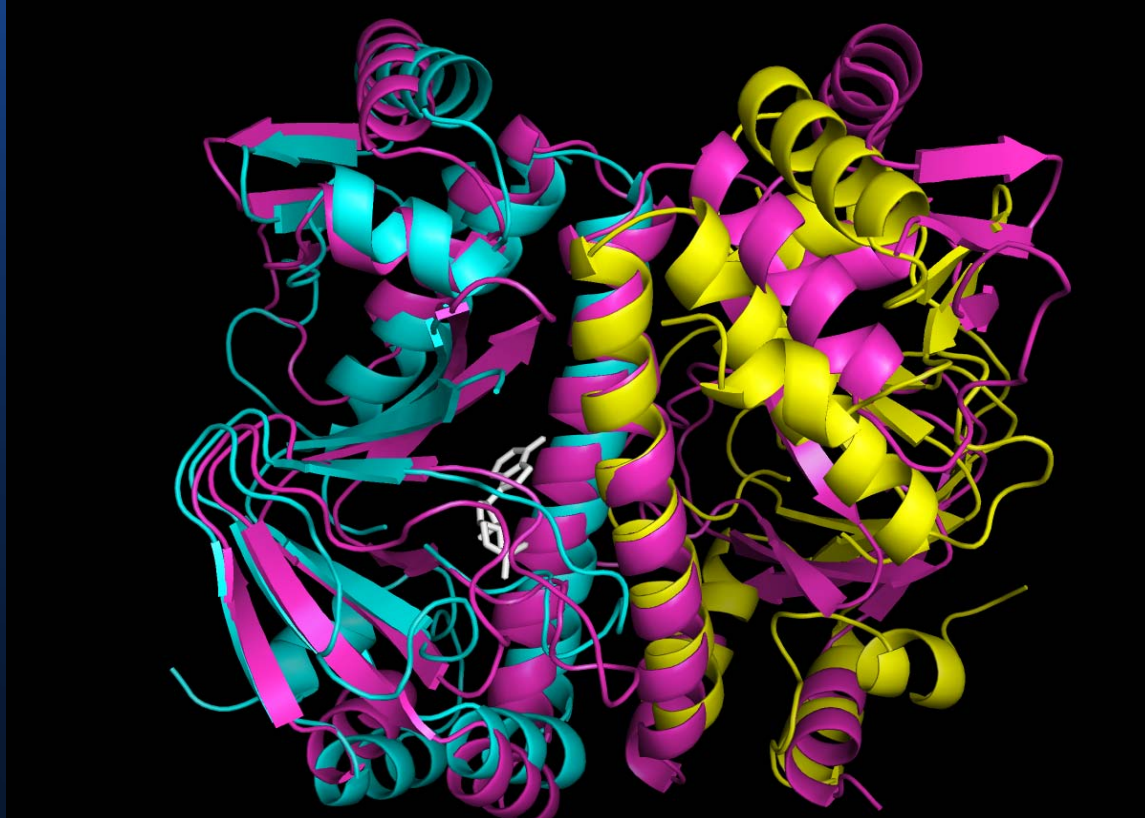


# *Project 1: Angle-Based Pairwise Structural Alignment Using Dynamic Programming*

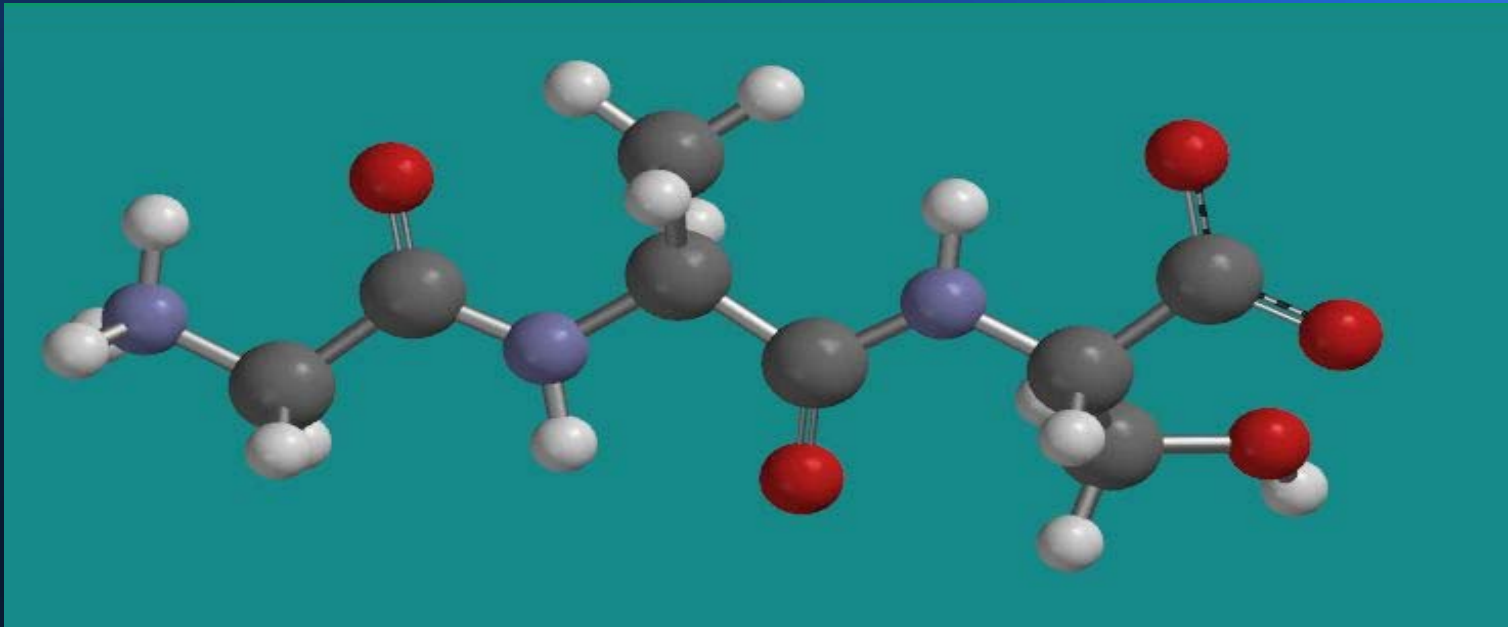


*Project Group:*

*Greg Beckham, Ray Terryn, Richie Hoch*

# Primary Structure

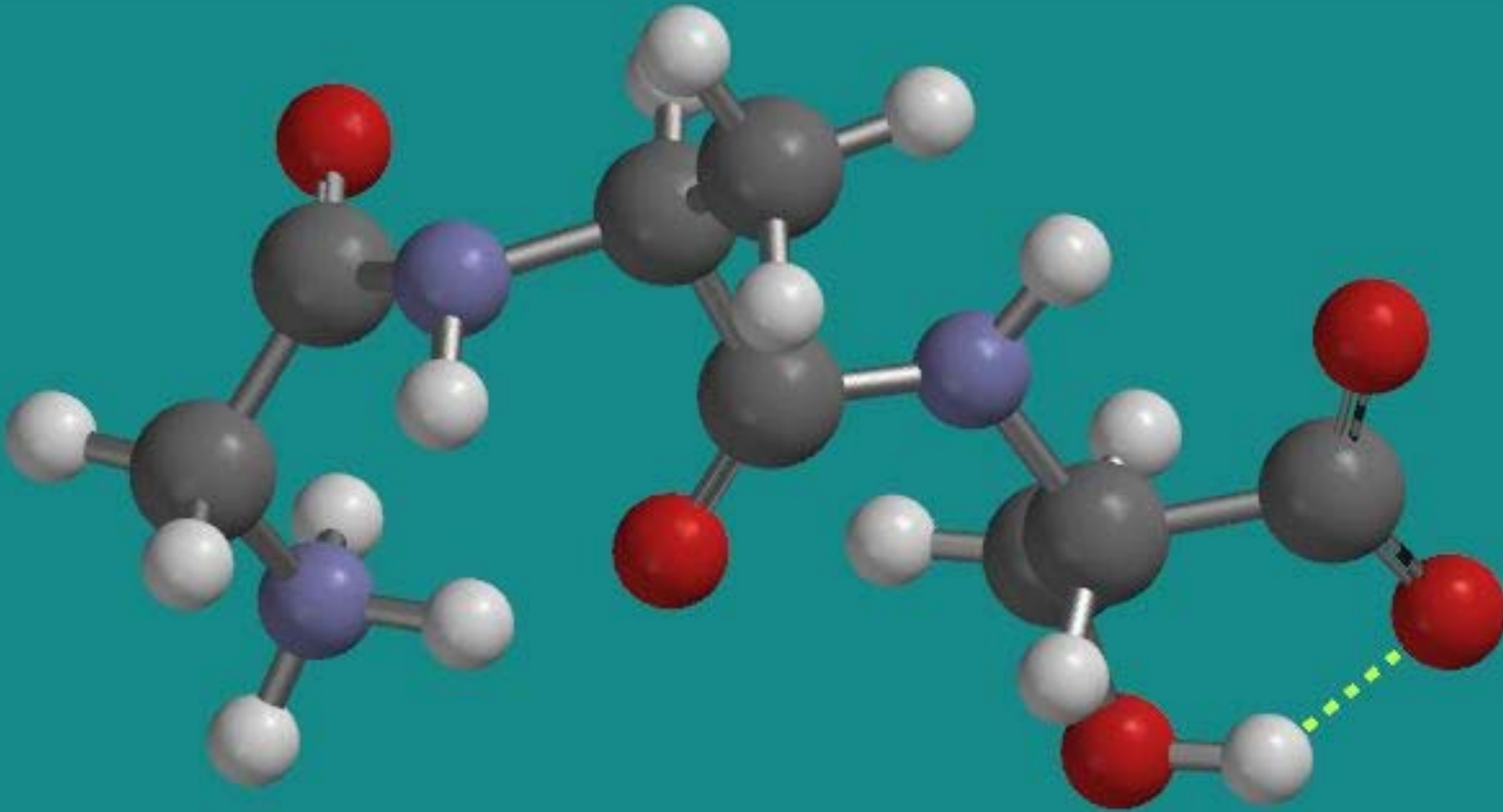
- *Primary Structure: the type and sequence of amino acids along the poly-peptide chain*



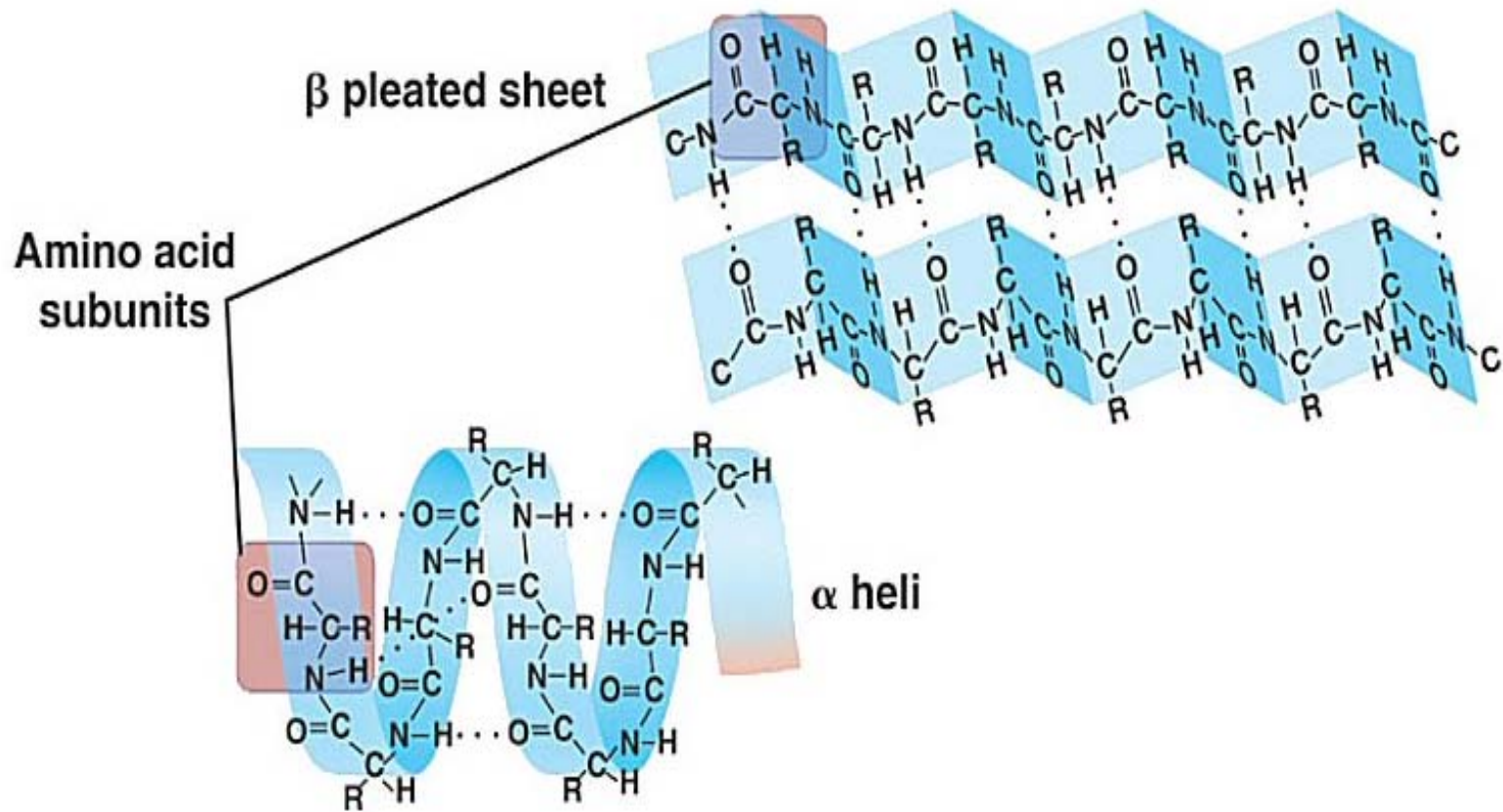
# *Secondary Structure*

- *Secondary Structure: “folding” of the protein via torsions about single bonds.*
- *Torsions mandated by minimization of repulsive forces and maximization of attractive forces, until a balance is reached that provides the lowest energy conformation for the protein as a whole.*

# *Secondary Structure*



# Secondary Structural Sub-units

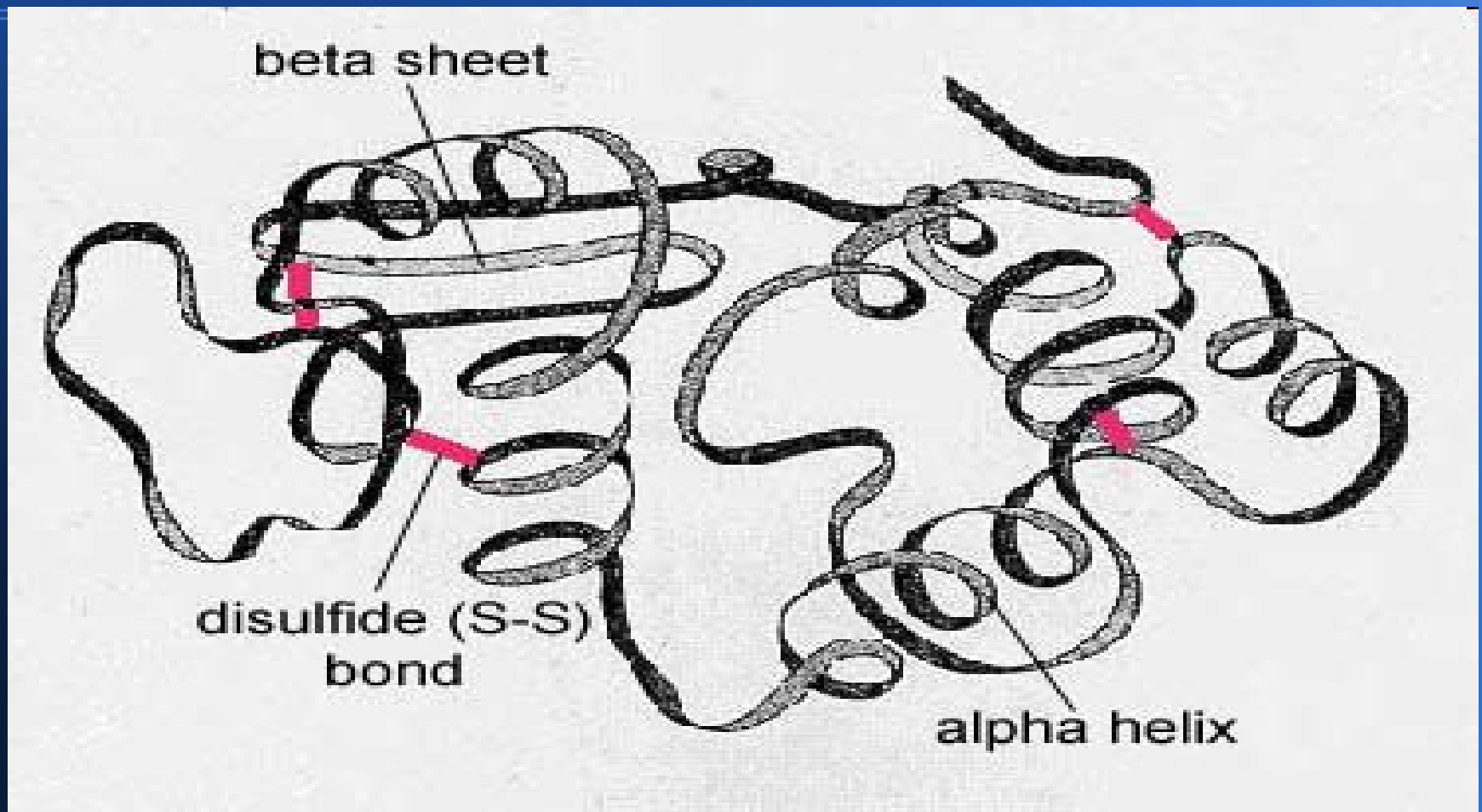


# *Tertiary Structure*

- *Tertiary Structure: the overall 3-D arrangement of all atoms in the protein.*
- *Formed by rearrangement of the subunits based again on attraction and repulsion.*
- *Responsible for function.*



# Tertiary Structure

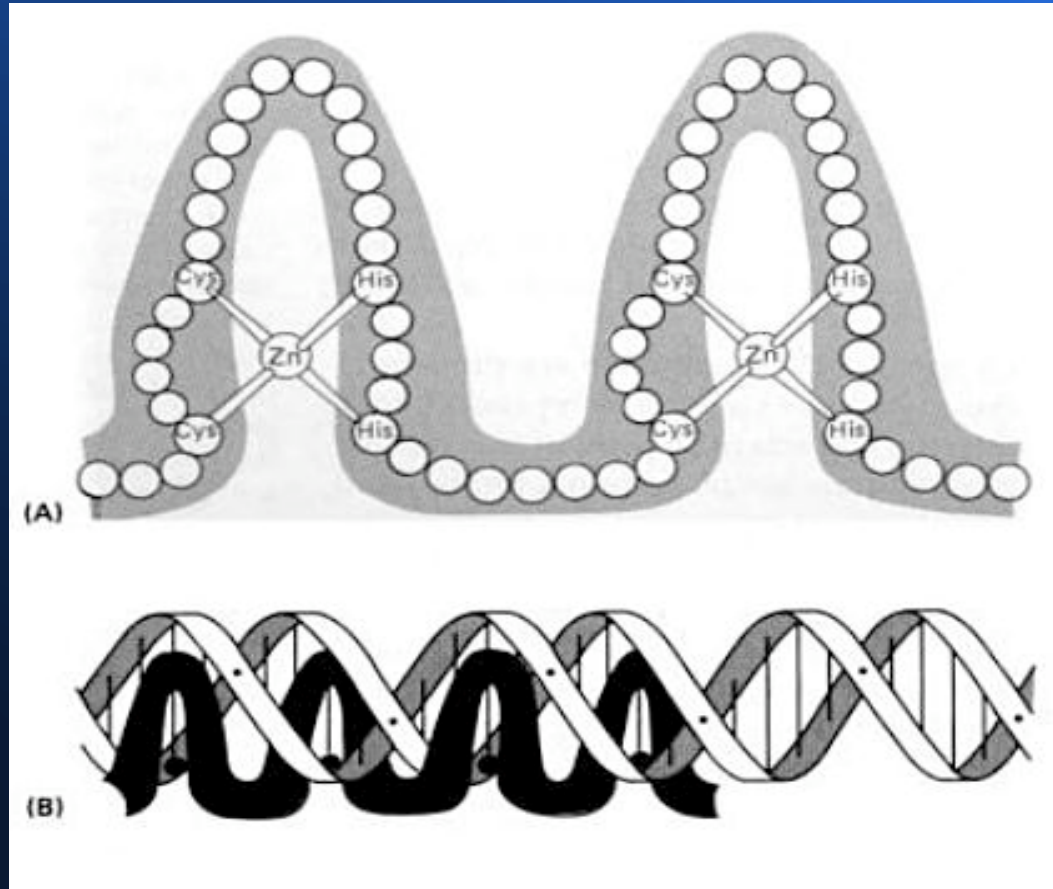


# Transcription Factor

*The Zinc Finger:*

General conformation of the DNA-binding

Domain 20-30 amino acids from the “finger



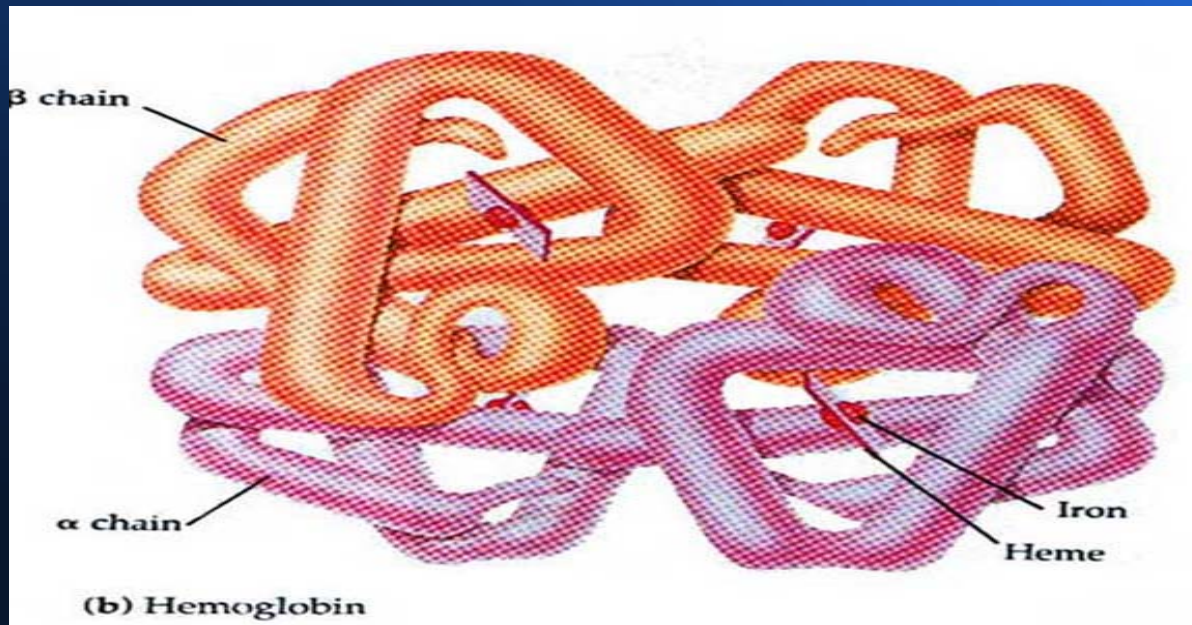
*Schematic view showing how four fingers bind to a specific DNA sequence*

*-form “recognition element”*



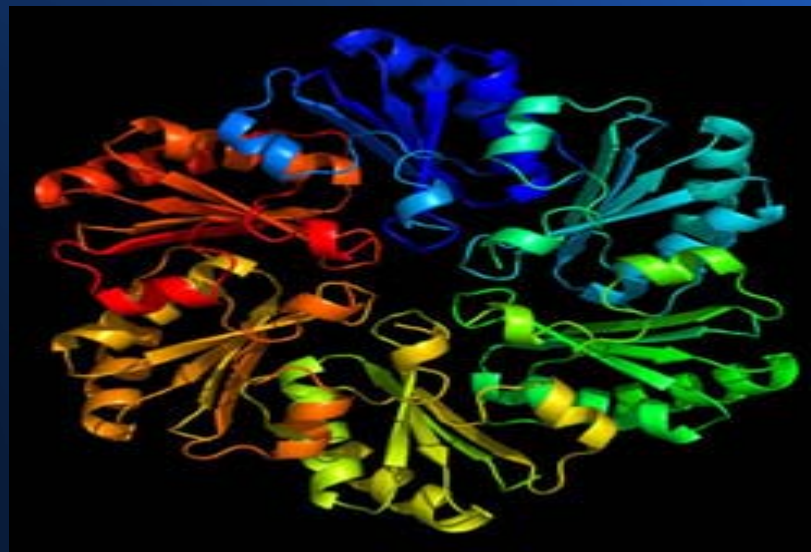
# Quaternary Structure

- *Quaternary Structure: the 3-D formed by the combination of multiple, separate polypeptide chains into a single complex*



# 3-D Structure Similarities

- *Prediction of function*
- *Realization of evolutionary details*
- *Structure better preserved than sequence*



# *3D Structure Analysis*

- *Large amount of 3D protein structures in the Protein Data Bank (PDB)*
- *Many different techniques to analyze them*
- *Generally, given two structures, find a rigid motion of one backbone onto the other such that large, contiguous regions of the backbones are matched.*

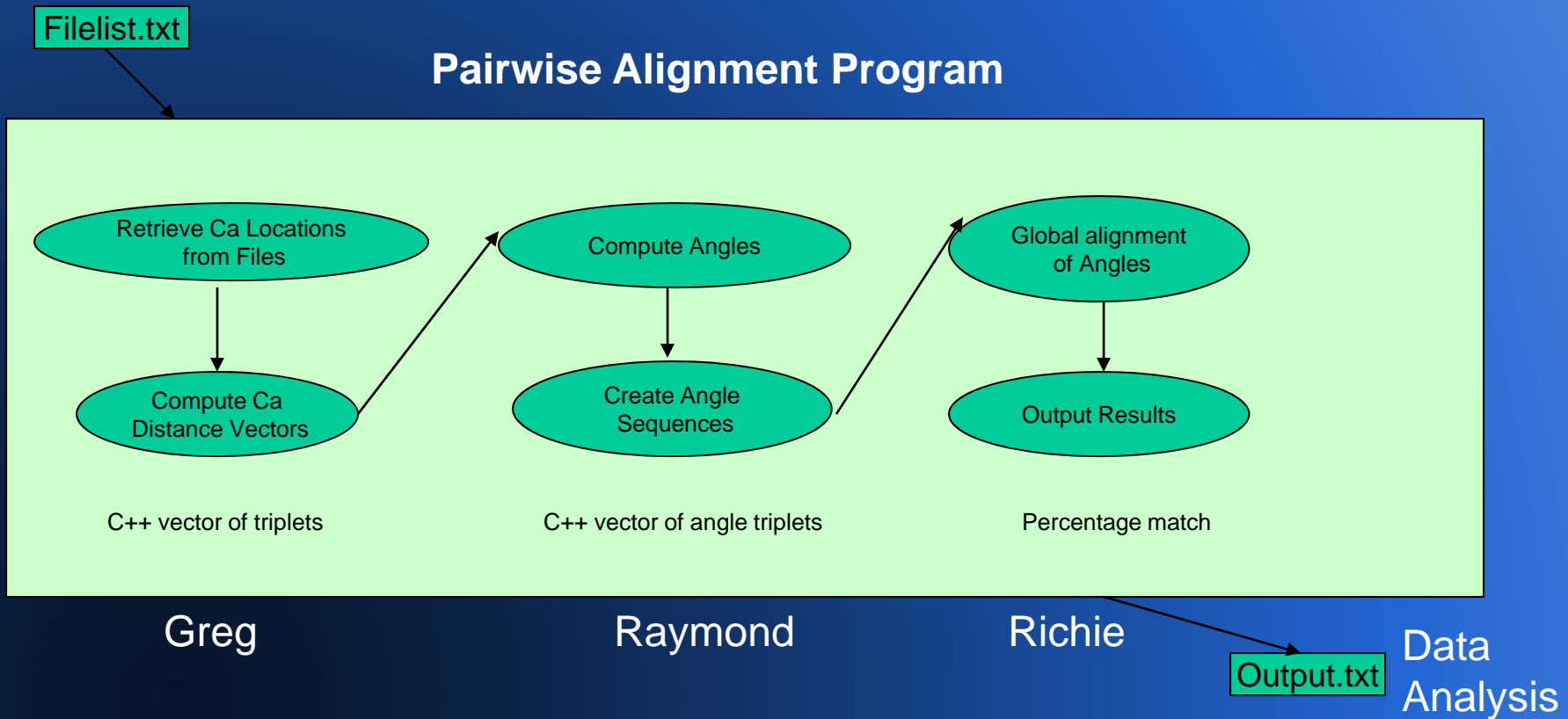
# Existing Techniques

- DALI – Uses distance matrices from C $\alpha$  backbones (align sub-matrices)
- LOCKE – Represents secondary structures as vectors (local alignment and DP)
- Chew et al. represent C $\alpha$  backbone as set of points on unit sphere (shift and combine)
- Others:
  - Combinatorial extension, geometric hashing, double dynamic programming

# *New Technique*

- *Focuses on representation of backbone as angles between  $C\alpha - C\alpha$  bonds*
- *Alignment is independent of locations in space*
- *Use dynamic programming to align triplet sequences*
- *Perfect alignment doesn't exist, so align substructures*

# Program Flow





# *Our Work*

- *Program - Three main sections:*
  - *Input processing (Greg)*
  - *Angle Calculation (Ray)*
  - *Dynamic Processing (Richie)*
- *Results*
  - *Similarity scores of proteins*
  - *Clustering of scores*

# *Program Input*

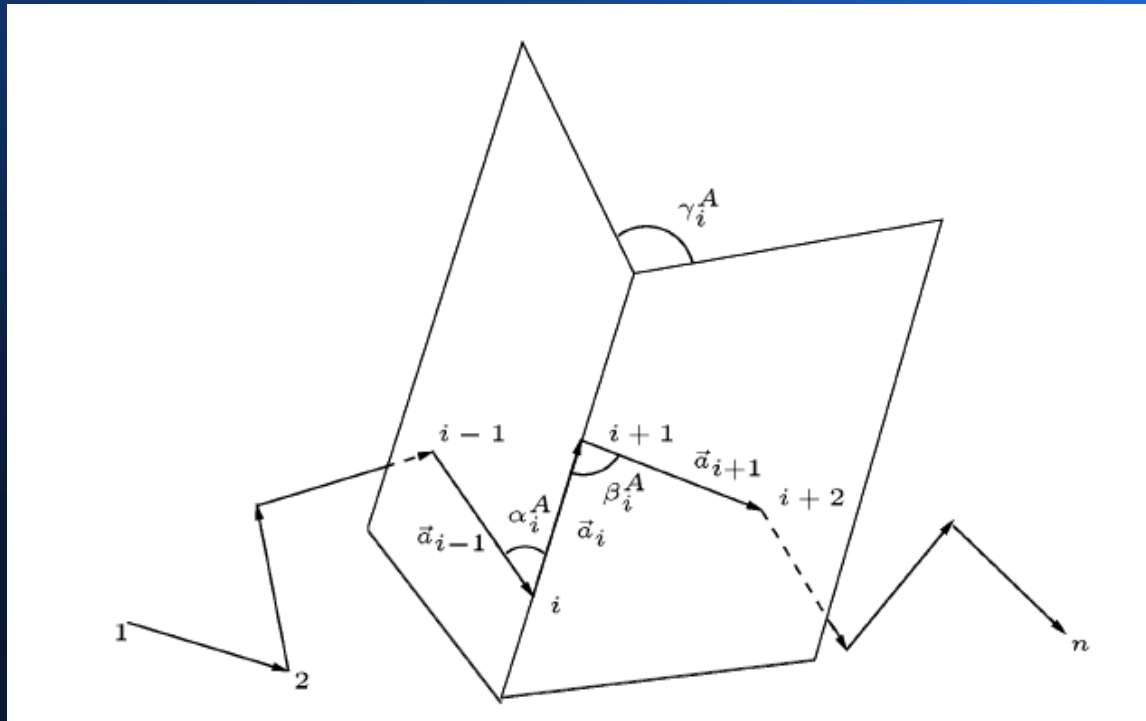
- *PDB Example file*
- *Process multiple files*
- *Convert to sequences of locations*

# *Vector Calculation*

- *Vectors from Ca-Ca needed*
- *PDB format provides locations*
- *Subtract one position  $a_{i-1}$  from  $a_i$  to achieve vector  $a_{i-1}$*
- *All vectors have a magnitude of  $\approx 3.8\text{\AA}$*

# Compute Local Geometric Representation

- Define vectors along C $\alpha$ -C $\alpha$  backbone
- Compute angles ( $\alpha_i$ ,  $\beta_i$ ,  $\gamma_i$ )



ComputeAngles( CaVectors )

for ( i ← 0 to #CaVectors )

$$\alpha_i \leftarrow \arccos \left[ \frac{-C\alpha\text{Vector}_{i-1} \bullet C\alpha\text{Vector}_i}{3.92^2} \right]$$

$$\beta_i \leftarrow \arccos \left[ \frac{-C\alpha\text{Vector}_i \bullet C\alpha\text{Vector}_{i+1}}{3.92^2} \right]$$

$$\text{normVector1} \leftarrow [-C\alpha\text{Vector}_{i-1} \times C\alpha\text{Vector}_i]$$

$$\text{normVector2} \leftarrow [-C\alpha\text{Vector}_i \times C\alpha\text{Vector}_{i+1}]$$

$$\text{normVector3} \leftarrow [-\text{normVector1} \times \text{normVector2}]$$

$$\gamma_i \leftarrow \arccos \left[ \frac{\text{normVector1} \bullet \text{normVector2}}{3.92^2} \right]$$

if ( normVector3 not in same direction as bondVector<sub>i</sub> )

$$\gamma_i \leftarrow 2\pi - \gamma_i$$

else

$$\gamma_i \leftarrow \gamma_i$$

angleTriplet<sub>i</sub> ← α<sub>i</sub> β<sub>i</sub> γ<sub>i</sub>

return 0

# Algorithm Details

- *Zero end gap penalty*
- *Internal Gap Penalty:  $h(k) = a + k*b$* 
  - *$k$ =gap length,  $a=0.2$ ,  $b=0.2$  ( $a,b$  found experimentally)*
- *Scoring function:  $S(A_i, B_i) = K - d(A_i, B_i)$* 
  - *$K=1.4$  (exp.),  $d(A_i, B_i)$  = Euclidean distance of  $A_i$  to  $B_i$*
  - *$d(A_i, B_i)$  is slightly different than expected*

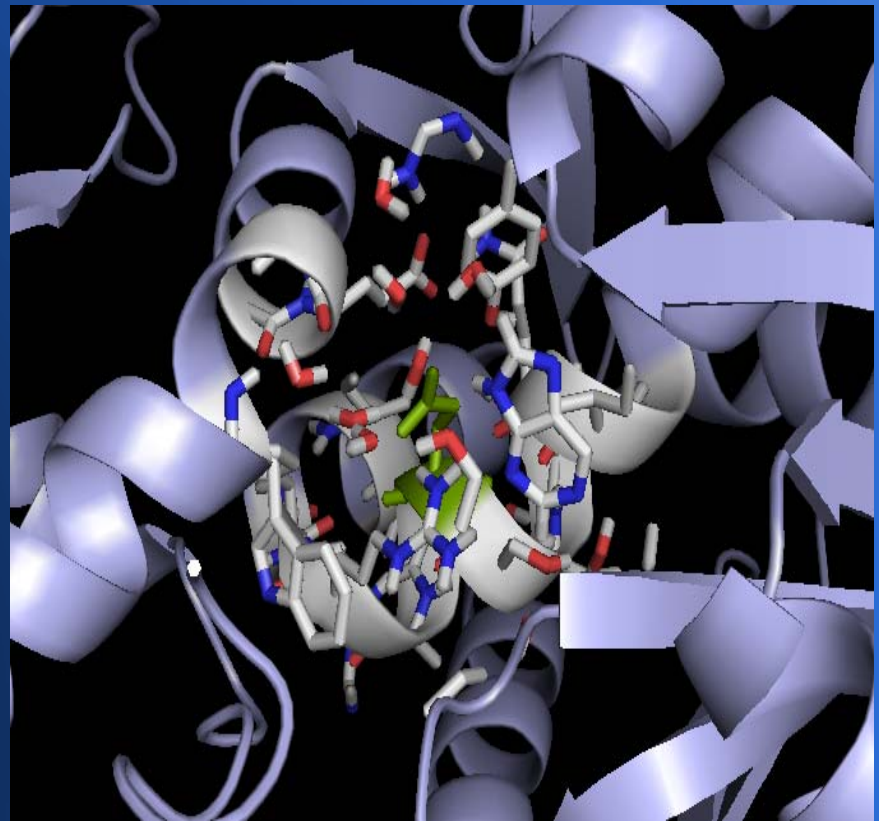
$$d(A_i, B_j) = \left( (\alpha_i^A - \alpha_j^B)^2 + (\beta_i^A - \beta_j^B)^2 + g(|\gamma_i^A - \gamma_j^B|)^2 \right)^{1/2}$$

$$g(x) = \min(2\pi - x, x)$$



# Cluster Results

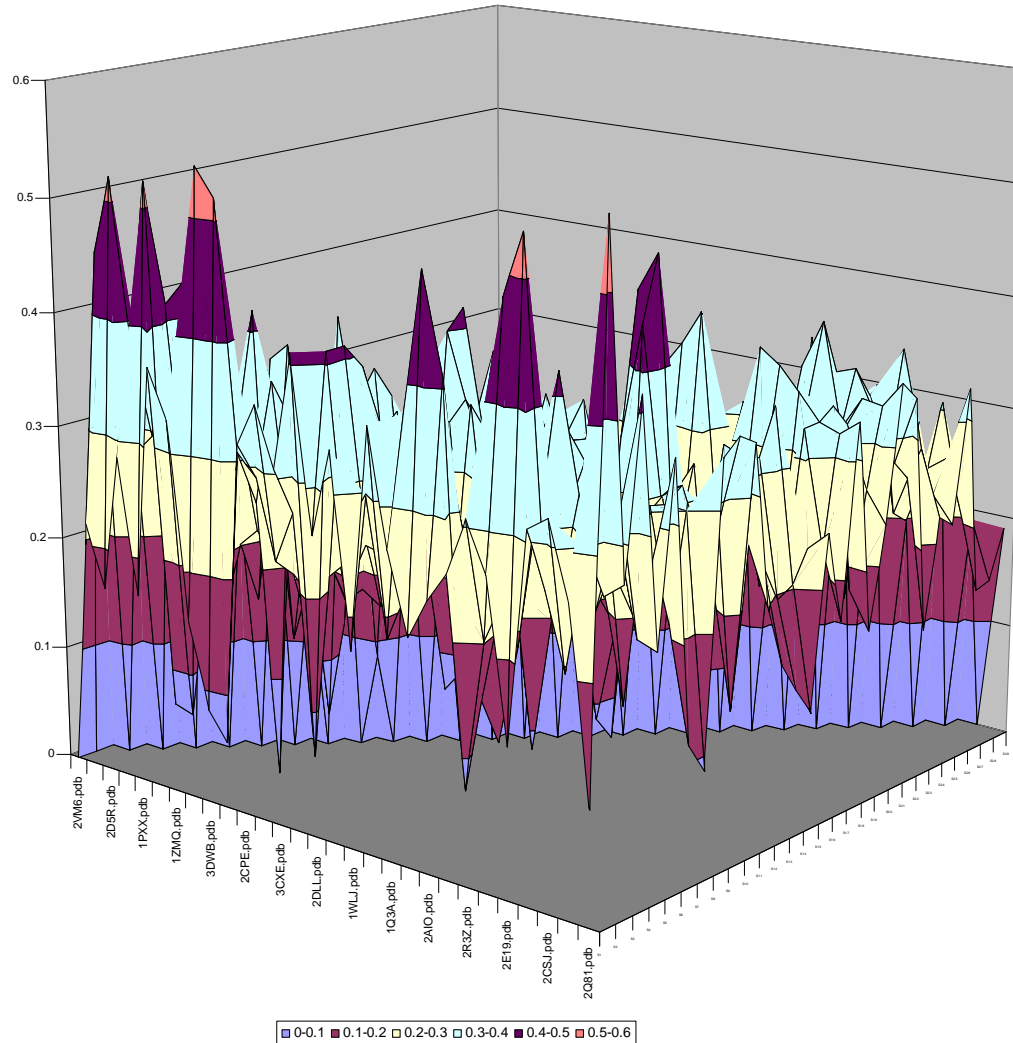
- *Use score from dynamic programming to determine protein similarity*
- *Can use percentage out of a “Max Score” to predict similar functionality*





# Results Graphed

Z – axis  
Percent  
Match



# 2VM6 vs 2Q81

Percentage Match: .56

Green: 2VM6

Red: 2Q81



# 2VM6 vs 2Q81

## 2VM6

### **Molecular Function**

- protein binding

### **Biological Process**

- apoptosis
- anti-apoptosis
- regulation of apoptosis
- protein binding
- microtubule binding
- in utero embryonic development
- kidney development
- induction of apoptosis
- cell-matrix adhesion
- activation of pro-apoptotic gene products
- post-embryonic development

## 2Q81

### **Molecular Function**

- nucleic acid binding
- DNA binding
- transcription factor activity
- protein binding
- zinc ion binding
- metal ion binding

### **Biological Process**

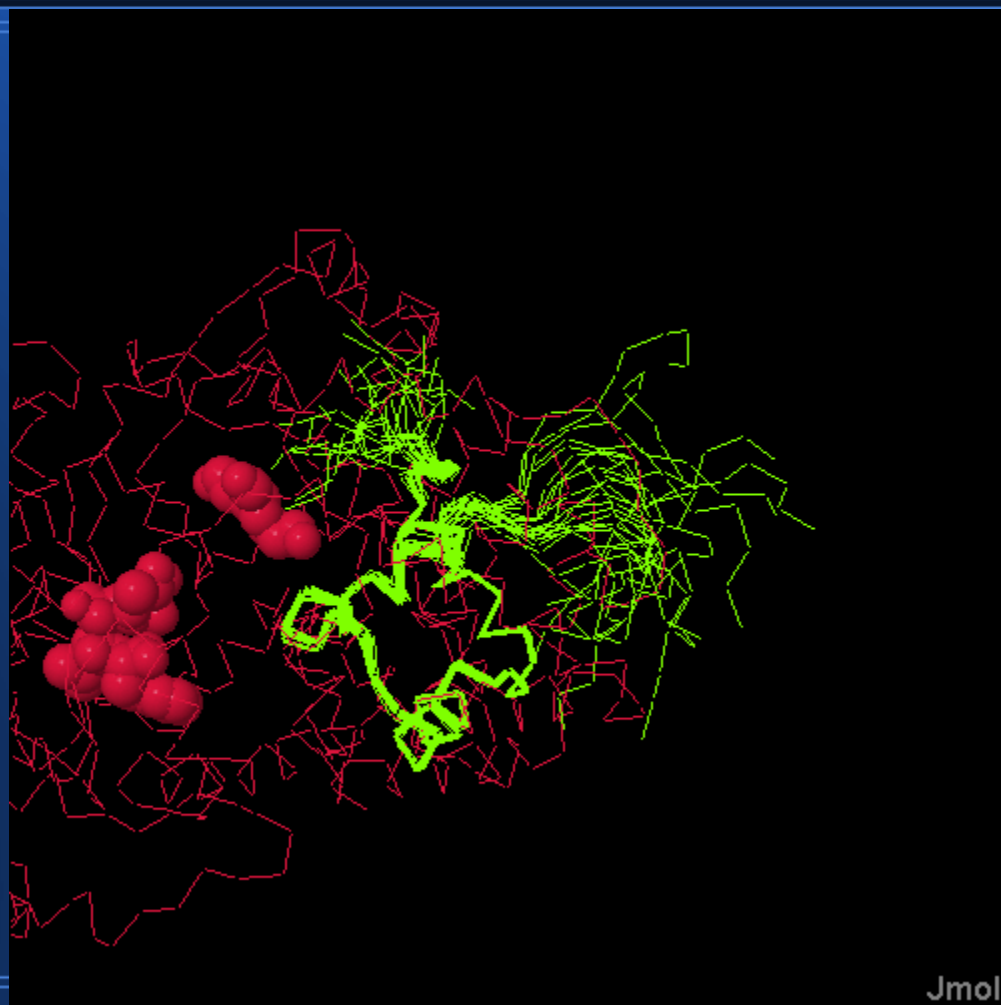
- gastrulation with mouth forming second
- transcription
- regulation of transcription, DNA-dependent
- multicellular organismal development
- ectoderm development
- negative regulation of cell cycle

# 2E19 vs 3DWB

Percentage Match: .49

Green: 2E19

Red: 3DWB





# 2E19 vs 3DWB

## 2E19

### **Molecular Function**

- nucleic acid binding
- DNA binding
- transcription factor activity
- transcription coactivator activity
- transcription corepressor activity
- zinc ion binding

### **Biological Process**

- negative regulation of transcription from RNA polymerase II promoter
- transcription
- regulation of transcription, DNA-dependent
- regulation of transcription from RNA polymerase II promoter
- immune response

## 3DWB

### **Molecular Function**

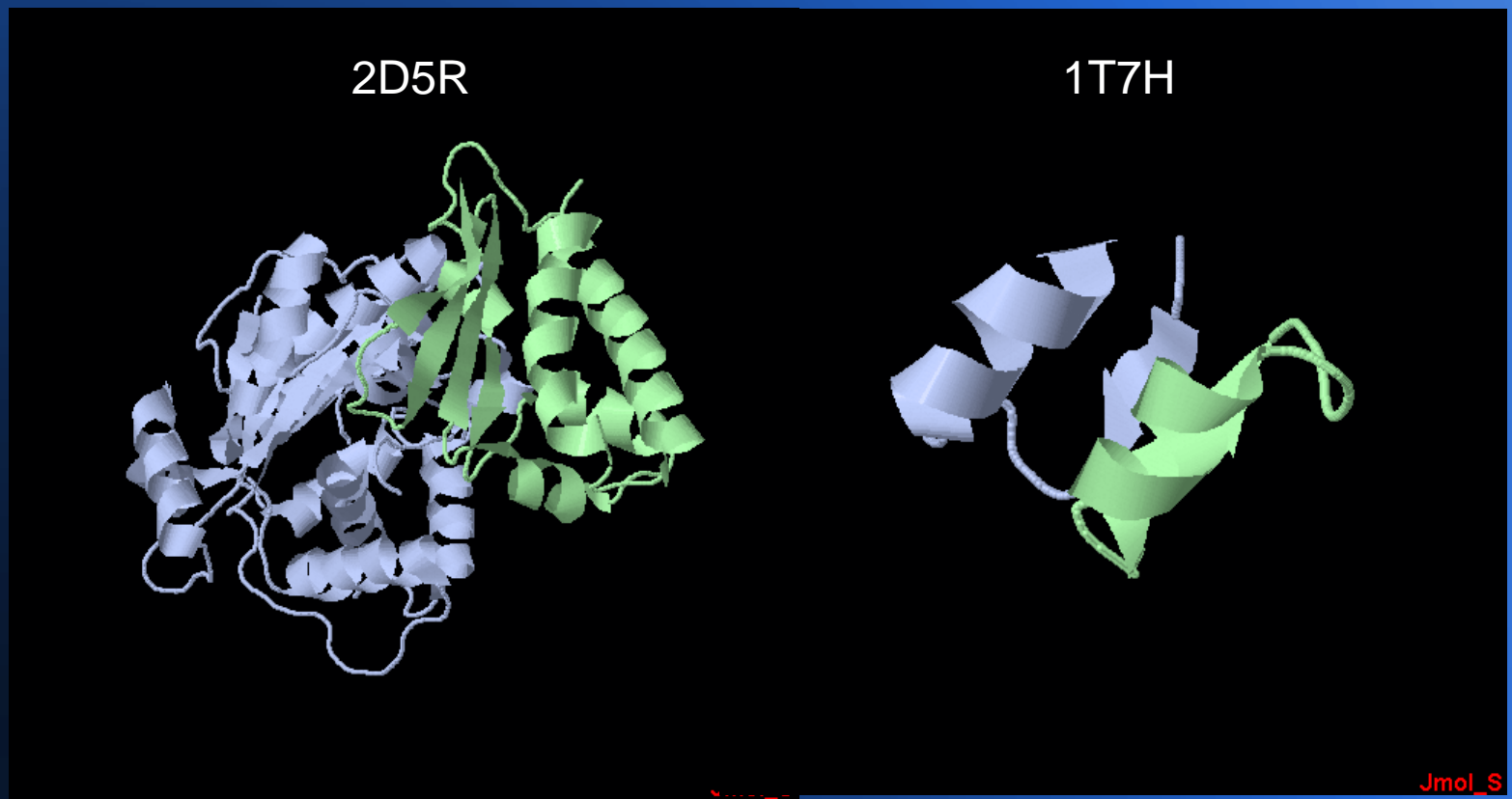
- endopeptidase activity
- metalloendopeptidase activity
- peptidase activity
- metallopeptidase activity
- zinc ion binding
- hydrolase activity
- peptide hormone binding
- protein homodimerization

### **Biological Process**

- response to hypoxia
- positive regulation of receptor recycling
- regulation of systemic arterial blood pressure by endothelin
- proteolysis
- apoptosis
- heart development

# 2D5R vs 1T7H

Percentage Match: .01



# 2D5R vs 1T7H

## 2D5R

### **Molecular Function**

- nucleic acid binding
- transcription factor activity
- signal transducer activity
- protein binding
- transcription activator activity

### **Biological Process**

- carbohydrate metabolic process
- transcription
- regulation of transcription, DNA-dependent
- signal transduction
- positive regulation of transcription from RNA polymerase II promoter

## 1T7H

### **Molecular Function**

- receptor binding
- hormone activity
- protein binding
- endothelin A receptor binding
- endothelin B receptor binding

### **Biological Process**

- skeletal system development
- patterning of blood vessels
- response to hypoxia
- in utero embryonic development
- regulation of systemic arterial blood pressure by endothelin
- regulation of pH
- cell surface receptor linked signal transduction