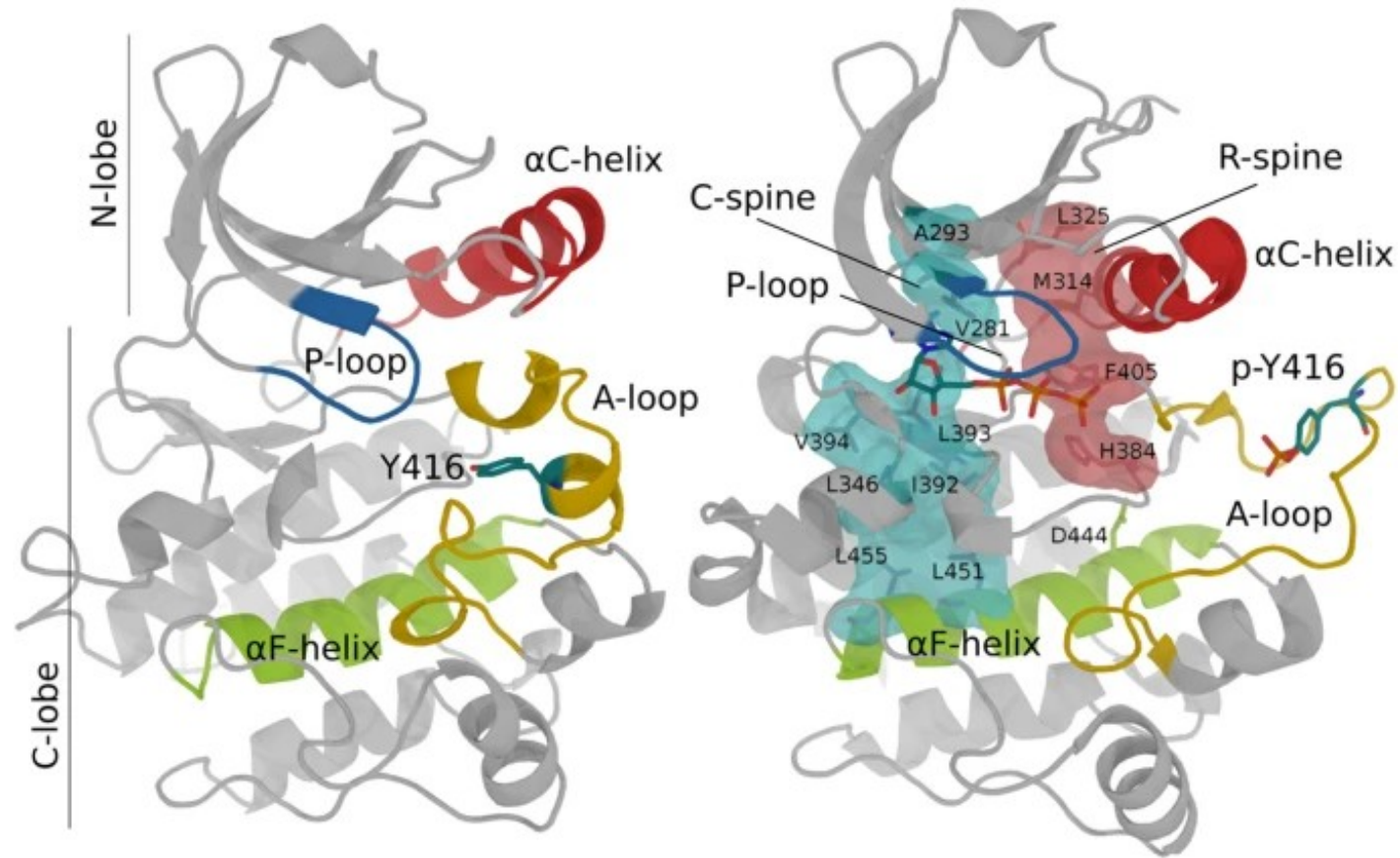


Inhibited protein kinase



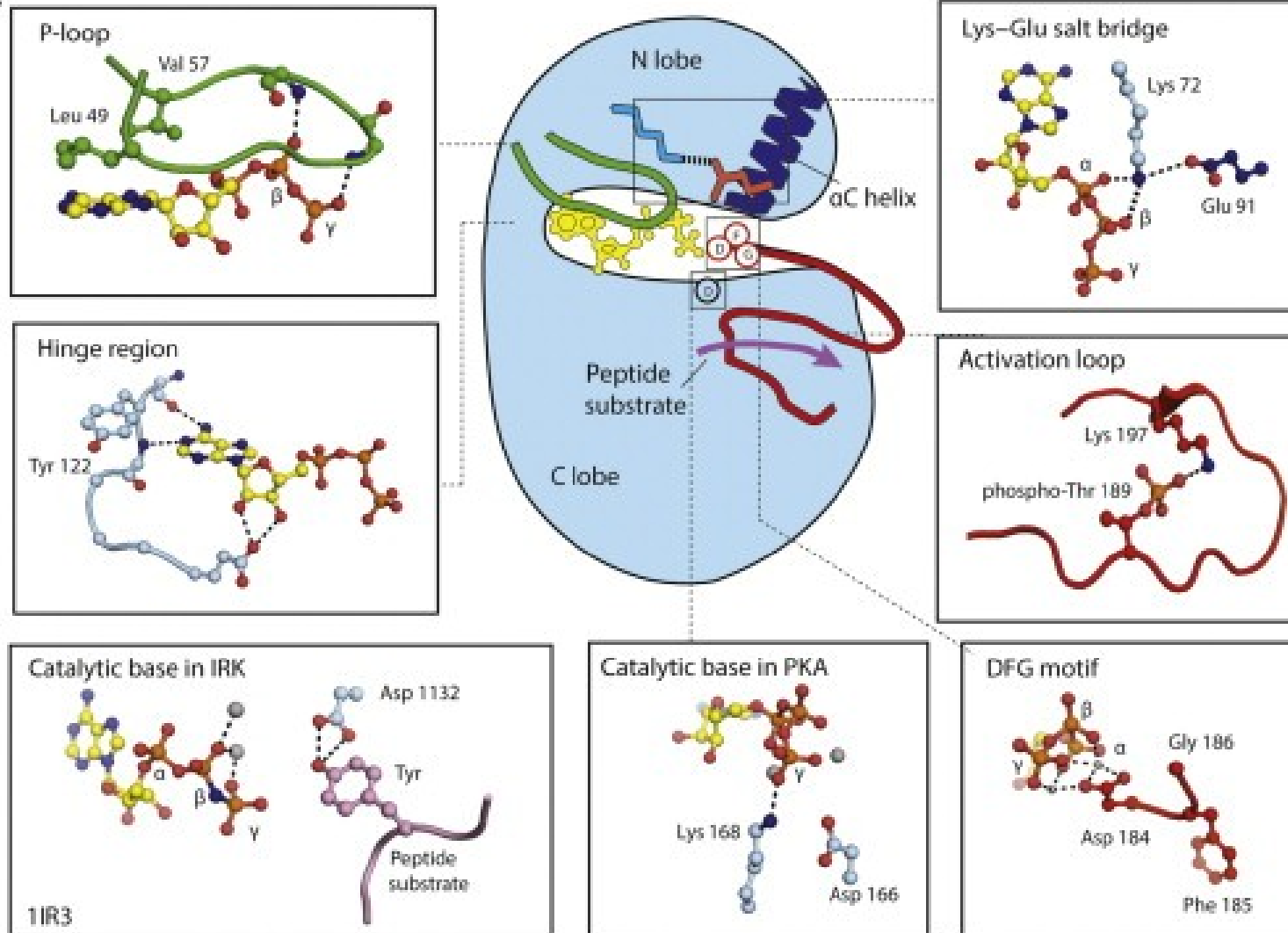
Remission

# structural elements

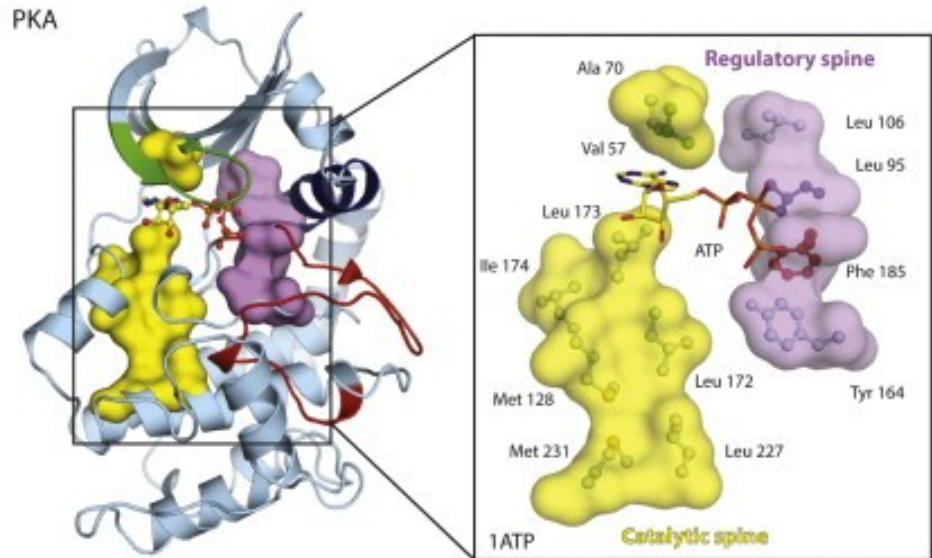


# structural elements

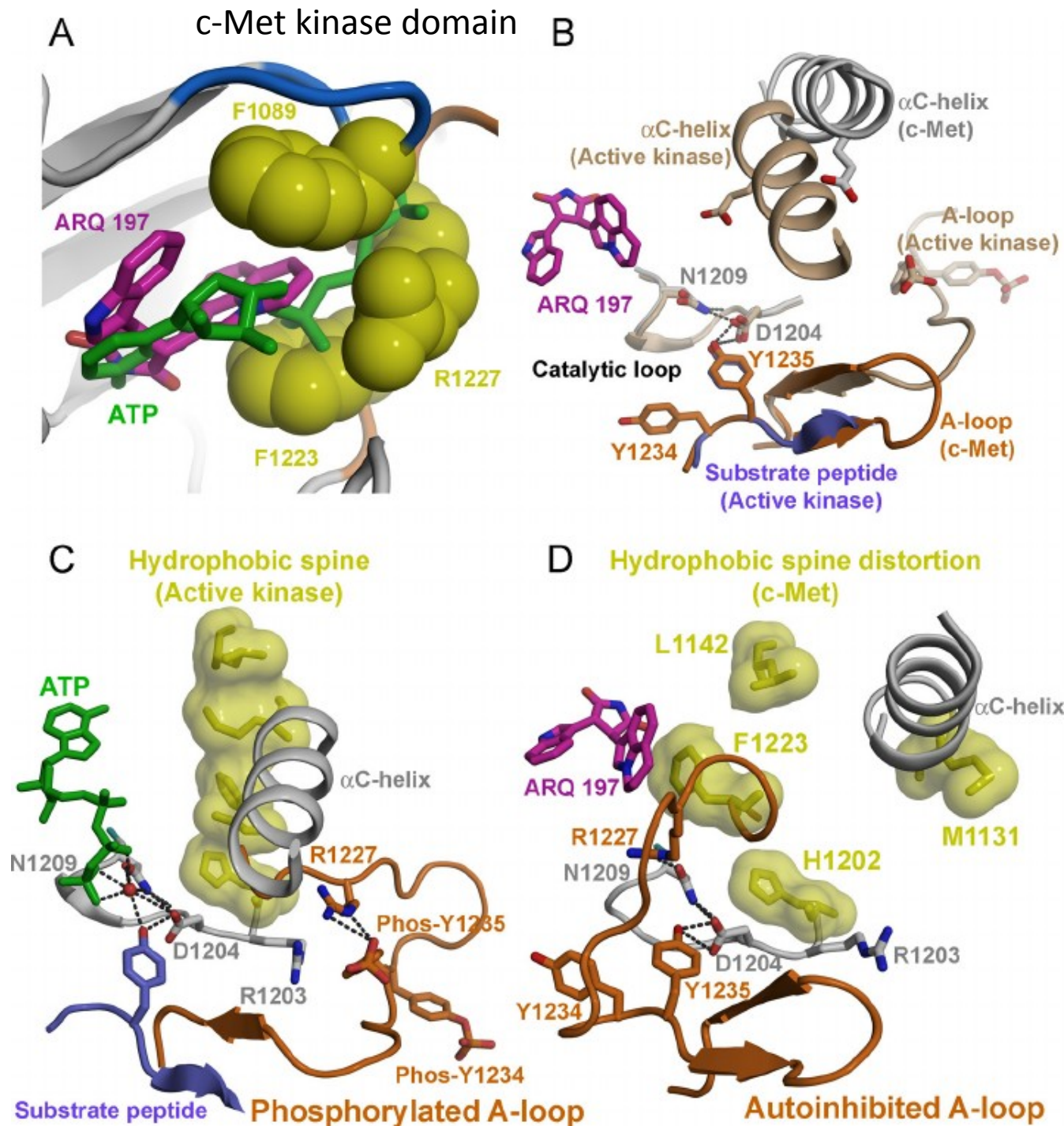
A



B  
PKA



# Intramolecular domain



# Intramolecular domain

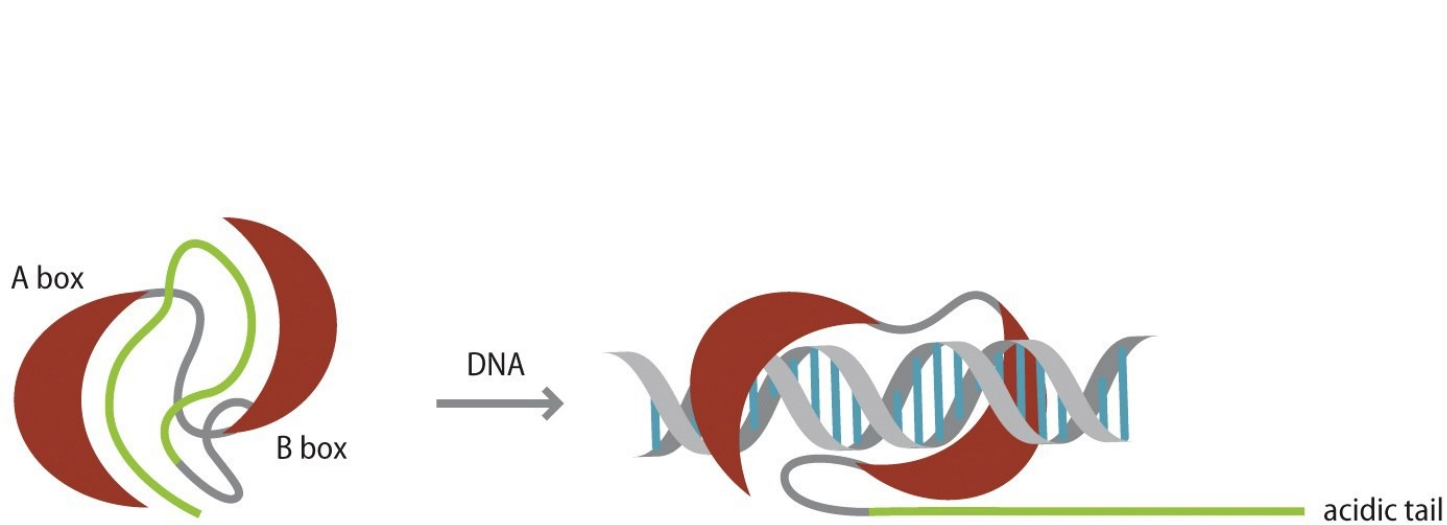
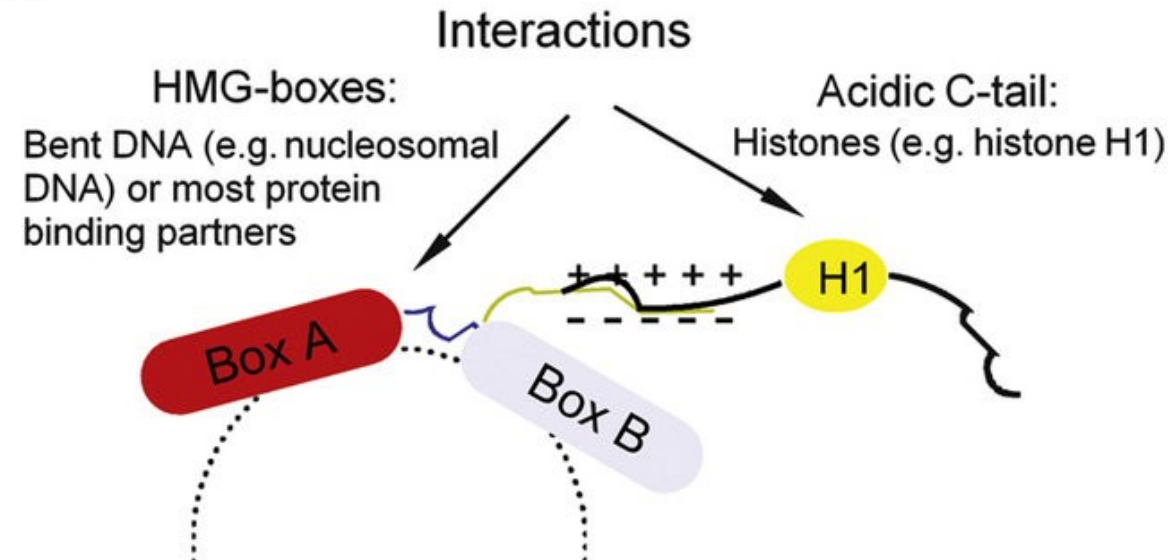
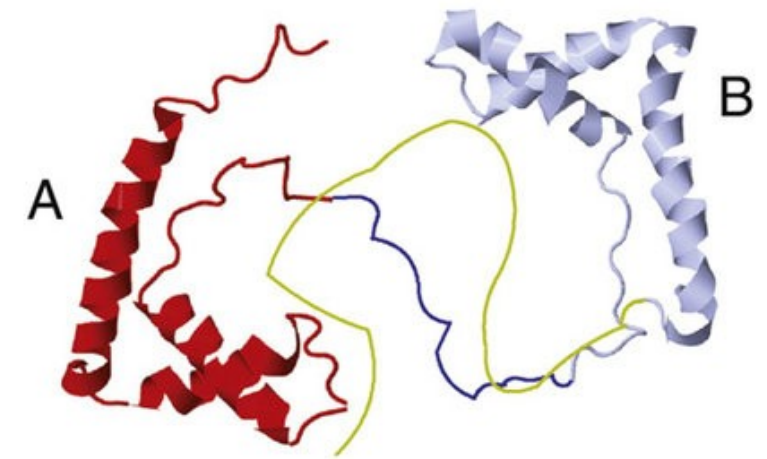
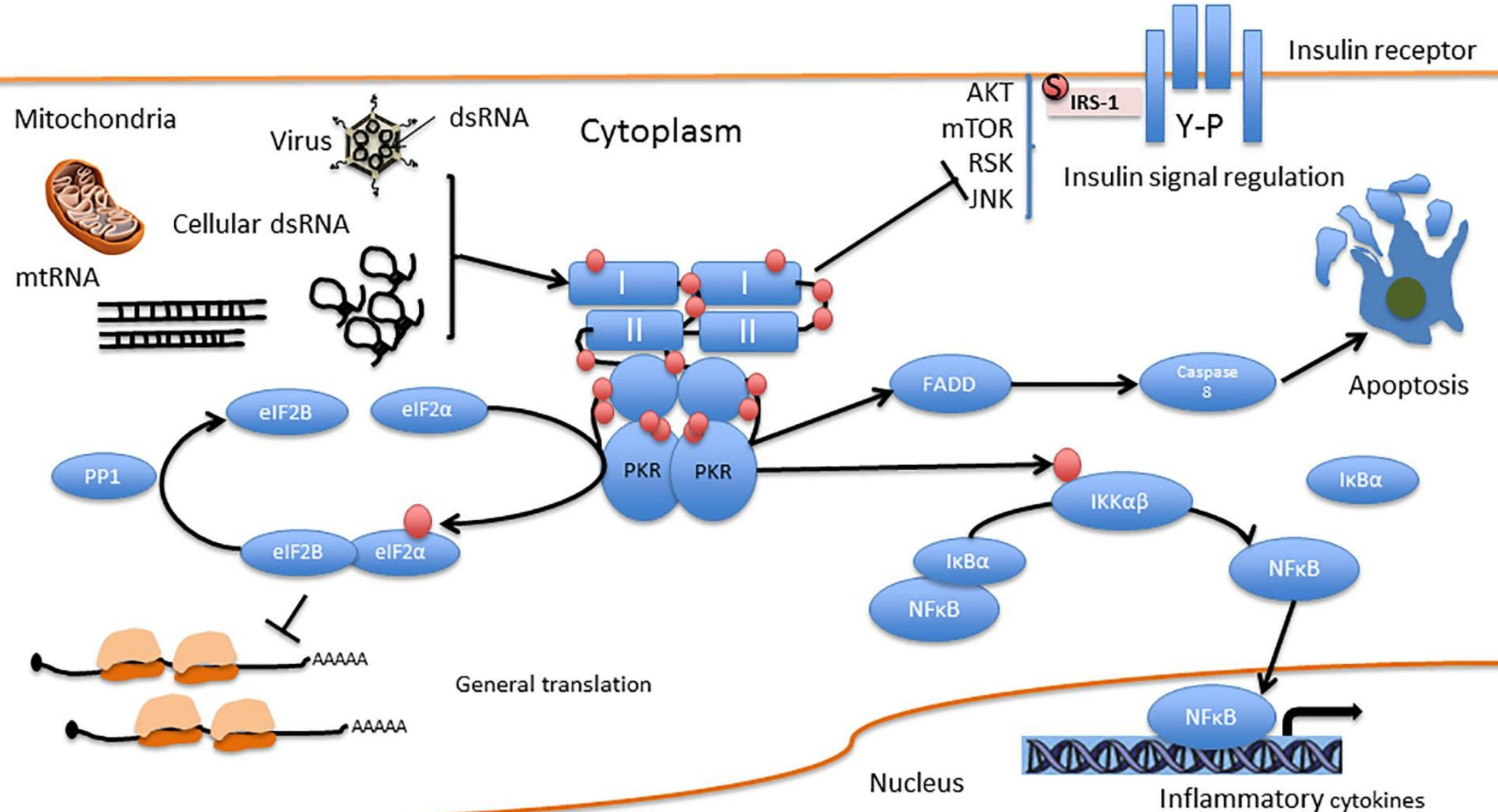


Figure 2.23 How Proteins Work (©2012 Garland Science)

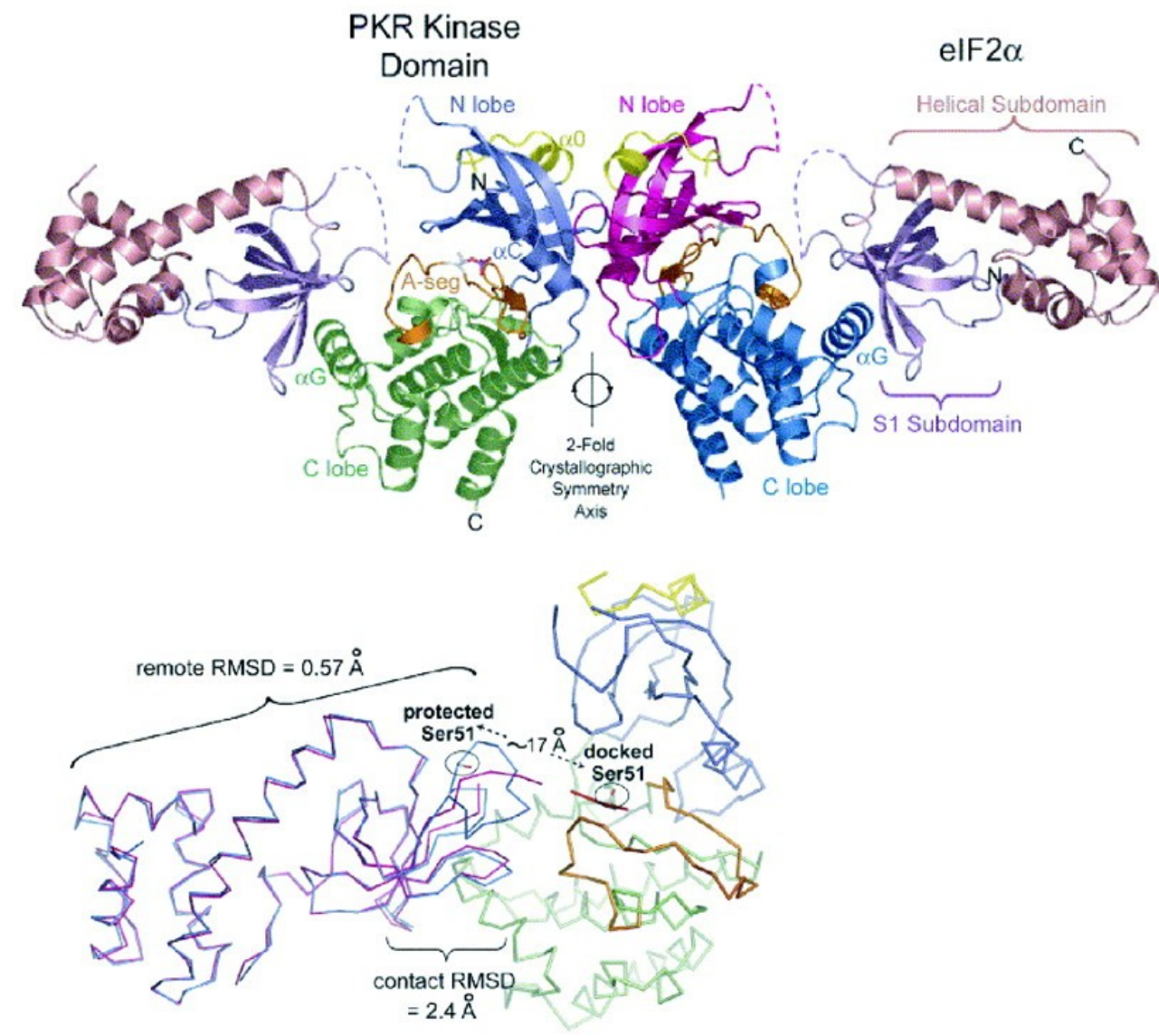
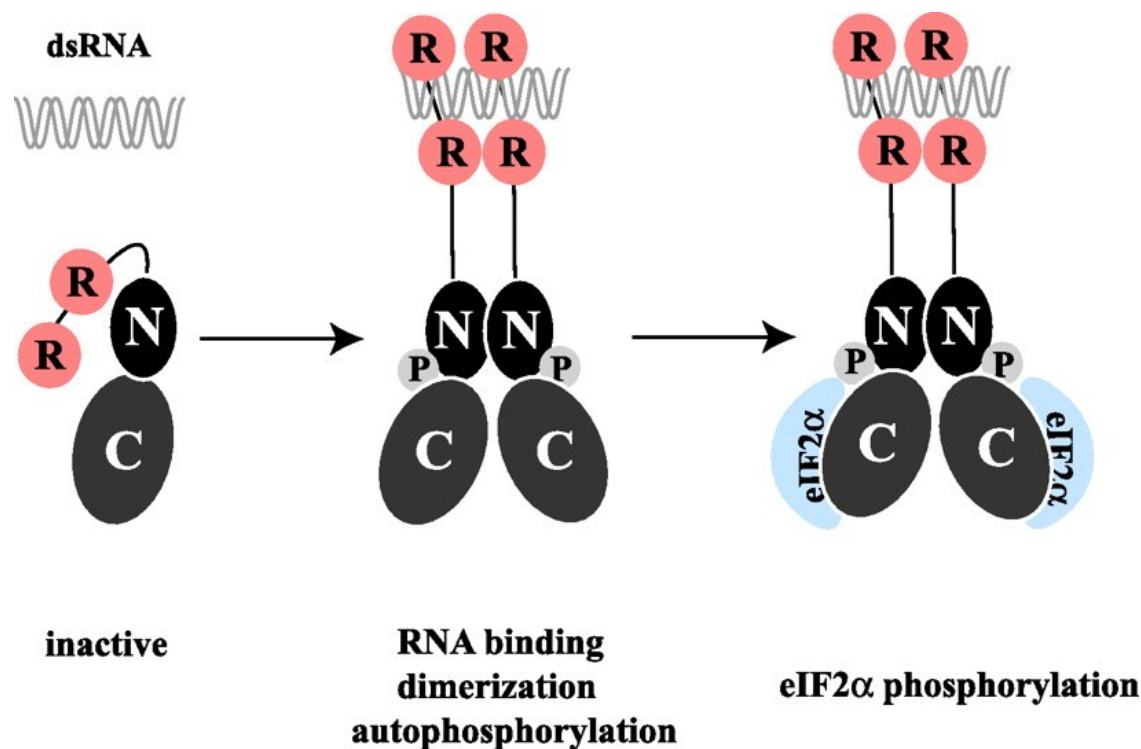


# Intramolecular domain

protein kinase R PKR regulate protein synthesis via the eIF2 $\alpha$  pathway.



# Intramolecular domain



M. A. García et al. Microbiol. Mol. Biol. Rev. 2006;  
doi:10.1128/MMBR.00027-06

# Binding specificity

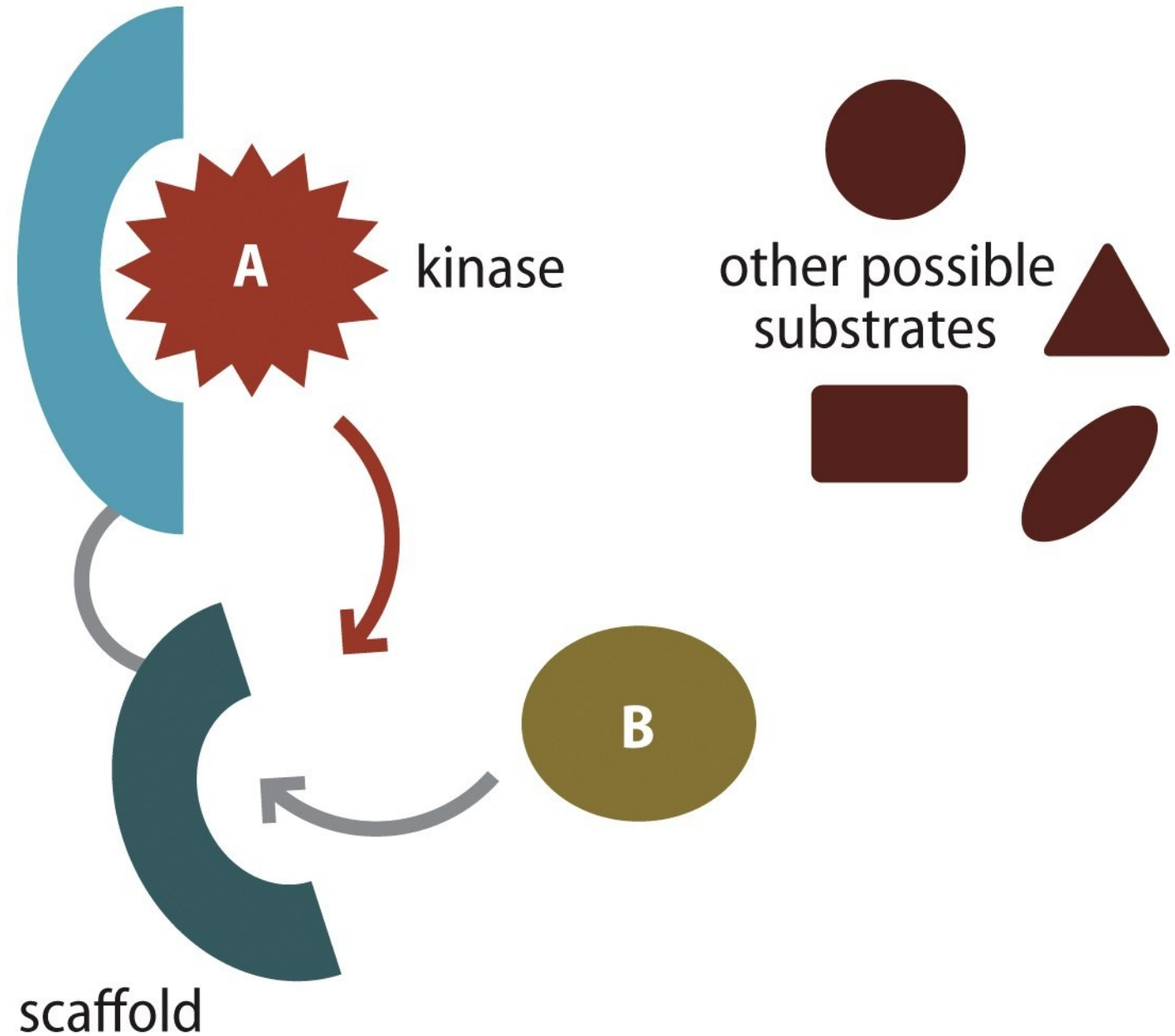
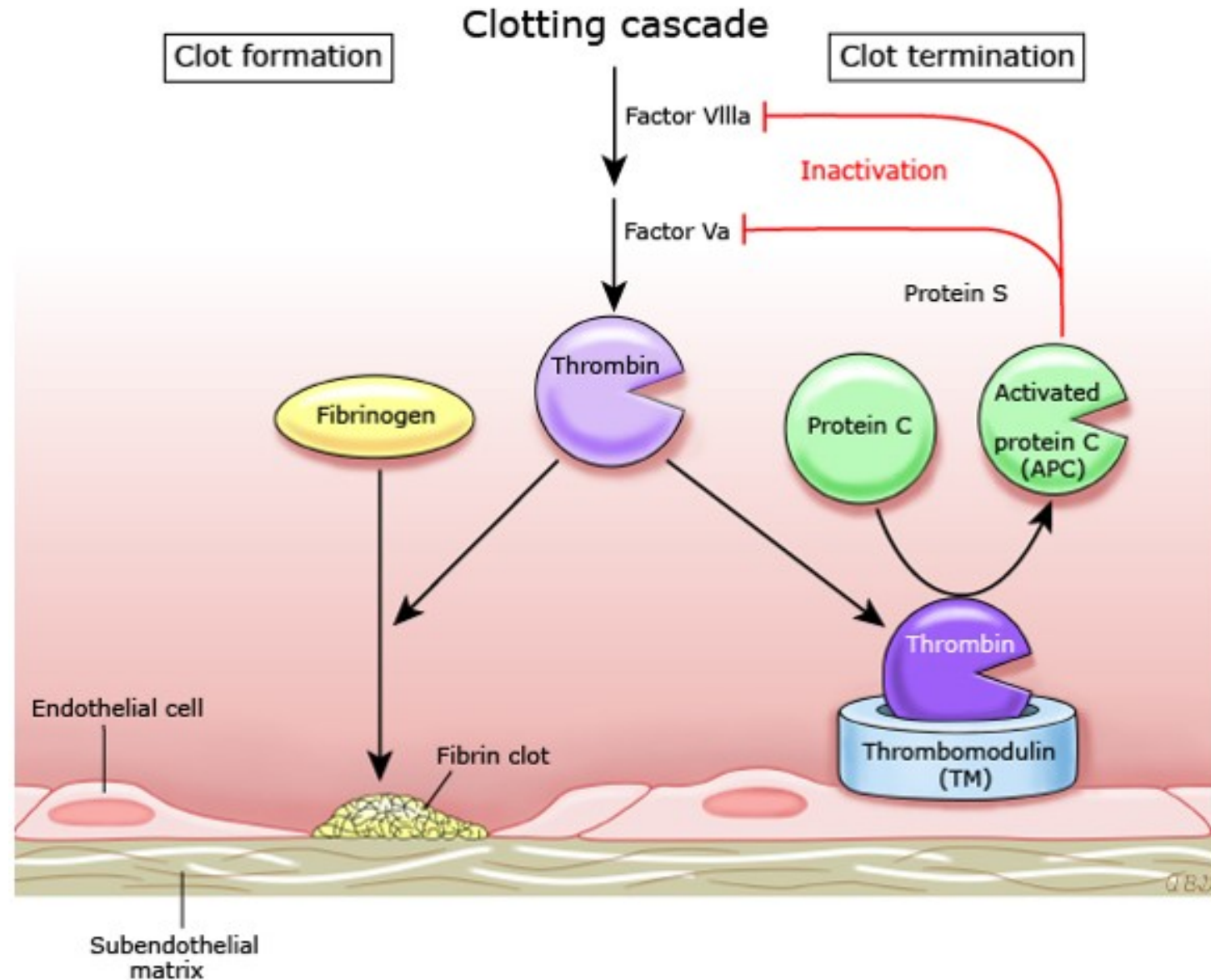


Figure 2.26 How Proteins Work (©2012 Garland Science)

# Binding specificity

θρομβομοντουλίνη



# Binding specificity

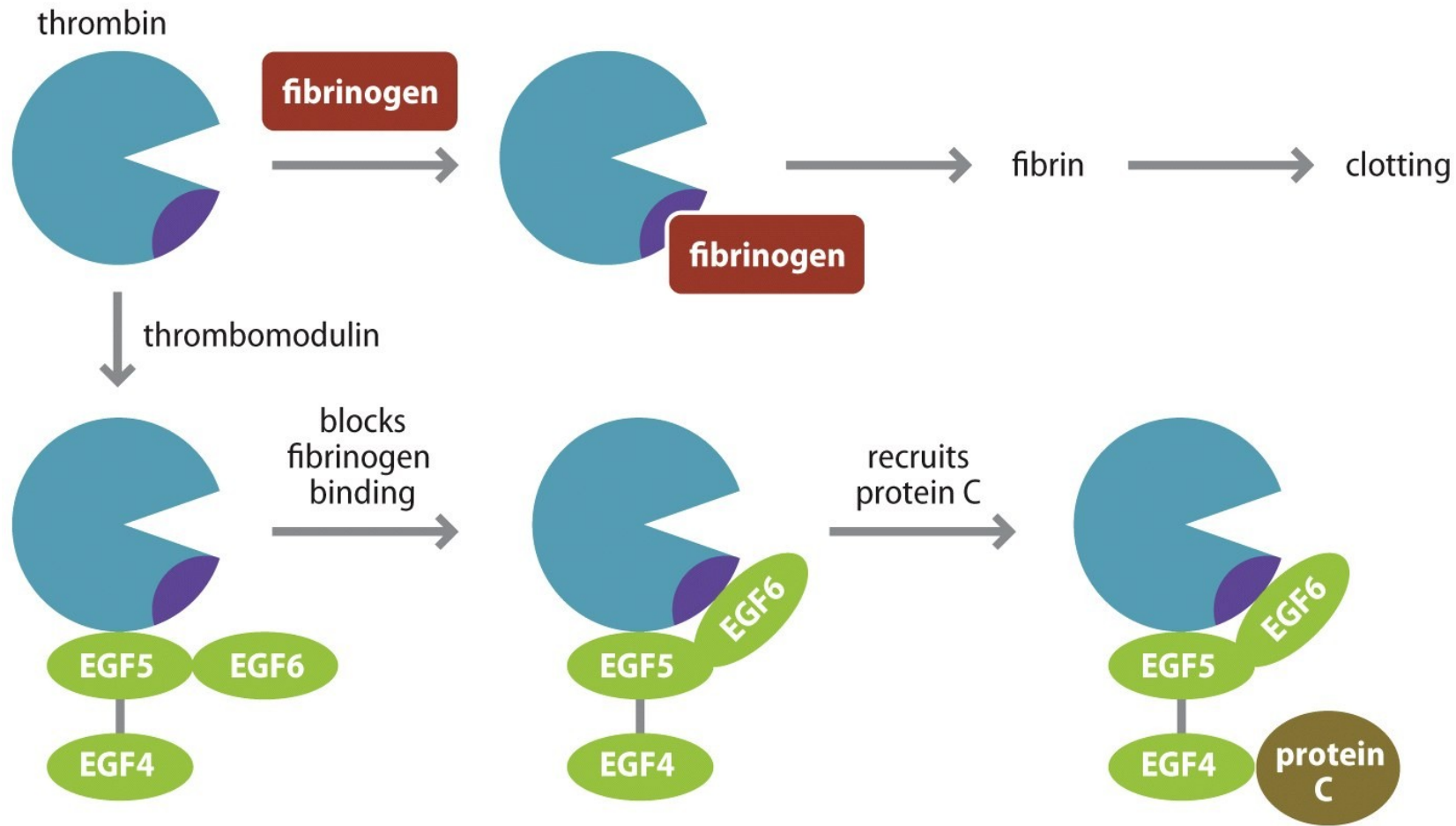
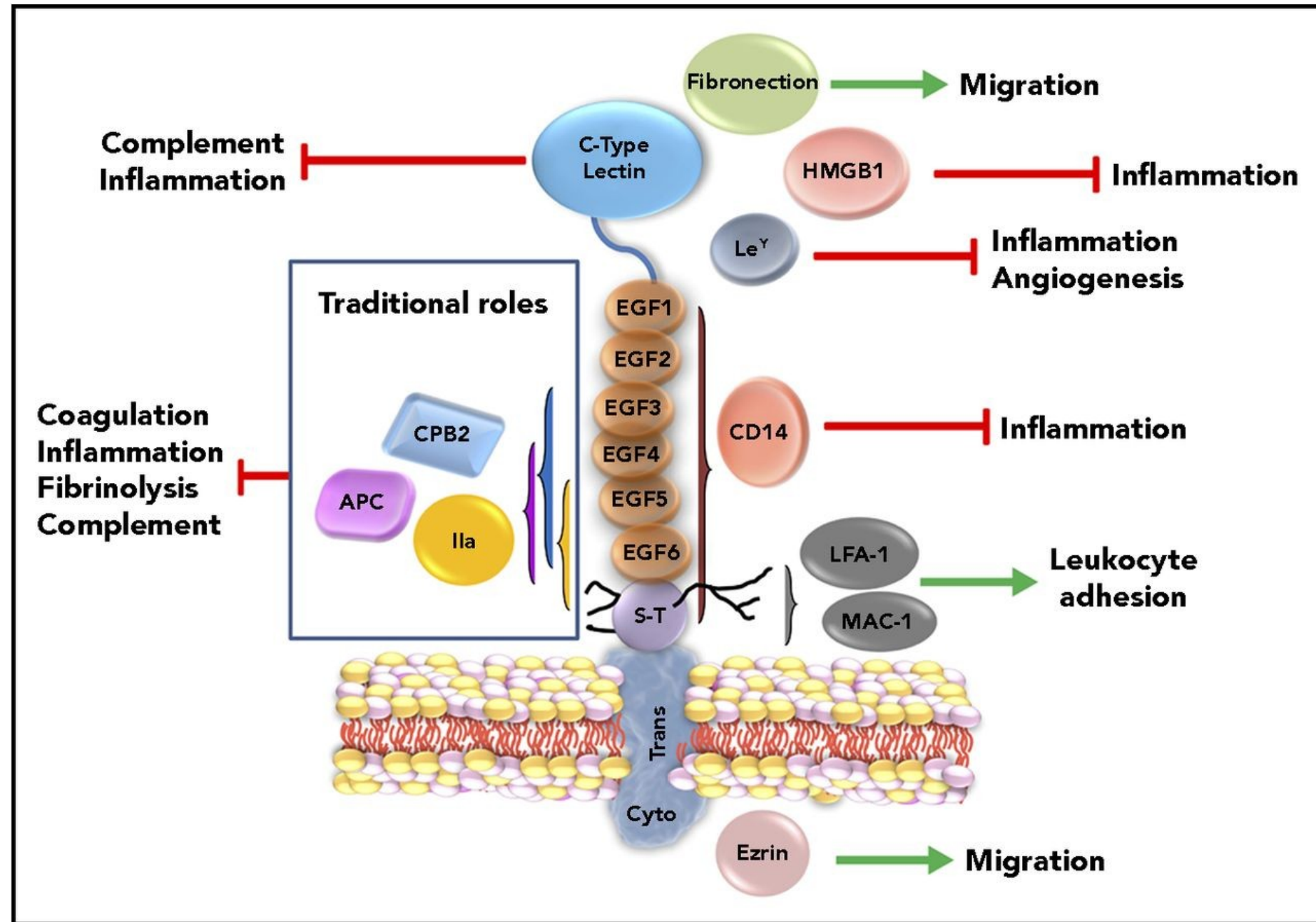


Figure 2.27 How Proteins Work (©2012 Garland Science)

# Binding specificity



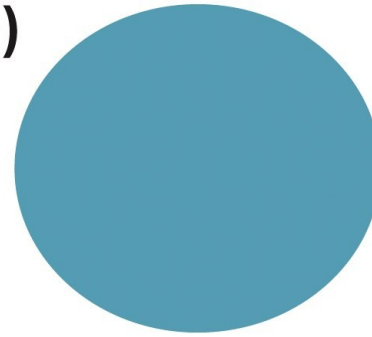
Houra Loghmani, Edward M. Conway,  
Exploring traditional and  
nontraditional roles for  
thrombomodulin, *Blood*, 2018



American Society of Hematology  
Helping hematologists conquer blood diseases worldwide

# Control and regulation

(a)



(b)

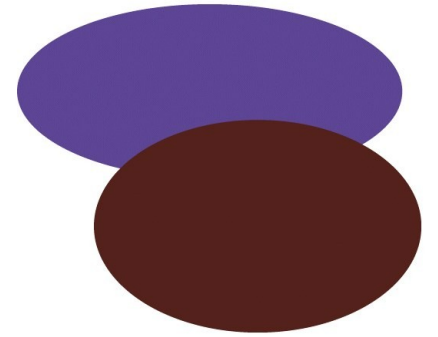


Figure 2.32 How Proteins Work (©2012 Garland Science)

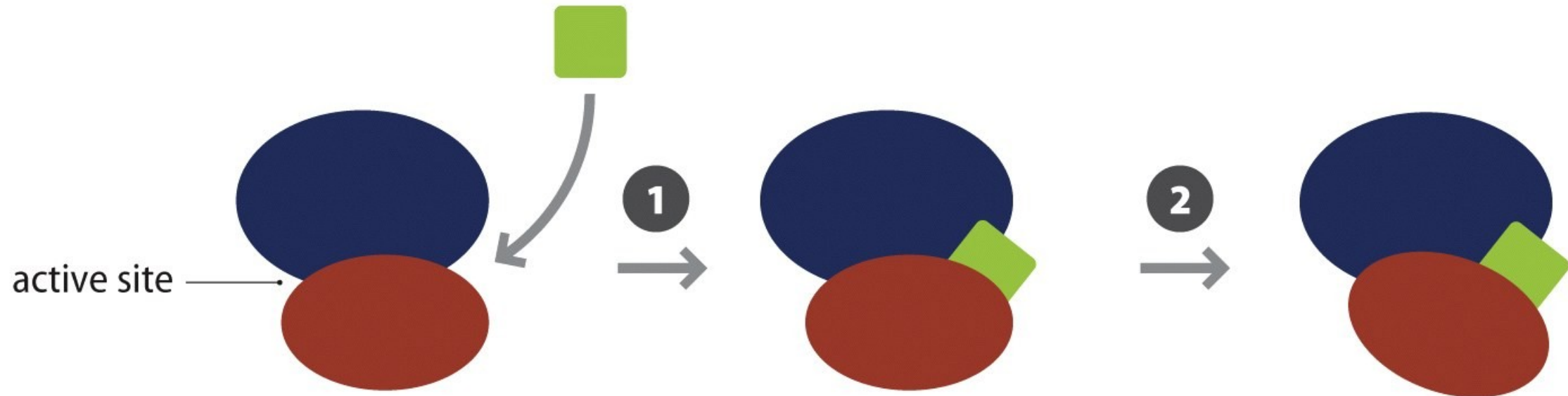


Figure 2.31 How Proteins Work (©2012 Garland Science)

# Enzyme regulation

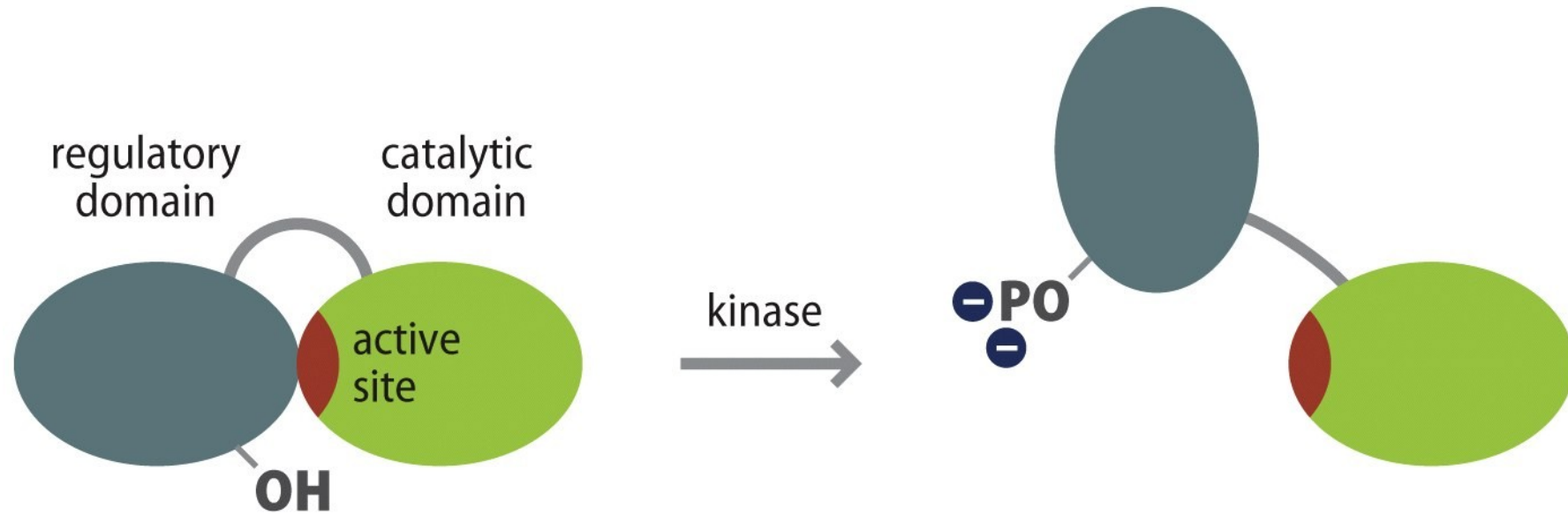


Figure 2.33 How Proteins Work (©2012 Garland Science)

# Bacterial two-component signaling system

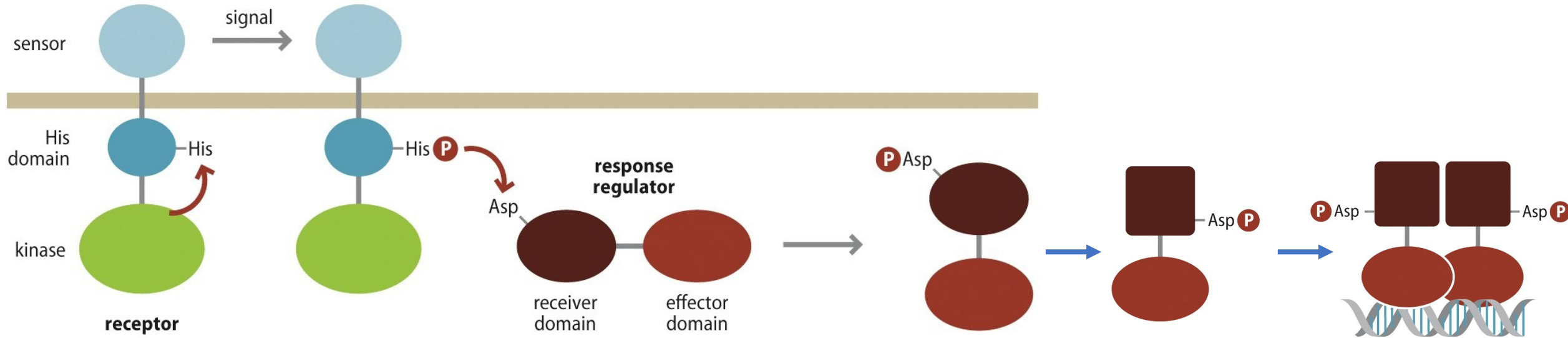


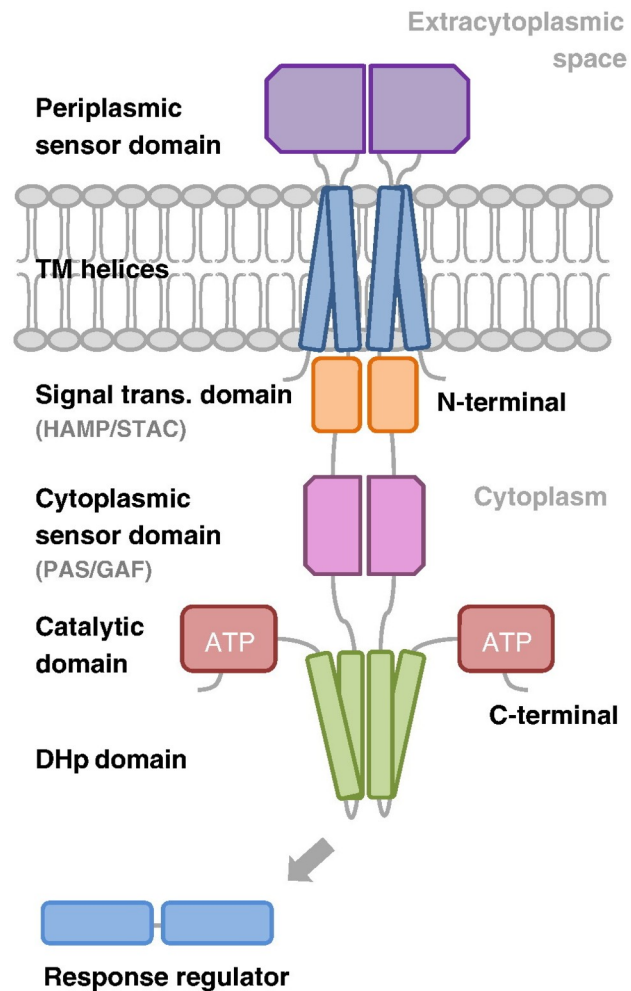
Figure 2.34 How Proteins Work (©2012 Garland Science)

Figure 2.36 How Proteins Work (©2012 Garland Science)

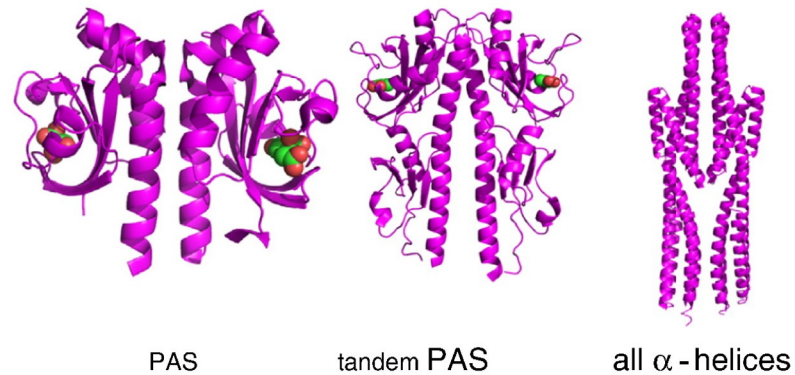
The proteins comprising the bacterial two-component system (TCS) are the sensor histidine kinase and the response regulator. These two factors are among the most abundant proteins in the sequence databases, owing to a wide distribution across the bacterial and archaeal kingdom and to significant amplification within bacterial and archaeal genomes

# Bacterial two-component signaling system

**(a) TCS domains**



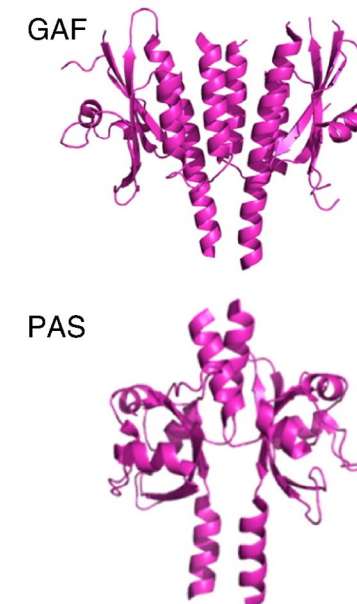
**(b) Extracytoplasmic sensor domain**



**(c) HAMP domain**



**(d) Cytoplasmic SD**



**(e) Kinase core**

