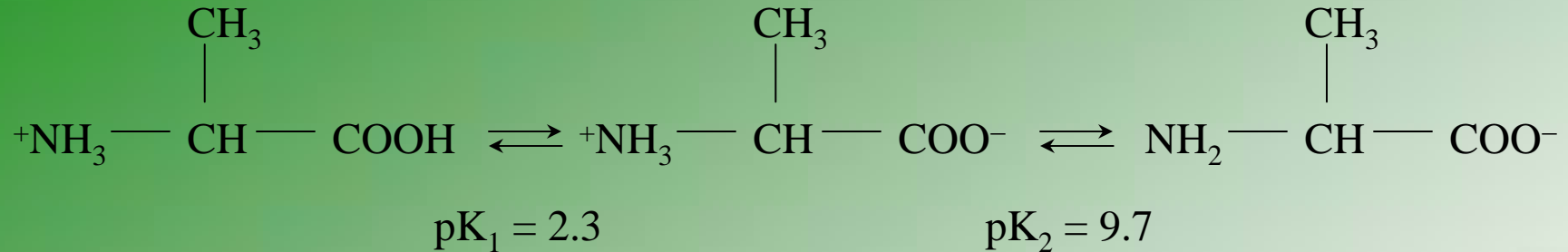


# Calculating Isoelectric Point

## Alanine:



The net charge of Ala is zero, when:

$$f_{\text{+HAH}} = f_{\text{A}^-}$$

$$f_{\text{NH}_3^+} \times f_{\text{COOH}} = f_{\text{NH}_2} \times f_{\text{COO}^-}$$

$$\left( \frac{\text{H}^+}{\text{K}_{\text{a}_1} + \text{H}^+} \right) \left( \frac{\text{H}^+}{\text{K}_{\text{a}_2} + \text{H}^+} \right) = \left( \frac{\text{K}_{\text{a}_1}}{\text{K}_{\text{a}_1} + \text{H}^+} \right) \left( \frac{\text{K}_{\text{a}_2}}{\text{K}_{\text{a}_2} + \text{H}^+} \right)$$

$$(\text{H}^+)^2 = \text{K}_{\text{a}_1} \times \text{K}_{\text{a}_2}$$

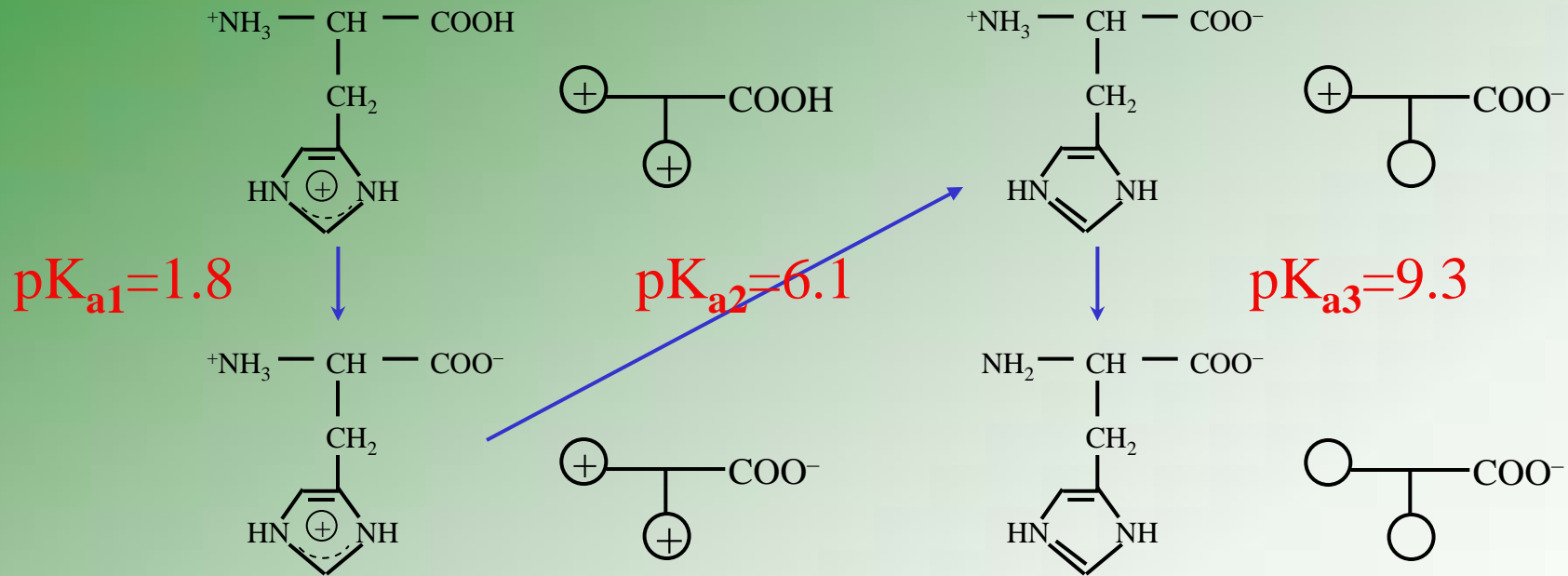
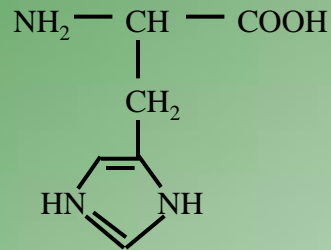
$$\text{pH}_{0 \text{ net charge}} = \text{pI} = \frac{\text{pK}_{\text{a}_1} + \text{pK}_{\text{a}_2}}{2}$$

$$\text{pI}_{\text{Ala}} = \frac{2.3 + 9.7}{2} = 6.0$$

# Calculating the Isoelectric Point of a Protein

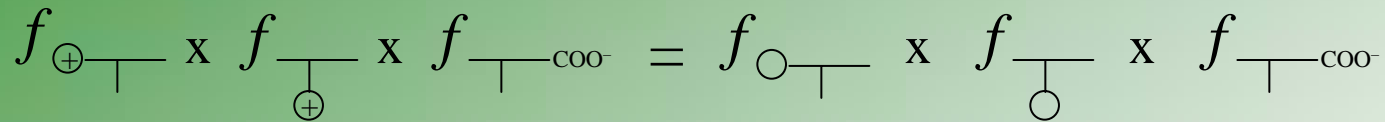
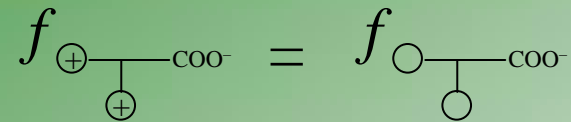
Another example...

**Histidine**



# Calculating Isoelectric Point

Calculations... At the isoelectric point,



$$\left( \frac{K_{a_1}}{K_{a_1} + H^+} \right) \left( \frac{H^+}{K_{a_2} + H^+} \right) \left( \frac{H^+}{K_{a_3} + H^+} \right) = \left( \frac{K_{a_1}}{K_{a_1} + H^+} \right) \left( \frac{K_{a_2}}{K_{a_2} + H^+} \right) \left( \frac{K_{a_3}}{K_{a_3} + H^+} \right)$$

$$(H^+)^2 = K_{a_2} \times K_{a_3} \quad \text{(Please correct + for x sign)}$$

$$pI = \frac{pK_{a_2} + pK_{a_3}}{2}$$

$$pI_{\text{His}} = \frac{6.1 + 9.3}{2} = 7.7$$

Let us calculate the pI for a protein...

# Calculating Isoelectric Point

Consider a typical protein: **CheY** from *E. coli*

Sequence...

1 ADKELKFLVV DDFSTMRRIV RNLLKELGFN NVEEAEDGVD ALNKLQAGGY  
 51 GJVISDWNMP NMDGLELLKT IRADGAMSAL PVLMTAEAK KENIIAAAQA  
 101 GASGYVVKPF TAATLEELN KIFEKLGGM

Molecular Weight = 13966.04 Da  
 Average Residue Weight = 109.110 Da  
 Residues = 128

Residue	Number	Mole Percent
A = Ala	16	12.500
B = Asx	0	0.000
C = Cys	0	0.000
D = Asp	8	6.250
E = Glu	11	8.594
F = Phe	6	4.688
G = Gly	10	7.813
H = His	0	0.000
I = Ile	6	4.688
K = Lys	11	8.594
L = Leu	15	11.719
M = Met	6	4.688
N = Asn	8	6.250
P = Pro	3	2.344
Q = Gln	2	1.563
R = Arg	4	3.125
S = Ser	4	3.125
T = Thr	5	3.906
V = Val	10	7.813
W = Trp	1	0.781
Y = Tyr	2	1.563
Z = Glx	0	0.000

Residue	Number	Mole Percent
A + G	26	20.313
S + T	9	7.031
D + E	19	14.844
D + E + N + Q	29	22.656
H + K + R	15	11.719
D + E + H + K + R	34	26.563
I + L + M + V	37	28.906
F + W + Y	9	7.031

Let us calculate the pI  
 from the provided  
 information...

# Calculating Isoelectric Point

Negative charges	1 COOH	-1	-20	At pH ~ 7, there is a <b>-4 net charge</b> since pKs of COOH, Glu, and Asp are <5,
	11 Glu	-11		
	8 Asp	-8		
Positive charges	1 NH <sub>3</sub> <sup>+</sup>	+1	+16	
	11 Lys	+11		
	4 Arg	+4		

pI is pH for which:

$$[11 \times f_{\text{Glu}^-}] + [8 \times f_{\text{Asp}^-}] + [1 \times f_{\text{COO}^-}] = 16$$

Try pH = 5...

# Calculating Isoelectric Point

Glu  $pK_a = 4.5$

$K_a = 3 \times 10^{-5} \text{ M}$

Asp  $pK_a = 4$

$K_a = 1 \times 10^{-4} \text{ M}$

C-term  $pK_a = 3.5$

$K_a = 3 \times 10^{-4} \text{ M}$

$$11 \times \frac{K_{\text{Glu}}}{K_{\text{Glu}} + H^+} = 11 \times \frac{3 \times 10^{-5}}{3 \times 10^{-5} + 1 \times 10^{-5}} = 11 \times \frac{3}{4} \approx 8$$

$$8 \times \frac{K_{\text{Asp}}}{K_{\text{Asp}} + H^+} = 8 \times \frac{10 \times 10^{-5}}{10 \times 10^{-5} + 1 \times 10^{-5}} = 8 \times \frac{10}{11} \approx 7.2$$

C-terminus = -1

Asp = -7.2

Glu = -8

-16.2 So... pI ~ 5

# Protein Structure

# Structural Biology

---

**Structural biology** is now an essential aspect of biochemistry and cell biology

- 2 - 3(5 - 7) new structures are solved every day
- Total structural genomics:
  - determine the folding pattern for every **protein family**
- As we will see soon, the structure allows a sophisticated, chemical understanding of:
  - Enzyme mechanisms
  - Binding of chemical signal-molecules that regulate function
  - Design of new inhibitors and drugs

# Structural Biology

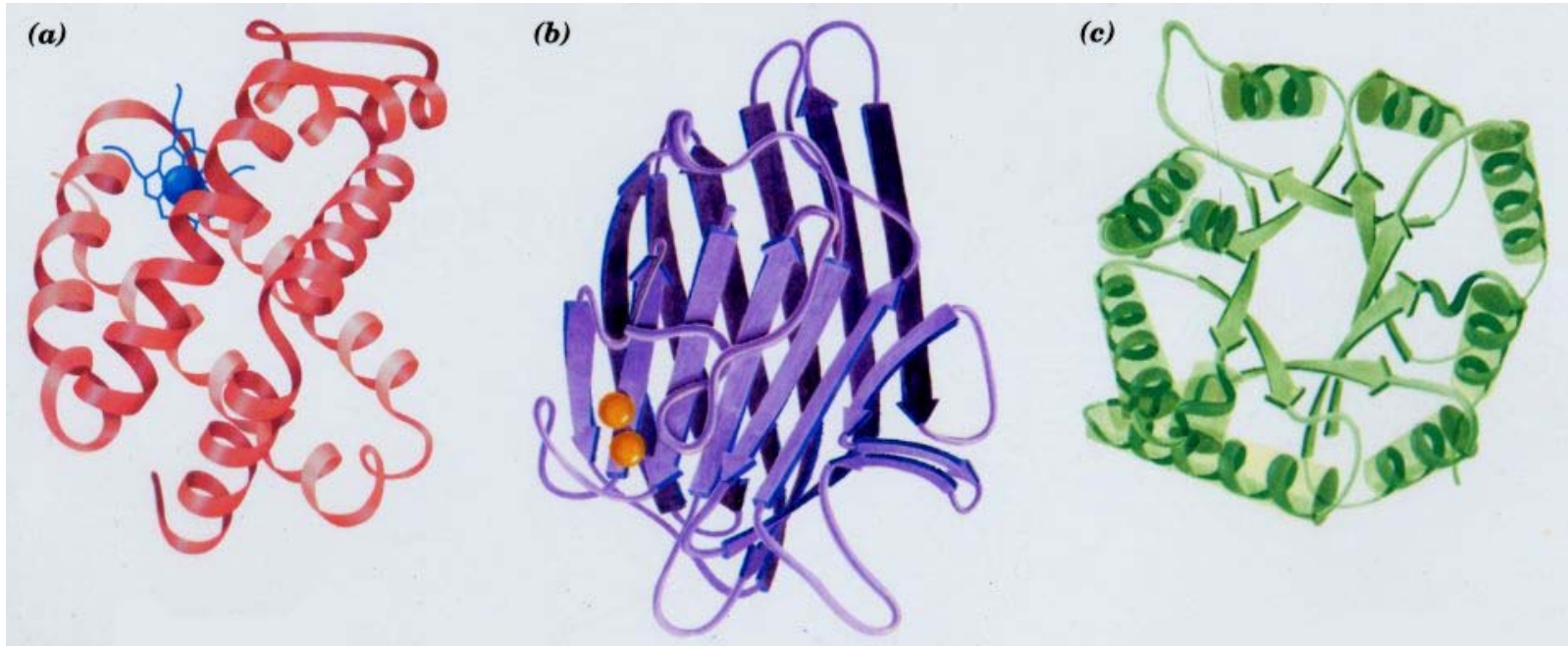
---

## Motifs and Domains

Helices and sheets often combine in various ways. Some proteins are purely helical, and some are purely  $\beta$ -sheets. Most are combinations of  $\alpha$  and  $\beta$ .

- Certain combinations of  $\alpha$  and  $\beta$  are called **supersecondary structures** or **MOTIFS**
- These occur in many unrelated globular proteins, as revealed by the thousands of structures solved to date

examples...



A Hb subunit

Concavalin A

Triose  
Phosphate  
Isomerase

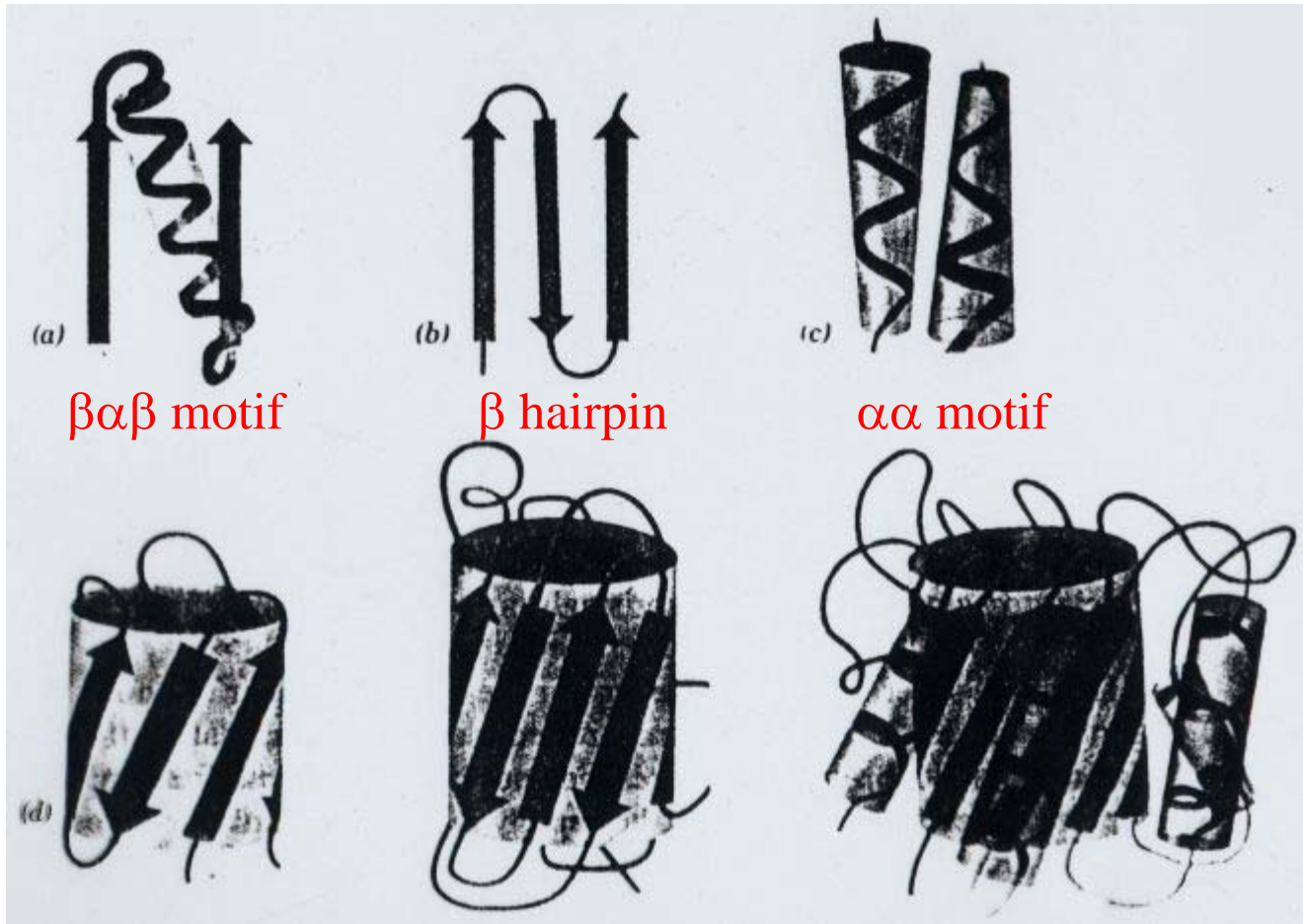
7 motifs...

# Structural Biology. Motifs

---

- ❑  $\beta\alpha\beta$  motif: the most common
- ❑  $\beta$  haripin: anti-parallel strands connected by reverse turns
- ❑  $\alpha\alpha$  motif: two successive anti-parallel  $\alpha$  helices, which are packed with their axes inclined
- ❑  $\beta$ -barrels:  $\beta$ -sheets rolling up into barrels
- ❑ Helix-loop-helix: helix-turn-helix
- ❑ Leucine zipper motif: dimerization/trimerization domain

representations of above...



$\beta$  barrels –  
different  
types

These motifs are incorporated into larger **DOMAINS**

# Structural Biology

---

## Domains

- Despite the huge number of possible sequences, certain arrangements or patterns of secondary structures appear to repeat over and over.
- These occur in regions that show little sequence similarity and show one or more combinations of motifs

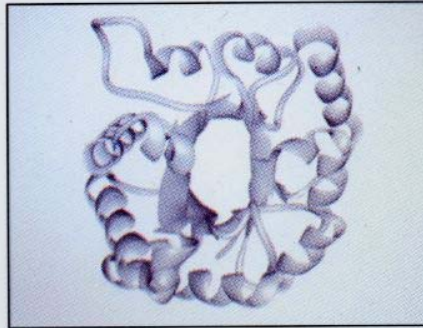
- $\alpha/\beta$  barrel

- Four helix bundle

- $\alpha/\beta$  saddle

- $\beta/\beta$  sandwich

**$\alpha/\beta$  Barrel**

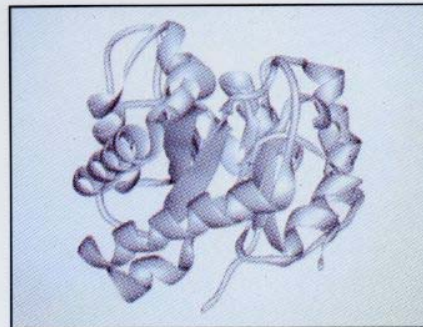


Triose phosphate  
isomerase  
(top view)

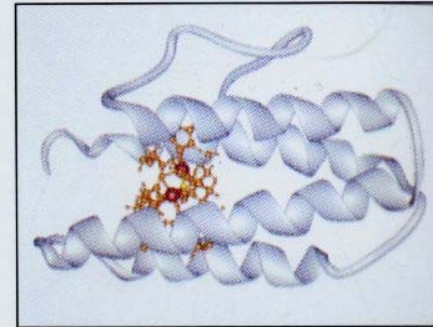
**Four-helix bundle**



Myohemerythrin  
(top view)



Triose phosphate  
isomerase  
(side view)

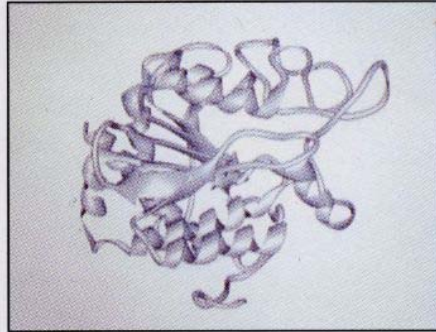


Myohemerythrin  
(side view)

**(a)**

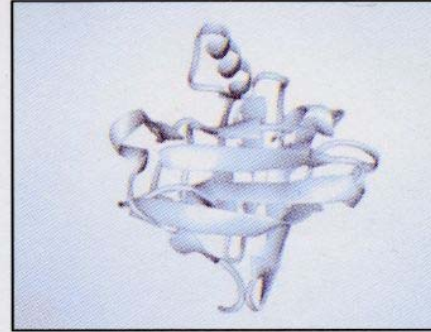
**(b)**

**$\alpha\beta$  with saddle at core**

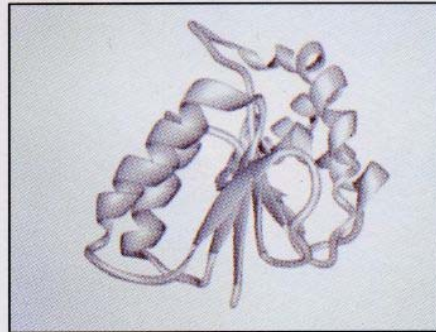


Carboxypeptidase

**$\beta\text{-}\beta$  Sandwich**

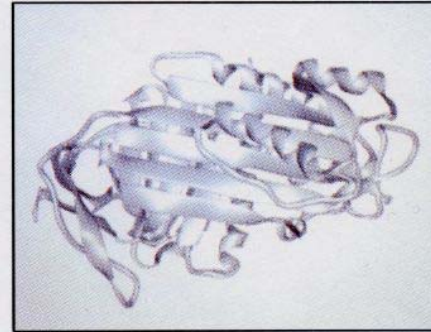


Insecticyanin



Lactate  
dehydrogenase  
domain 1

**(c)**



$\alpha_1$ -Antitrypsin

**(d)**

# Structural Biology

---

## **Domains or 'folds':** (see fig. 6-30)

- Compact globular structures. Domains are *structurally independent* units that have the **characteristics of a small globular protein**
- 2 or more often connected by a loop or a hinge. Homologous proteins from another organism may be expressed as 2 separate autonomously folded subunits
- Typical size is 100-200 aa, and the average diameter  $\sim 25 \text{ \AA}$

## **Basic Rules:**

- Non-polar side-chains are buried in the interior of domains
- Surface of domains are a mixture of polar and non-polar residues
- Assembly of domains into longer clusters uses both hydrophobic contacts and complementary charge interactions
- Often it is possible to cleave the loops or hinges and retains structure, and sometimes function

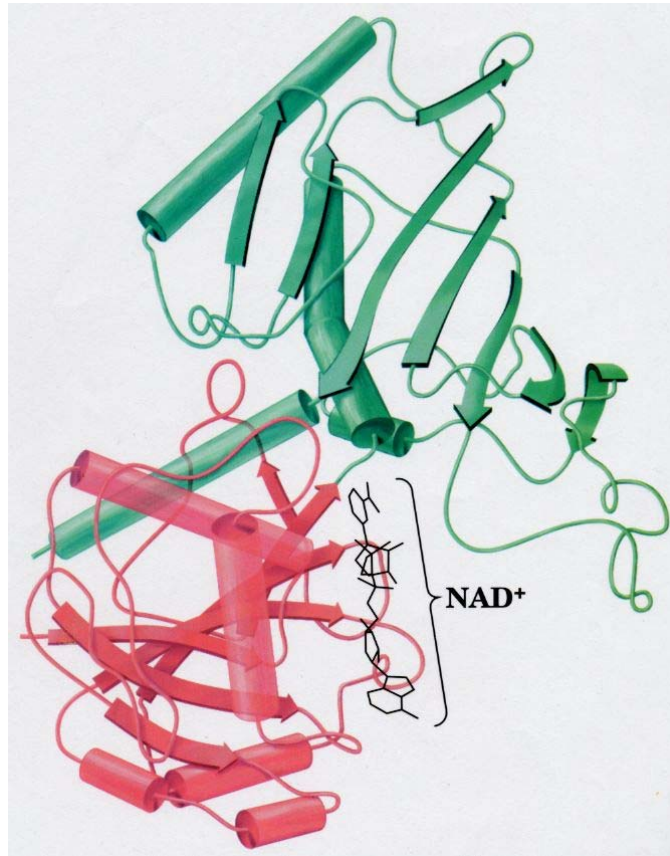
# Structural Biology

---

How is this multi-domain architecture used in biochemistry?

- Often the **domains** seem to possess **specialized functions**:
  - DNA binding domains
  - RNA binding domains
  - Ligand (regulatory) domains
  - Oligomerization domains
- In other instances ligands or substrates bind in the cleft (groove) between two domains and each separate domain contributes one or more catalytic residues to the active site

Example of multi-domain protein...



Glyceraldehyde-3-Phosphate Dehydrogenase

# Structural Biology

---

## Protein Families

The PDB (protein data bank) is a publicly available data base of all solved protein structures:

<http://www.rcsb.org/pdb>

# Structural Biology

---

- The data bank can be searched for multiple occurrences of motifs and whole folds or domains. You'll find:
  - A small number of deposits are clearly related, sharing motifs or domains. These are defined as a **FAMILY**
  - Unrelated proteins often are found to belong to different families
- There are currently several hundred unique protein domains or folds. We expect ~1000 unique (distinct) domains or folds, once all structures are known
- Within a family, structure is closely conserved to a far greater extent than is the case for sequence

# Structural Biology

---

## Tertiary Structure Prediction:

How well can we predict 3<sup>o</sup> structure from the aa sequence?

1. It would help if we could first predict 2<sup>o</sup> structure

2. Idea: Look for trends in the data base

- Chow and Fasman did (see table 6-1)
- Gives a quantitative estimation of whether an aa prefers  $\alpha$ -helix or  $\beta$ -sheet
- Add these up as a running average to see whether a given run has a local preference for helix or sheet
- Only about 70% accurate because 3<sup>o</sup> contacts have a major impact on the formulation of 2<sup>o</sup> structure

**The so-called Folding Problem probably has no solution in the strict sense.**

# Structural Biology

---

Propensity of aa residues  
To adopt  $\alpha$  helical or  $\beta$   
Sheet conformations:

Residue	$P_{\alpha}$	$P_{\beta}$
Ala	1.42	0.83
Arg	0.98	0.93
Asn	0.67	0.89
Asp	1.01	0.54
Cys	0.70	1.19
Gln	1.11	1.10
Glu	1.51	0.37
Gly	0.57	0.75
His	1.00	0.87
Ile	1.08	1.60
Leu	1.21	1.30
Lys	1.16	0.74
Met	1.45	1.05
Phe	1.13	1.38
Pro	0.57	0.55
Ser	0.77	0.75
Thr	0.83	1.19
Trp	1.08	1.37
Tyr	0.69	1.47
Val	1.06	1.70