Protein Modeling

Regina Zibuck
rzibuck@wayne.edu

What is Protein Modeling?

- Developed by the Milwaukee School of Engineering Center for BioMolecular Modeling
- Understand basic features of protein structure
- Explore and manipulate protein structures using the online protein visualization webpages
- Create physical models using a foam covered wire called a Mini-Toober

The Competition (3 parts)

- Pre-build model
  - Students will bring their pre-built models to the assigned impound
- Computer Exploration of Protein Structure
  - Students will explore a new protein structure, guided by visualization of 3-D coordinates from the RCSB PDB, using Jmol/JSmol program and answer questions related to the structure
- Exam
  - Students will answer questions on a written test

Scoring

- Pre-build (40 %)
  - Rubric gets more detailed as the competition progresses
    - Invitational
    - Regional
    - States
    - Nationals
- On-site Computer Exploration (30 %)
- Written Test (30 %)

What should the team bring?

- Pre-build model, fit within 61.0 cm X 61.0 cm X 61.0 cm
- Judges will pick up and rotate your model
- ONE 4 by 6 inch note card explaining their additions to the model, in the form of a table headed: what is displayed, how it is displayed, why it is important

What each **STUDENT** should bring?

Three students on a team

- pens or pencils to write the exam
- ONE double-sided 8 ½ by 11 page of notes, may be in a sheet protector or laminated

What the event supervisor provides?

- For the On-site Computer Exploration
  - PDB file of a new protein
  - A computer connected to the internet

- The Exam

YOUR JOB:

Be familiar with protein structure. Learn about levels of protein structure - primary, secondary, tertiary, and quaternary structure

Properties of amino acid side chains – which ones are acidic vs basic and hydrophobic vs hydrophilic etc.

Interactions in protein structure – covalent (peptide bonds and disulfide bridges) and non-covalent (hydrogen bonds, hydrophobic interactions, charge based interactions), metal coordination.

Learn to visualize and explore protein structures.
Learn about the Protein Data Bank (www.rcsb.org)
Learn to use Jmol/JSmol to visualize structures from the PDB

Practice protein modeling

- Mark the toober (or other similar material) in 2 cm segments. Each of these segments denotes one amino acids in the protein. This corresponds to the primary structure.
- Visualize the different alpha helix and beta strand elements in Jmol/JSmol and fold corresponding regions of the toober or other material accordingly. This is the secondary structure.

Practice protein modeling

- Fold all secondary structural elements with respect to each other so that they come together to form the 3-D shape of the protein. This is the tertiary structure of the region being modeled.
- Identify functionally important structural features to answer questions and/or add creative enhancements to your model.

How to determine the functionally important structural features?

– Read the manuscript describing the structure, review article or Molecule of the Month feature provided. Pay attention to any specific amino acid residues, secondary structural elements etc. that are discussed in the context of the protein function.

How to determine the functionally important structural features?

- Visualize the full molecule at the RCSB PDB. Pay close attention to all polymer chains in the structure and how they interact with each other. Usually amino acid residues at the interaction interface have important roles in the function of the protein. Also pay attention to chemical principles in protein folding and stability – hydrophobic amino acids in the core of globular structures, polar and charged amino acids on the surface etc.

Start Here

http://cbm.msoe.edu/scienceOlympiad/participant1.php

http://scioly.org/wiki/index.php/Protein_Modeling

https://www.shop3dmoleculardesigns.com/Science-Olympiad-Protein-Modeling-Event-Kits-p/sopb.htm?1=1&CartID=0

The 2019-2020 Science Olympiad Modeling Event focuses on a modified CRISPR system that allows for correcting single base mutations in DNA. Teams will model a portion of the cytosine deaminase protein. This protein is used in conjunction with the CRISPR-Cas9 system to target a specific C-G base pair and convert it to a T-A base pair.

Science Olympiad

Participant Resources

1 - Overview of the Event

★ 2 - This Year's Theme

3 - Building Your Model

The Pre-build Visualization Environment

The Practice Visualization Environment

Supervisor Resources

Overview of the Event

★ Ordering Event Materials

Additional Resources for Event Supervisors and Coaches

Past Protein Modeling Events

Congratulations to the 2016 National Science Olympiad winners!

Science Olympiad Protein Modeling Event

Protein Modeling is an event developed by the CBM for Science Olympiad. This event challenges students to explore protein structure/function by creating 3-dimensional models of proteins using Mini-Toobers, foam-covered wire that will hold its shape once folded.

Join the 2019-20

Protein Modeling Event Now!

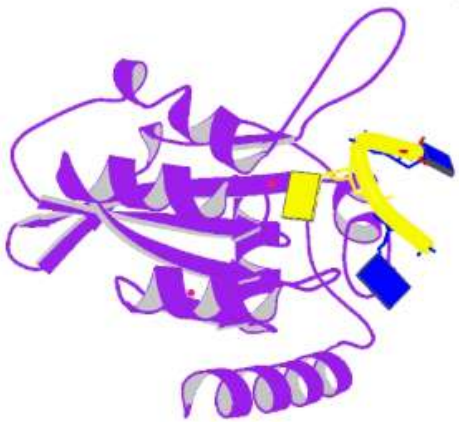
I am a Participant

I am an Event Supervisor

PDB for 5keg

Structure Summary | 3D View | Annotations | Sequence | Sequence Similarity | Structure Similarity | Experiment

Biological Assembly 1



3D View: Structure | Electron Density | Ligand Interaction

Standalone Viewers
Protein Workshop | Ligand Explorer

Global Symmetry: Asymmetric - C1

Display Files | Download Files

5KEG

Crystal structure of APOBEC3A in complex with a single-stranded DNA

DOI: 10.2210/pdb5KEG/pdb **NDB:** 5KEG

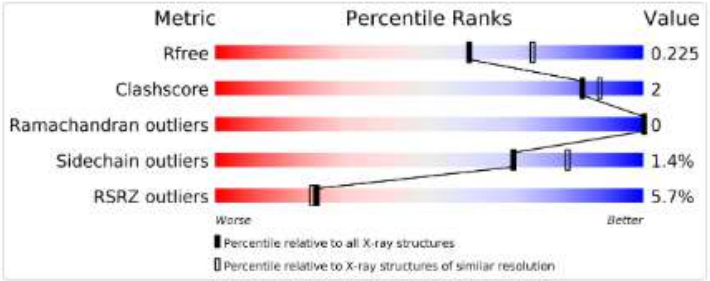
Classification: HYDROLASE/DNA
Organism(s): Homo sapiens, synthetic construct
Expression System: Escherichia coli
Mutation(s): 2

Deposited: 2016-06-09 **Released:** 2017-05-10
Deposition Author(s): Kouno, T., Hilbert, B.J., Silvas, T., Royer, W.E., Matsuo, H., Schiffer, C.A.
Funding Organization(s): National Institutes of Health/National Institute of General Medical Sciences

Experimental Data Snapshot

Method: X-RAY DIFFRACTION
Resolution: 2.2 Å
R-Value Free: 0.225
R-Value Work: 0.177

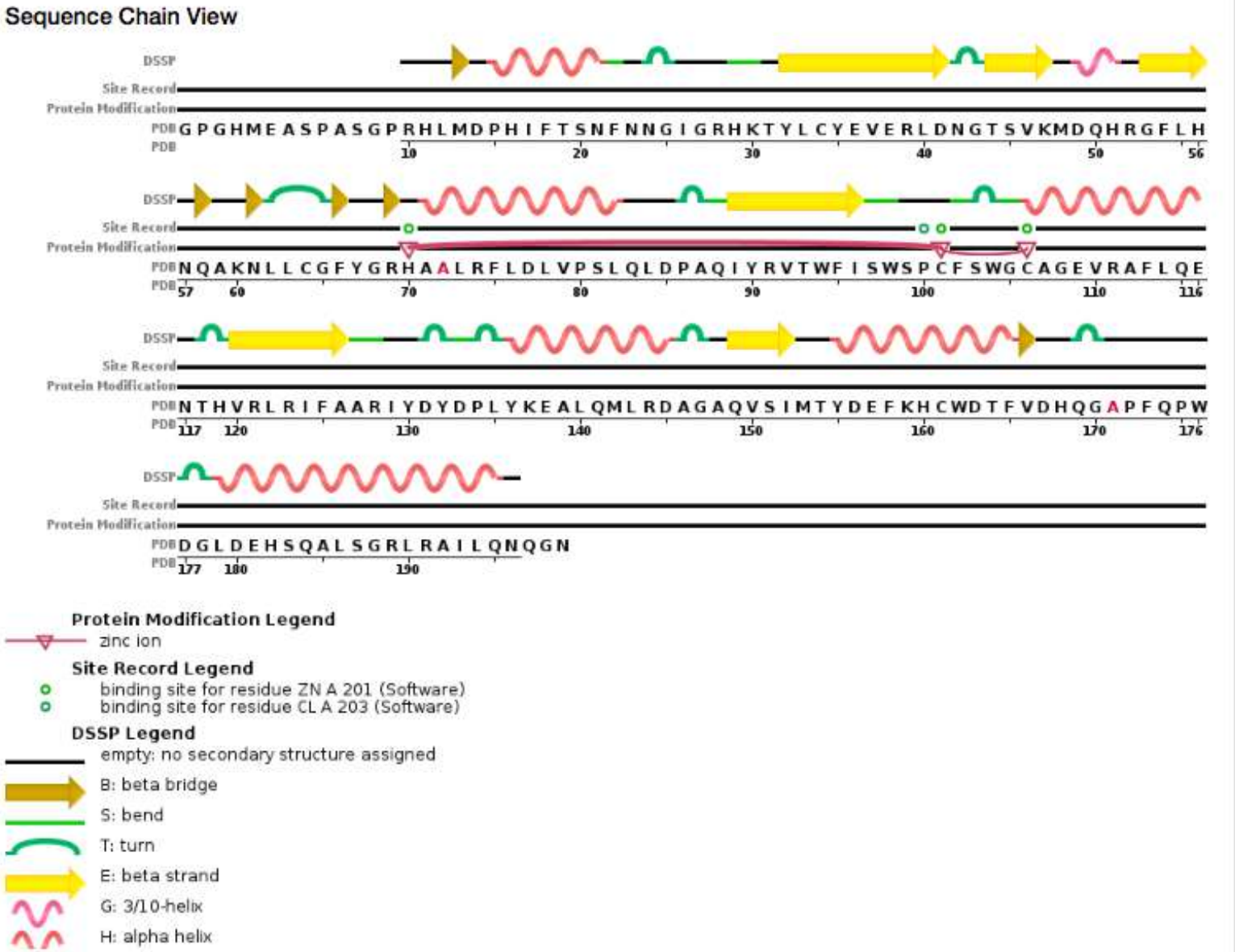
wwPDB Validation 3D Report | Full Report



www.rcsb.org

- https://pdb101.rcsb.org/learn/guide-to-understanding-pdb-data/primary-sequences-and-the-pdb-format

Chain A residues 31-145



Background

- Fun posters and videos

- http://www.rcsb.org/pdb/101/static101.do?p=education_discussion/educational_resources/index.html

2019-20 Pre-Build Model

Based on *5keg.pdb*

Directions:

The 2019-20 Pre-Build Model should represent **amino acids 31 - 145 of chain A** of the cytidine deaminase protein based on the PDB file 5keg.pdb. This section will require a toober that is 228cm long at the scale of 1 amino acid = 2cm.

★★ NOTE: This is a different structure/sequence than originally stated in the Science Olympiad rules. An official rule change has been submitted. The correct structure/sequence is amino acids 31-145 of chain A of 5keg.pdb★★

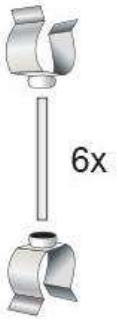
The pre-build model is worth 40% of a team's final score for the event. The pre-build model will be created using a purchased "pre-build" Mini-Toober kit from *www.3dmoleculardesigns.com*, or with found materials of the participant's choosing such as *KwikTwist* tie-down ropes. The same pre-build model will be used at each level of the competition (invitational, regional and state) and participants will take their pre-build model home after each event.

Resources:

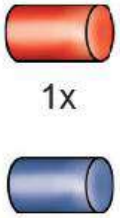
Jmol Quick Reference Sheet - A Jmol Quick Reference Sheet that highlights key commands and useful resources for designing protein models in Jmol on one side and a clear chart of the 20 amino acids on the other side.

To learn Jmol and how to use this competition environment, visit *Section 3 - Building Your Physical Protein Models* of the MSOE Center for BioMolecular Modeling's Protein Modeling Event participant resources.

Parts List:



Support Posts

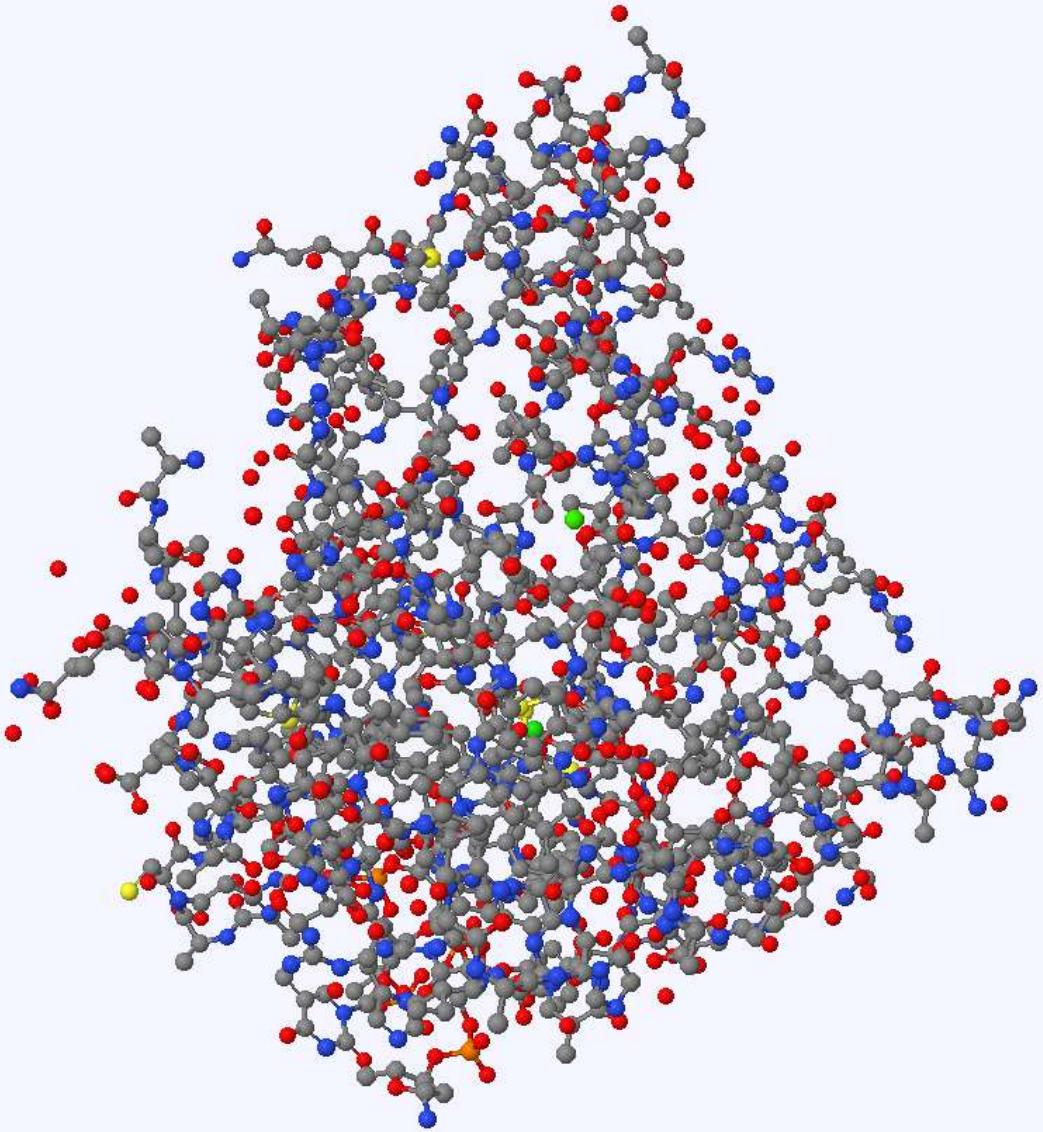


Endcaps

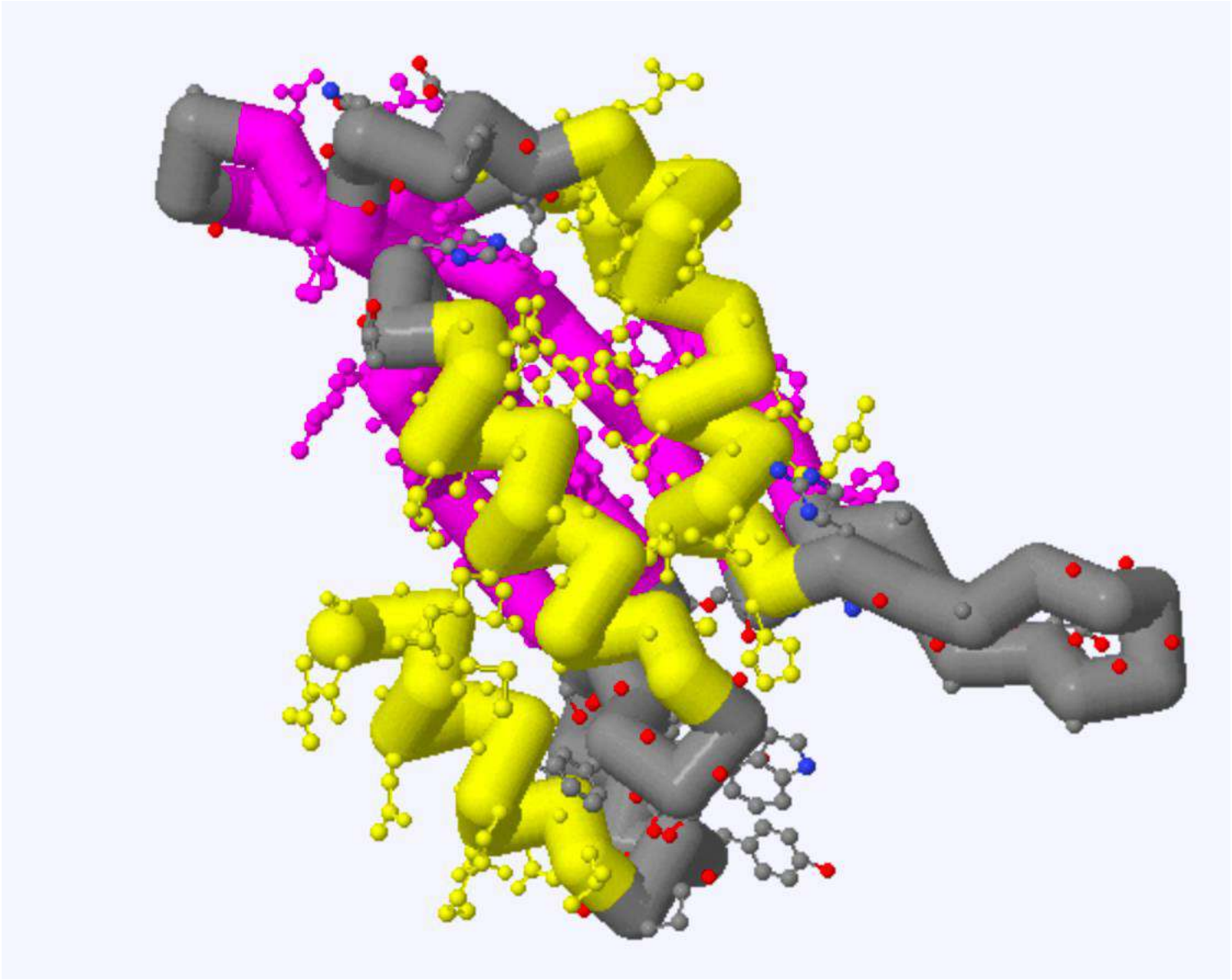
228 cm

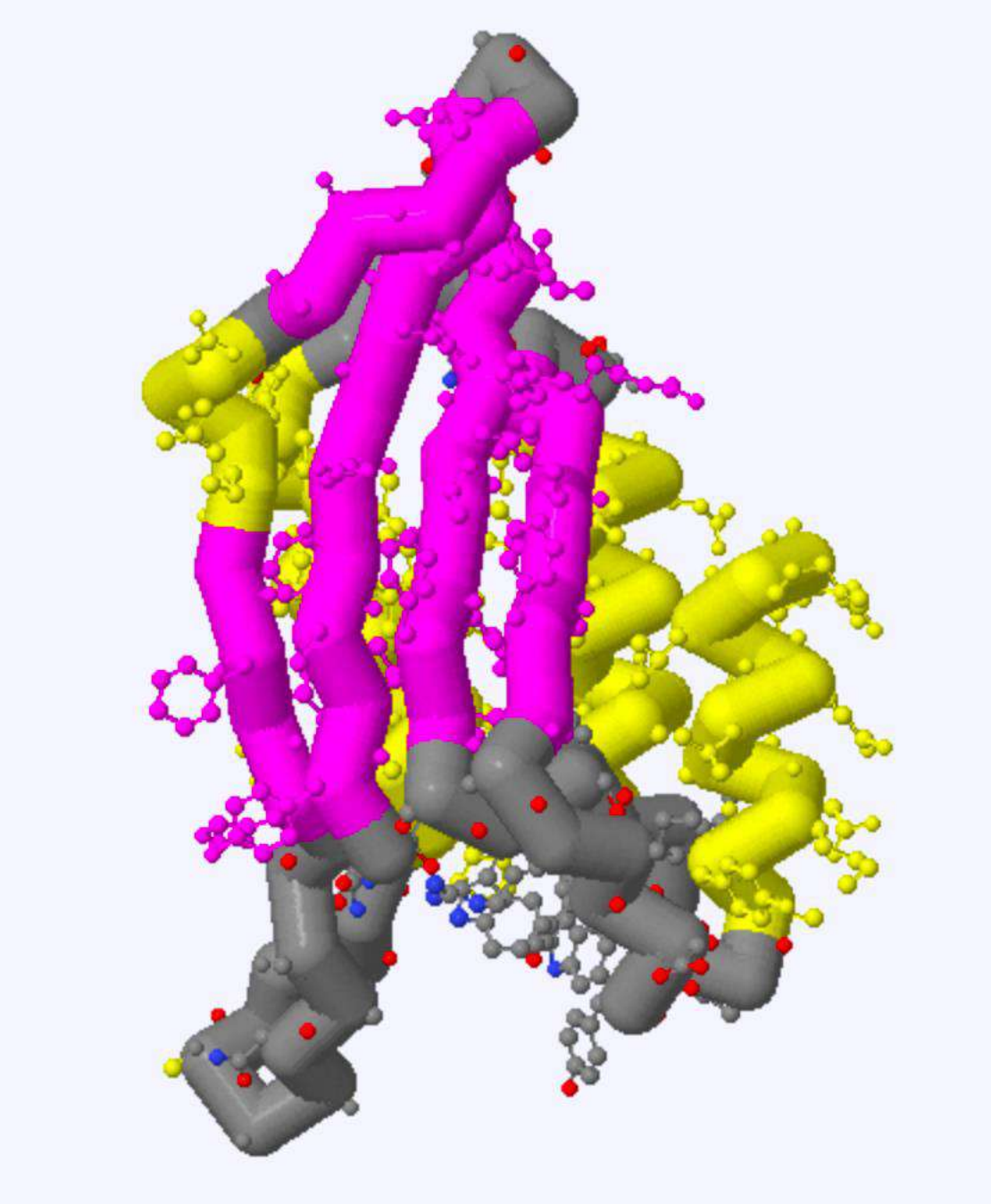
Mini Toober

Use amino acids 31-145 of chain A



Something like this?





What do you see?

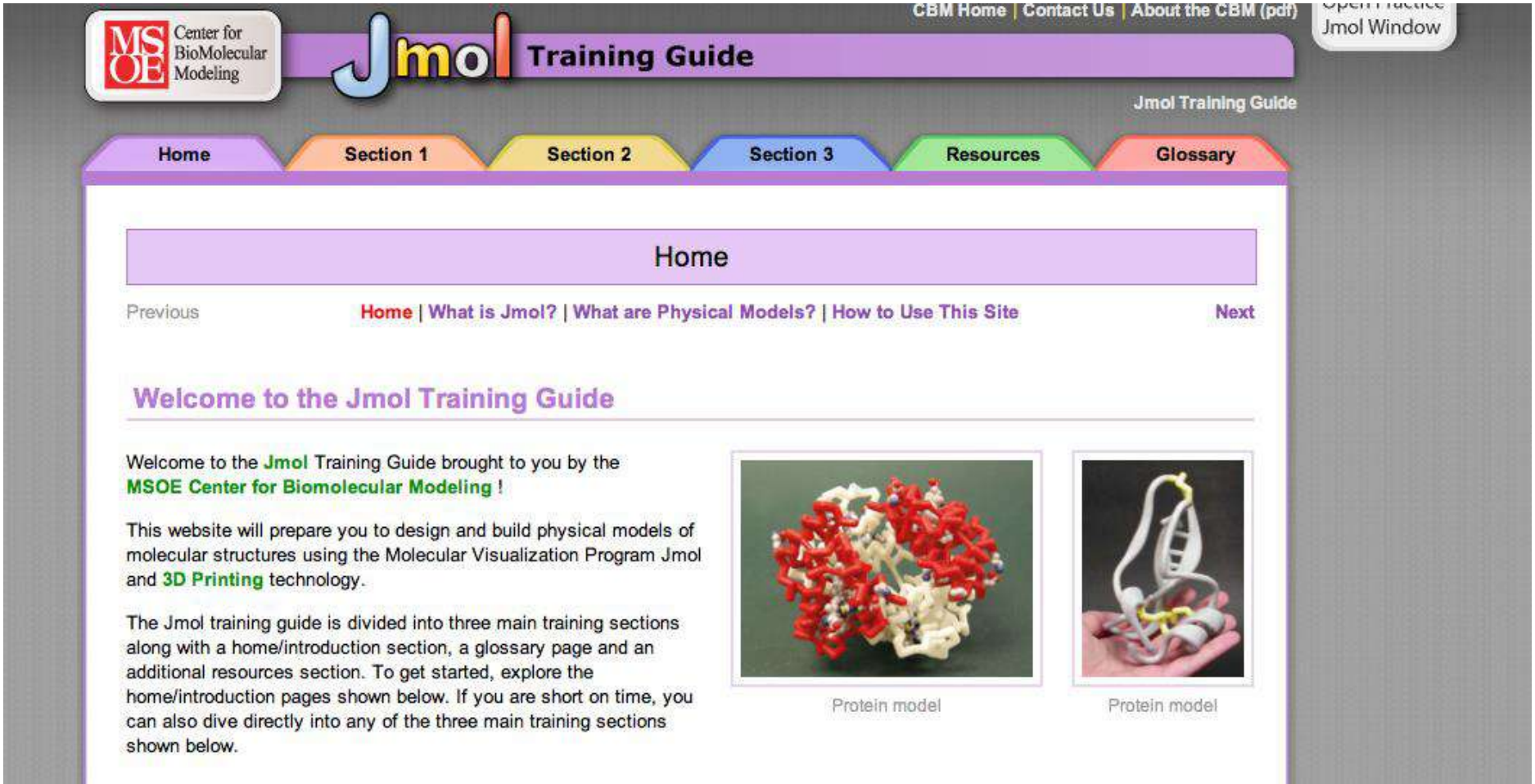
- There are 4 Alpha helixes
- One of those is a short 3/10 helix
- The other 3 α-helixes are about the same size
- There are 6 beta sheets
- There are 2 beta bridges
- There are 7 turns
- There 3 S bends
- Can you tell where each structure starts and ends?

Jmol Training Guide

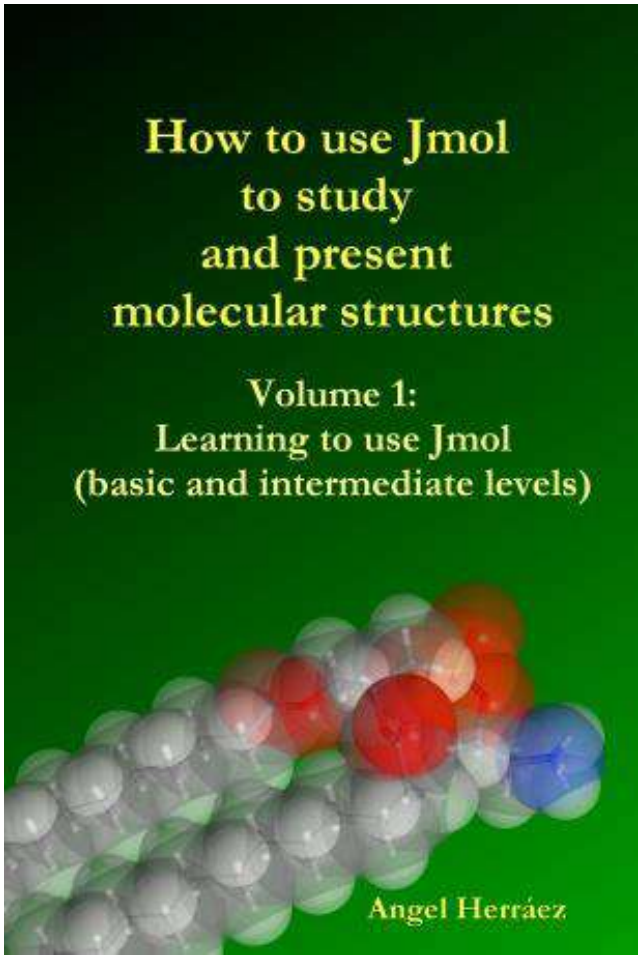
- http://cbm.msoe.edu/includes/pdf/crest/jmoltrainingall.pdf
- A free manual you can download

- Tutorial: http://cbm.msoe.edu/teachingResources/jmol/jmolTraining/started.html

http://cbm.msoe.edu/markMyweb/jmolWebsite/index.html



Jmol Book


How to use Jmol
to study
and present
molecular structures
Volume 1:
Learning to use Jmol
(basic and intermediate levels)
Angel Herráez

OnSite Jmol Activity

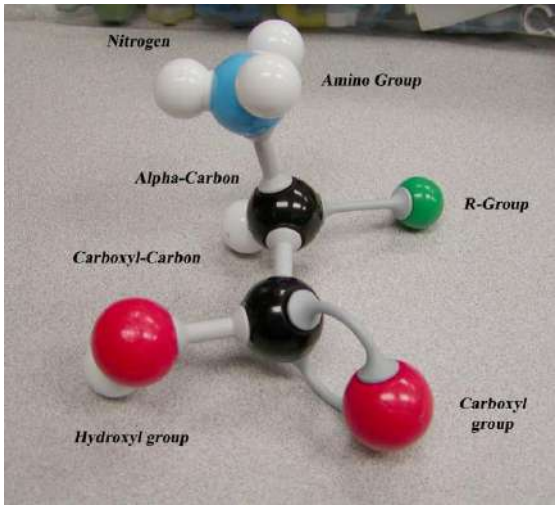
- Can you manipulate the JMOL/SJMOL program to be able to answer questions about the bonding in the prebuild molecule?
- Can you point out some of the areas that are hydrophobic in the molecule?
- Can you point out some of the areas that are hydrophilic?
- Which are the important attach points?
- Where are the alpha helixes?
- Where are the beta sheets?
- What holds the structure of the molecule together?

This is a Chemistry Event

- Know the chemistry of amino acids
- Know structure
- Know how to draw amino acids
- Know the bonding involved
- Know how to make an amide (peptide) bond
- Condensation reaction with water as the by-product

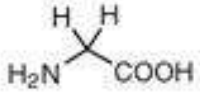
Background info about 20 amino acids

- Backbone consist of:
  - Amino group (NH_2 or NH_3^+)
  - Carbon atom, where the side chain is bound
  - Carboxyl group (COO^- or COOH)

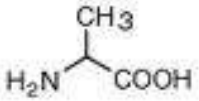


- Side chains are:
  - Hydrophobic
    - Have only carbon and hydrogen atoms
    - Usually buried inside the protein
    - Non-polar or apolar
  - Hydrophilic
    - have hydroxyl, carboxylic acid, or amine groups
    - Are generally on the outside of the protein
    - May be acidic or basic, or polar

Small

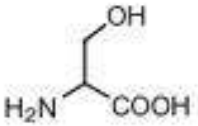


Glycine (Gly, G)
MW: 57.05

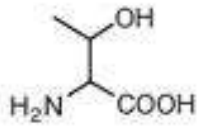


Alanine (Ala, A)
MW: 71.09

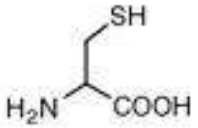
Nucleophilic



Serine (Ser, S)
MW: 87.08, pK_a ~ 16

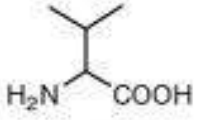


Threonine (Thr, T)
MW: 101.11, pK_a ~ 16

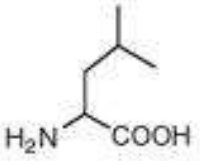


Cysteine (Cys, C)
MW: 103.15, pK_a = 8.35

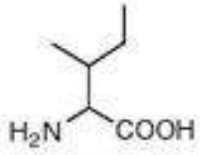
Hydrophobic



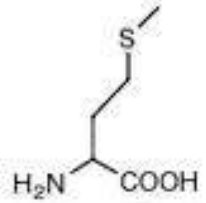
Valine (Val, V)
MW: 99.14



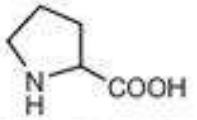
Leucine (Leu, L)
MW: 113.16



Isoleucine (Ile, I)
MW: 113.16

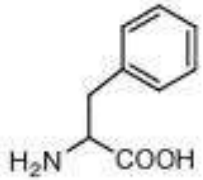


Methionine (Met, M)
MW: 131.19

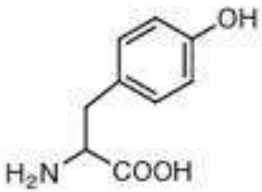


Proline (Pro, P)
MW: 97.12

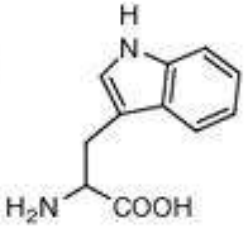
Aromatic



Phenylalanine (Phe, F)
MW: 147.18

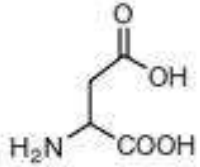


Tyrosine (Tyr, Y)
MW: 163.18

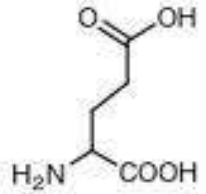


Tryptophan (Trp, W)
MW: 186.21

Acidic

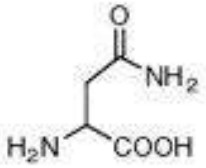


Aspartic Acid (Asp, D)
MW: 115.09, pK_a = 3.9

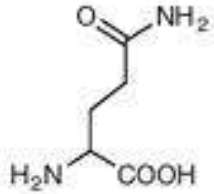


Glutamic Acid (Glu, E)
MW: 129.12, pK_a = 4.07

Amide

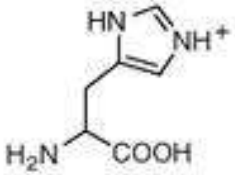


Asparagine (Asn, N)
MW: 114.11

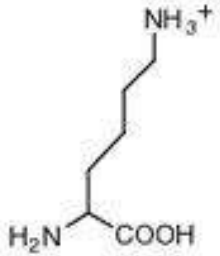


Glutamine (Gln, Q)
MW: 128.14

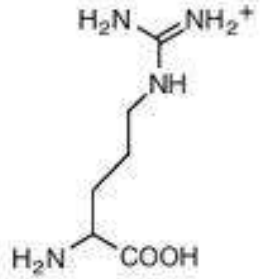
Basic



Histidine (His, H)
MW: 137.14, pK_a = 6.04



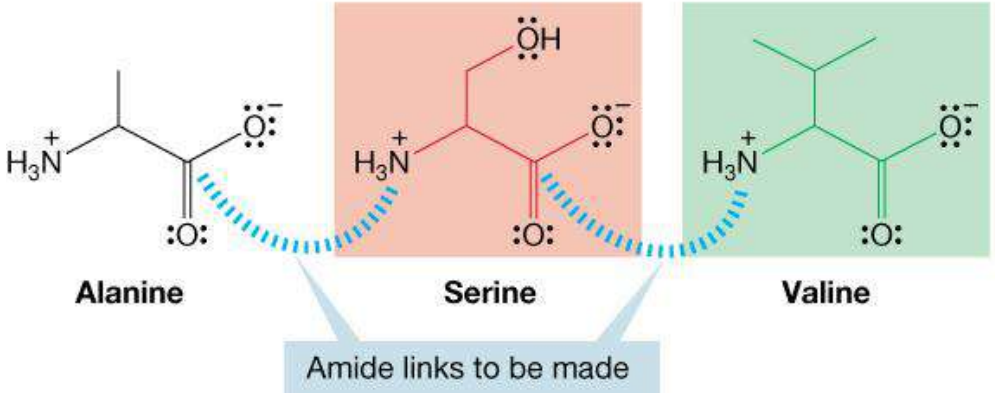
Lysine (Lys, K)
MW: 128.17, pK_a = 10.79

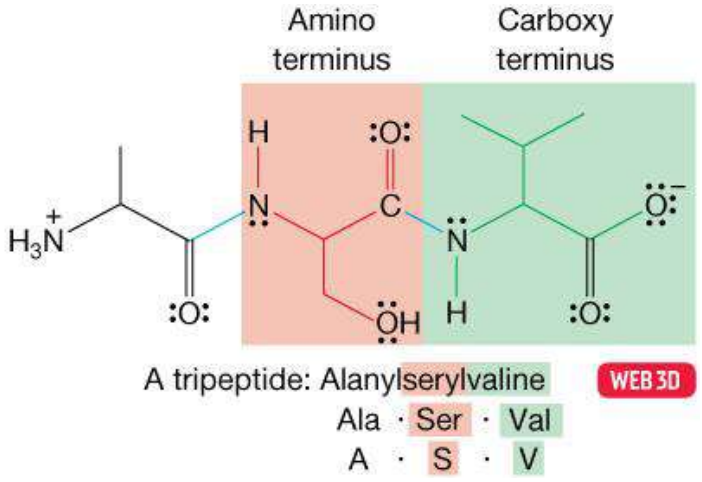


Arginine (Arg, R)
MW: 156.19, pK_a = 12.48

Peptide Chemistry

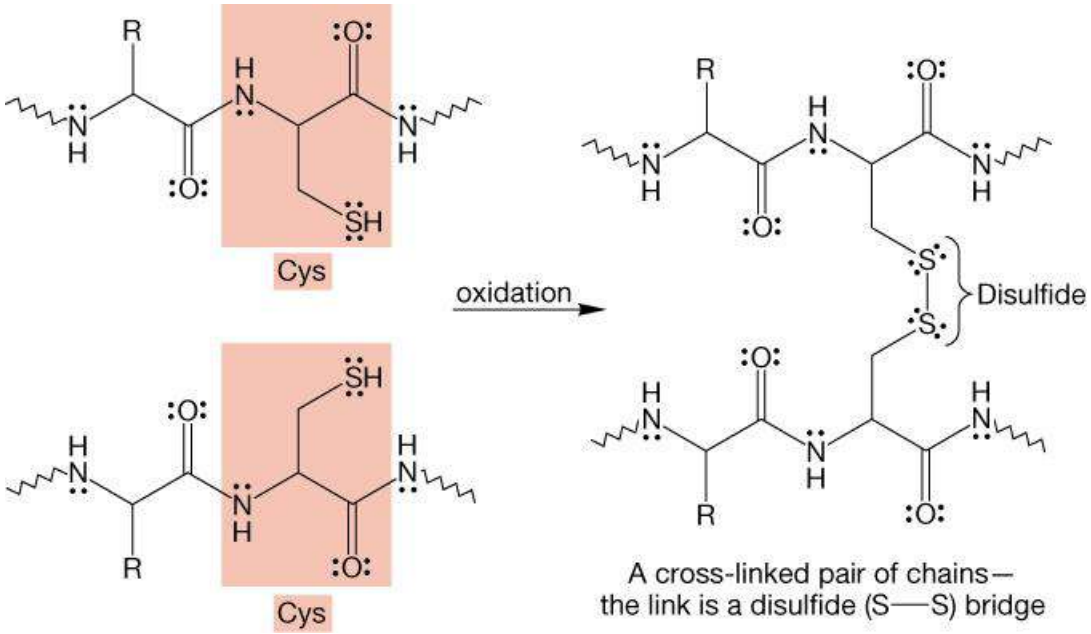
- Peptides usually written with the **amino terminus** on the left, proceeding toward the **carboxy terminus** on the right





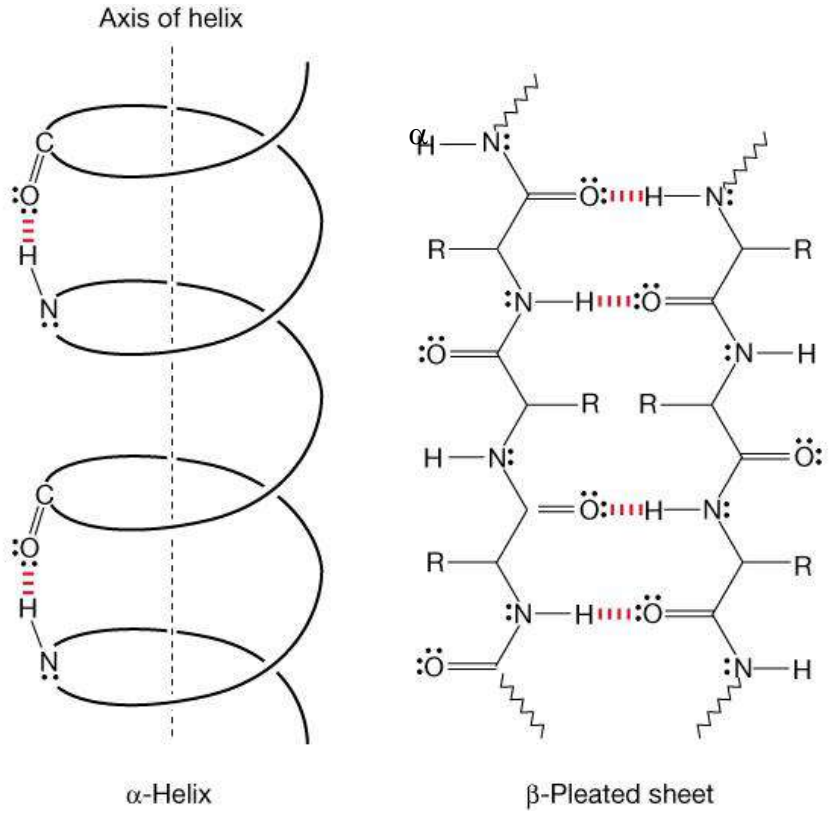
Primary Peptide Structures

- The sequence of amino acids in a peptide or protein gives the primary structure
- Connections between chains or within the same chain are often formed through disulfide bridges



Secondary Peptide Structure

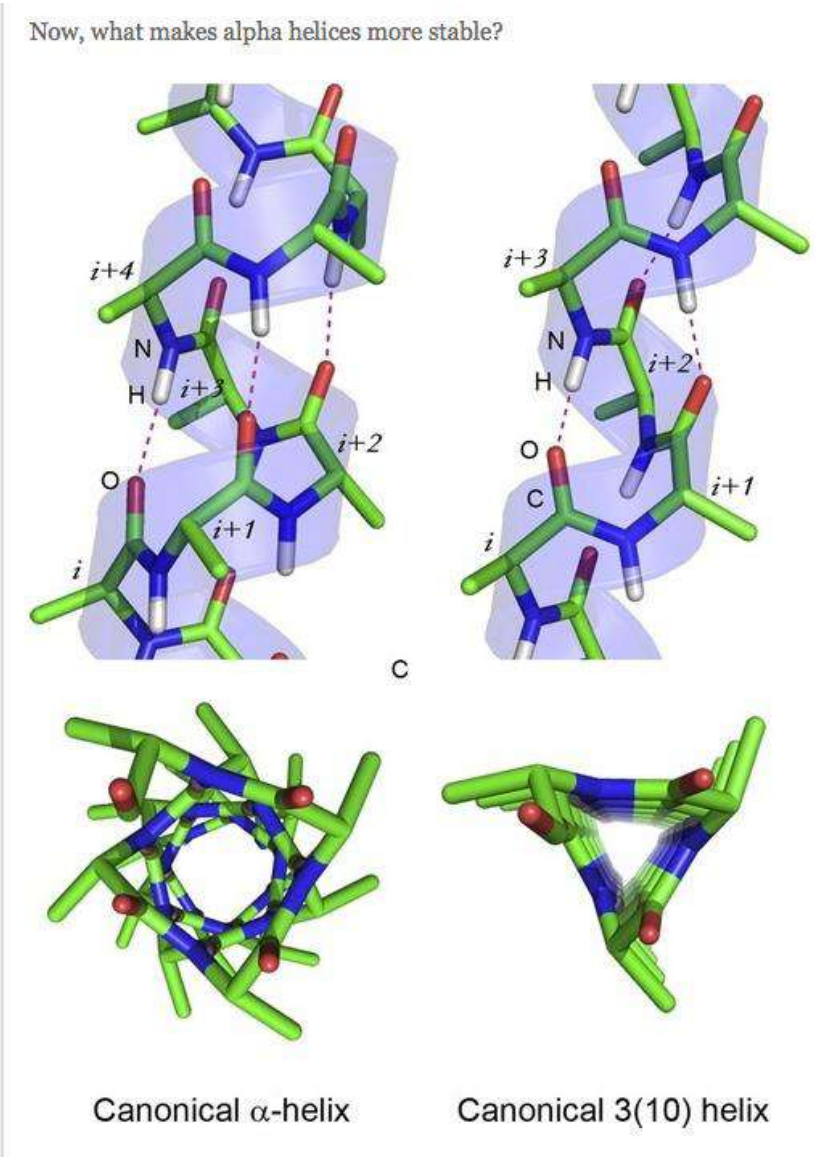
- α-helix, β-pleated sheets, or random coil of amino acids gives the secondary structure



Secondary Structure of a Protein

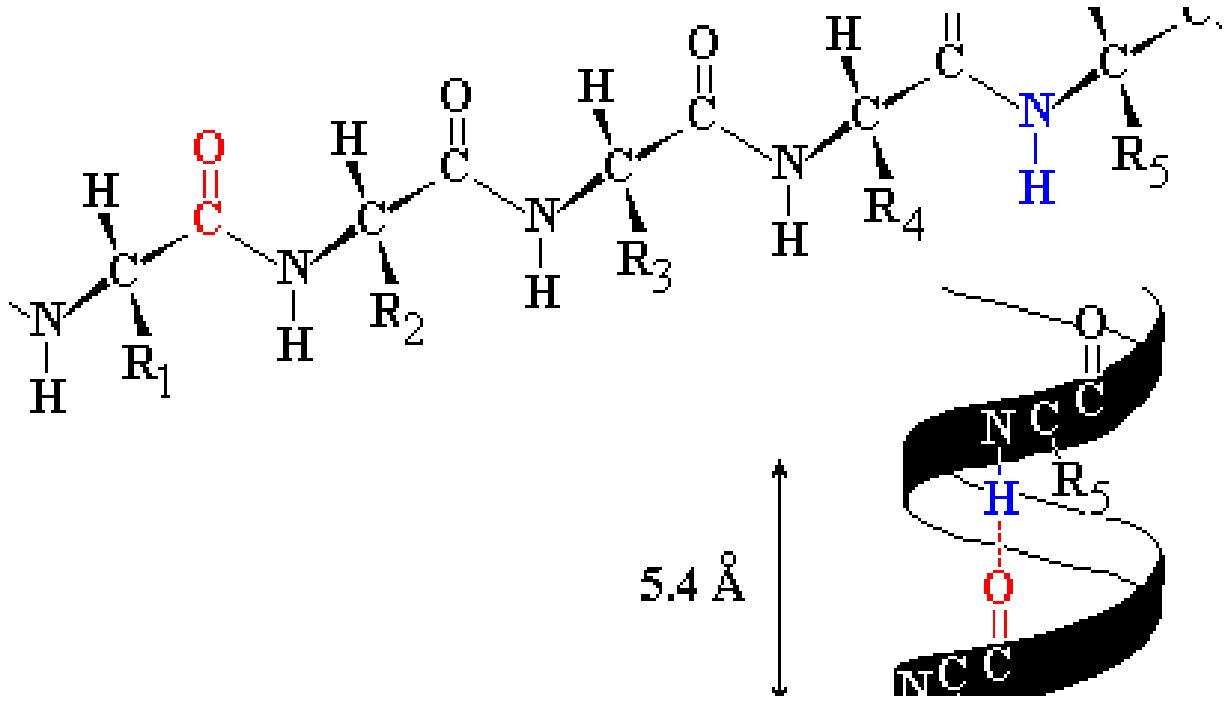
- The 2 most common types of secondary structure are the alpha helix and beta pleat
- The alpha helix ONLY coils right-handed (if you are going up the stairs, your right hand rests on the outside banister going up)
- The beta pleat should bend back and forth in a zigzag pattern of about 20° at each start of a new amino acid

α-helix or 3_{10} helix



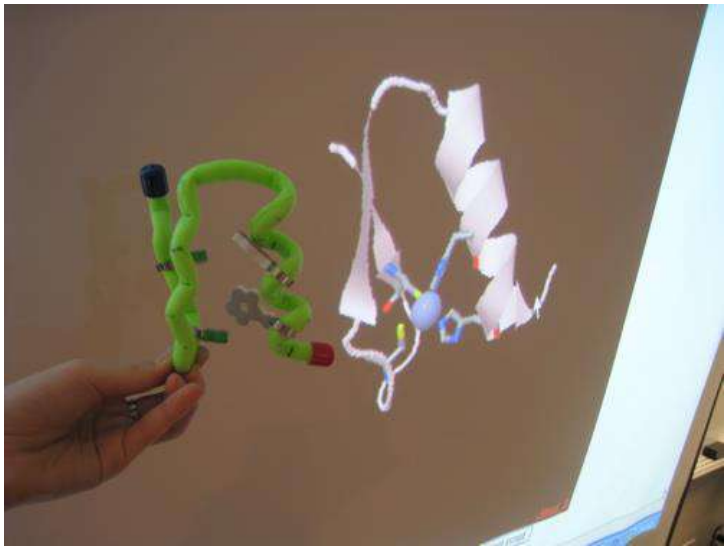
α-helix

Alpha helix: has **3.6 amino acids** per turn of the helix, which places the **C=O group of amino acid #1 exactly in line with the H-N group of amino acid #5** (and C=O #2 with H-N #6).



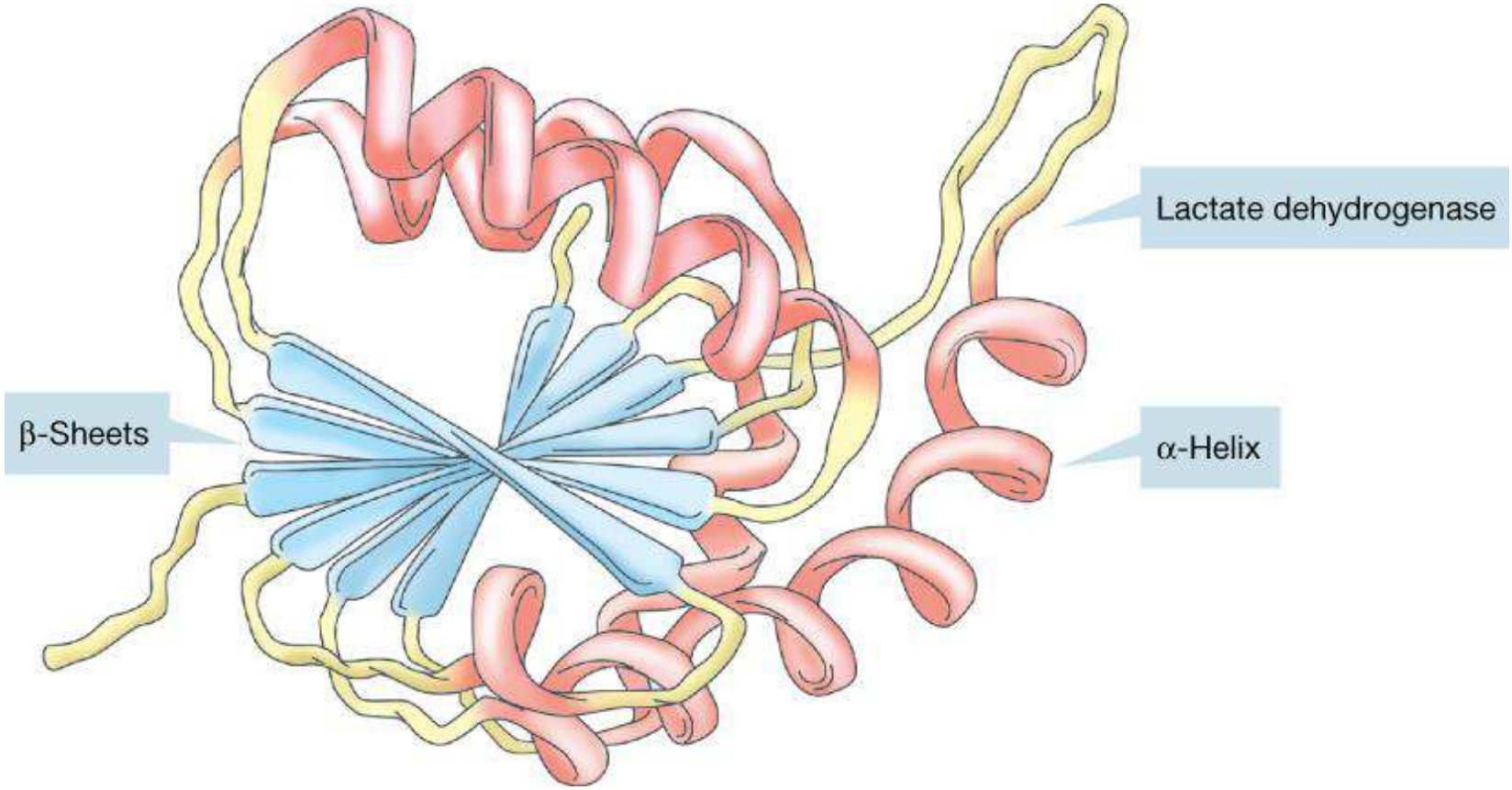
3(10) helix

- 3_{10} refers to a helix of 3 amino acids per turn, with hydrogen bonds from #1C=O to H-N #4; the hydrogen bond closes a loop of 10 atoms.



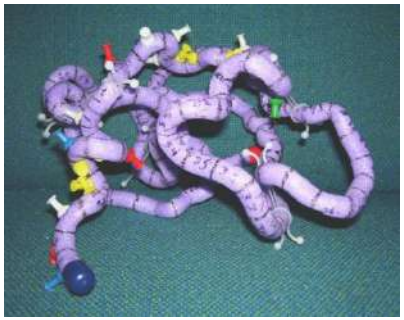
Tertiary Peptide Structure

- Peptides have three-dimensional structures



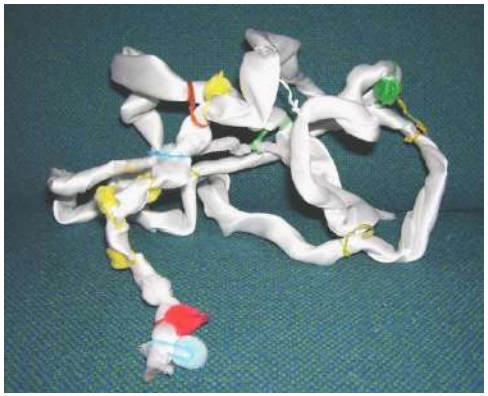
Tertiary Structure of a Protein

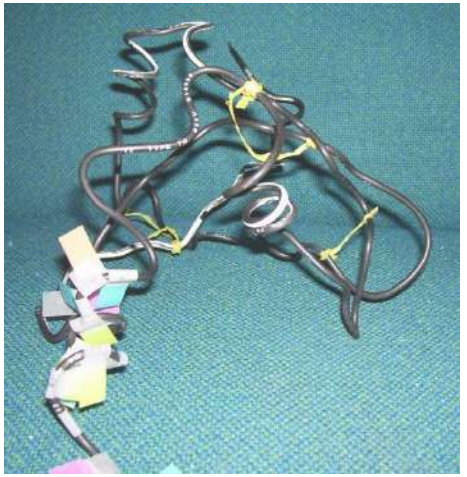
- The tertiary structure of the protein is the final folding that is the result of the molecular interactions formed by the primary and secondary structure.
- This is determined using the JMol program.
- What a finished pre-build might look



Toober

Pipe Cleaner





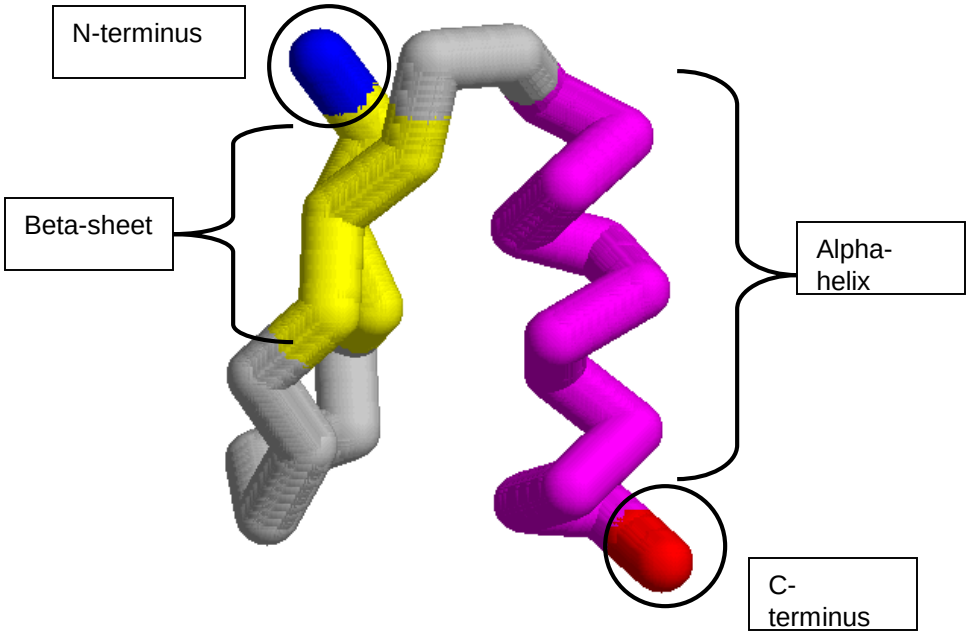
12 Gauge Wire

Past Competitions

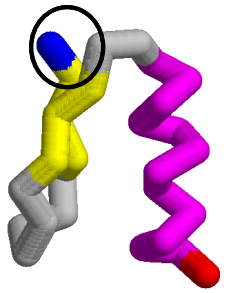
- 2010, 2011, 2012, 2015, 2016, and 2019
- http://cbm.msoe.edu/scienceOlympiad/scienceOlympiadPastEvents.php

Sample Rubric Sheet

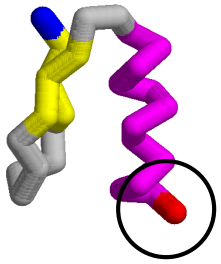
Whole Molecule Overview



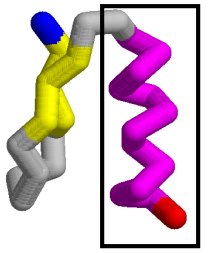
Rubric Details

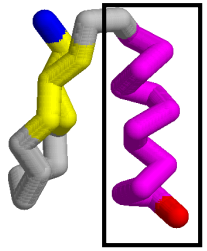


- Blue cap on N-terminal amino acid (Pro4) (1 pt)
  - To receive this point, the blue cap should be positioned on the fir amino acid. This should be next to the beta-strand. See picture t the right for correct placement of the blue end cap.
- Red cap on C-terminal amino acid (Gly31) (1 pt)
  - To receive this point, the red cap should be positioned on the last amino acid. This should be next to the alpha-helix. See picture to the right for correct placement of the red end cap.



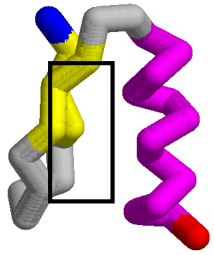
- Alpha helix (amino acids 19-31) is located at C-terminus of protein (2 pts)
  - There should be an alpha helix located at the C-terminus of the protein. See figure to right. On the model and in the figure, the alpha helix is colored magenta.





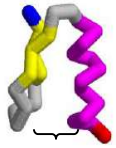
- Alpha helix is right-handed (2 pts)
  - Alpha helices are right-handed. Check the alpha helix in the model to confirm that the helix is right-handed. If the alpha helix is right-handed, the model is awarded two points.
  - To determine if the helix is right-handed, find one of the ends of the helix and imagine that the helix is a spiral staircase. Pretend that you are climbing that staircase and the helix is the hand-rail, which is always on the ourside edge of the staircase. If you would put your right hand on the toober as you go up the staircase, you have a right-handed helix. If you would put your left hand on the toober, you have a left-handed helix and the model would not receive the points.
- Alpha helix is properly formed (helix resembles a telephone cord) (1 pt)
  - The helix should be formed in such a way that it resembles a telephone cord stretched out slightly. The helix should not be compacted down so that there is not any space between the turns. It should also not be so stretched out that there is a lot of space between the turns
- Alpha helix is appropriate length (13 amino acids; ~3.5 turns) (2 pts)
  - The helix is 13 amino acids, and each turn in the helix is approximately 3.6 amino acids in length. Therefore, the length of this helix should be ~3.5 turns.

Rubric Details, Continued



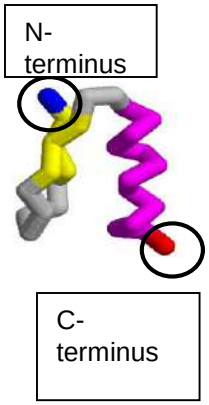
- Beta strand #1 (amino acids 5-7) (2 pts)
  - To receive these points, the model should have a beta strand from amino acids 5-7 (3 amino acids in length). The first beta strand should be located near the blue end cap. The model and the figure to the right have the beta strands colored yellow.
- . Beta strand #2 (amino acids 14-16) (2 pts)
  - To receive these points, the model should have a beta strand from amino acids 14-16 (3 amino acids in length). The second beta strand is located 6 amino acids (12 cm) away from the first. The model and the figure to the right have the beta strands colored yellow.
- Beta strand is formed properly (1 pt)
  - To receive this point, the model should have properly formed beta strands. The model can have the beta strands in a zig-zag shape (a bend every 2 cm) or it could have them be represented as straight regions to the model. There should not be any helical or coiled portions in this area

Rubric Details, Continued

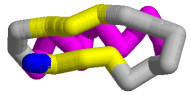


- Helix is arranged next to beta sheet (protein should be compact with a 2-stranded beta sheet lying next to an alpha helix; helix and sheet should not be too far apart) (2 pts)
  - To receive these points, the beta sheet and alpha helix should be located close to one another. There should only be enough space between the two secondary structure to allow for a zinc ion to coordinate between the 4 amino acids that bind the ion. In other words, there should not be much space between the alpha helix and the beta sheet
- 12. N-terminus (blue cap) and C-terminus (red cap) are pointed in opposite directions (2 pts)
  - To receive these points, the N-terminus and C-terminus of the protein should be facing away from each other. If you hold the model so that the beta sheet is facing the left and the helix is on the right (like the picture shown to the right), then the N-terminus should be pointing upward and the C-terminus is pointing downward.

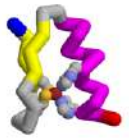
Please note that there is not much space between the two secondary structures.



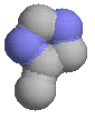
Rubric Details, Continued



- Model should be flat in that the beta strands and alpha helix are occupying the same plane (2 pts)
  - To receive these points, the alpha helix and beta sheet should be in the same plane (please see figure to the right). The model should be "flat" in that neither the helix nor the sheet protrudes upward or downward form the main axis. You should be able to look through the beta sheet and see the alpha helix



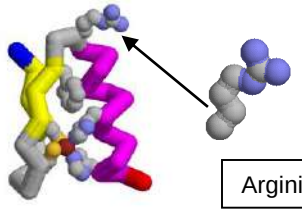
- Creative Additions to model (2 pts each):
  - Zinc ion
    - To receive these points, the model should have a zinc ion located between the alpha helix and beta sheet closer to the C-terminus than the N-terminus. Please model and figure to the right (zinc ion is colored dark red).
  - 2 Histidines (His 25 and 29) (coordinates Zn) 2 Cysteines (Cys 7 and 12) (coordinates Zn)
    - Cysteine
  - To receive these points, the model should have 2 Cysteines at positions #7 and #12.
    - If zinc ion is present, then these Cysteines should be connected to the zinc ion.
  - Arginine 18 (attaches to DNA)
    - Arginine
  - To receive these points, the model should have an Arginine at position 18.
    - If DNA is present on model, this amino acid should interact with the DNA.



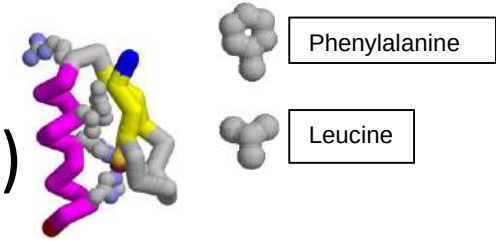
    - To receive these points, the model should have 2 Histidines at positions #25 and #29.
    - If zinc ion is present, then these Histidines should be connected to the zinc ion

Rubric Details 6: To receive point(s) …

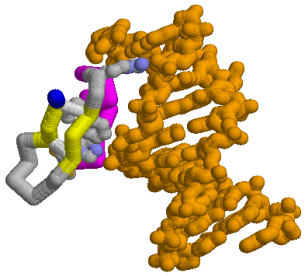
- Creative Additions to model (two points each):
  - Arginine 18 bound to DNA (2 pts)
    - The model should have an Arginine at position 18.
    - If DNA is present on model, this amino acid should interact with the DNA.



  - Hydrophobic amino acids (Phe16, Leu22) (2 pts)
    - These residues should face inward to create a stable hydrophobic core stabilizing protein



  - DNA attached to protein (2 pts)
    - The model should have DNA bound to the zinc finger.
    - Zinc finger should be in the major groove of the DNA.



Rubric Details 7: To receive point(s) …

- Creative additions are appropriate (2 pts)
  - The creative additions should be relevant to telling the functional story of the protein.
  - Any amino acid shown should play an important role in the stability (Zinc ion coordination, hydrophobic core) or function of the molecule (binding to DNA).
  - Models that have displayed all of the amino acids should not be awarded these points.
- Creative additions are accurate (2 pts)
  - The creative additions need to be accurate and reflect the scientific information that has been provided in Goodsell's Molecule of the Month, the PDB file or alternative resources.
- Students submitted a 3x5 card to explain model (2 pts)
  - A 3x5 card should be submitted along with the model, describing what additional features have been added to the model and what they represents.