

Bioinformatics
CS300
Prediction and
Modeling Protein Structure

Fall 2019
Oliver BONHAM-CARTER



Exam 2

- Monday 18th November 2019 during lab.
- Multiple choice and very similar to exam 1
- Ten questions, ten points a piece.
- Conceptually-oriented, concerning material from class discussion, slides and activities





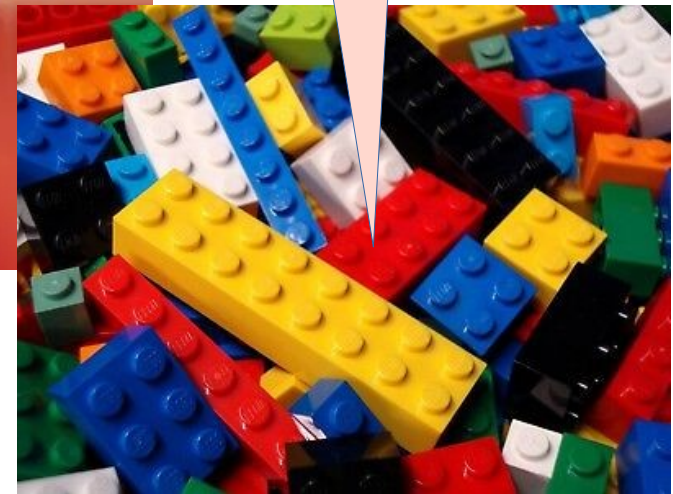
General Topics

- Gene sequencing
- Horizontal gene transfer
- Blast, tasks and outputs
- Annotation tasks for a newly sequenced genomes
- Gene prediction; concepts
- Genomes; main differences in terms of annotation
 - Prokaryotic and Eukaryotic
- Substitution matrices
- Translation of Genes in DNA
- Open reading frames, NCBI's ORF finder tool
- Gene prediction landmarks
- Protein Folding

Properties From Combining Pieces

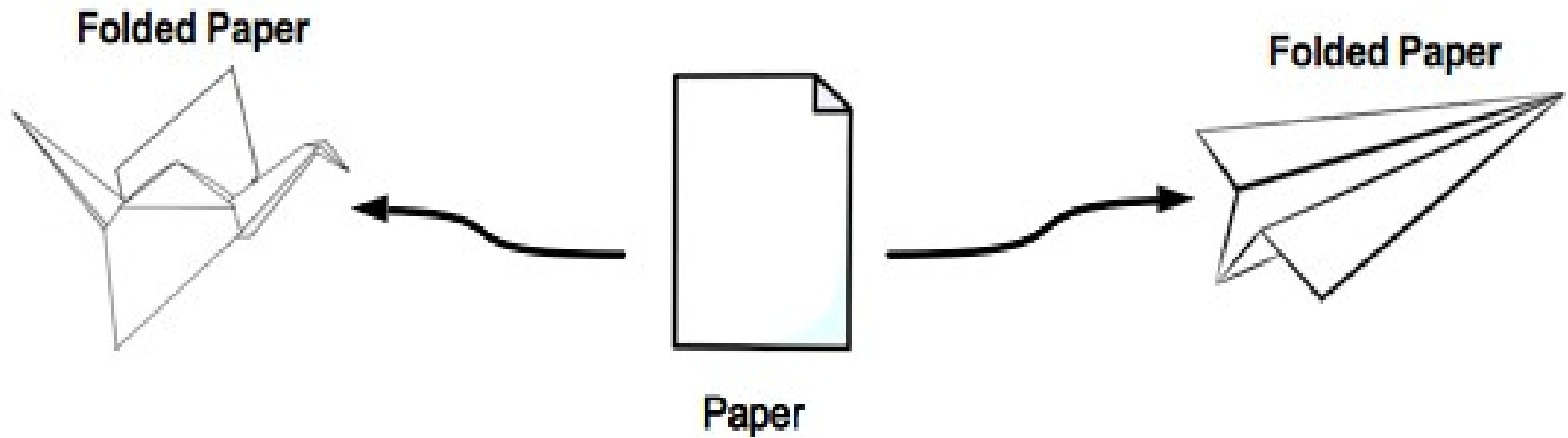


**From
these
pieces?**

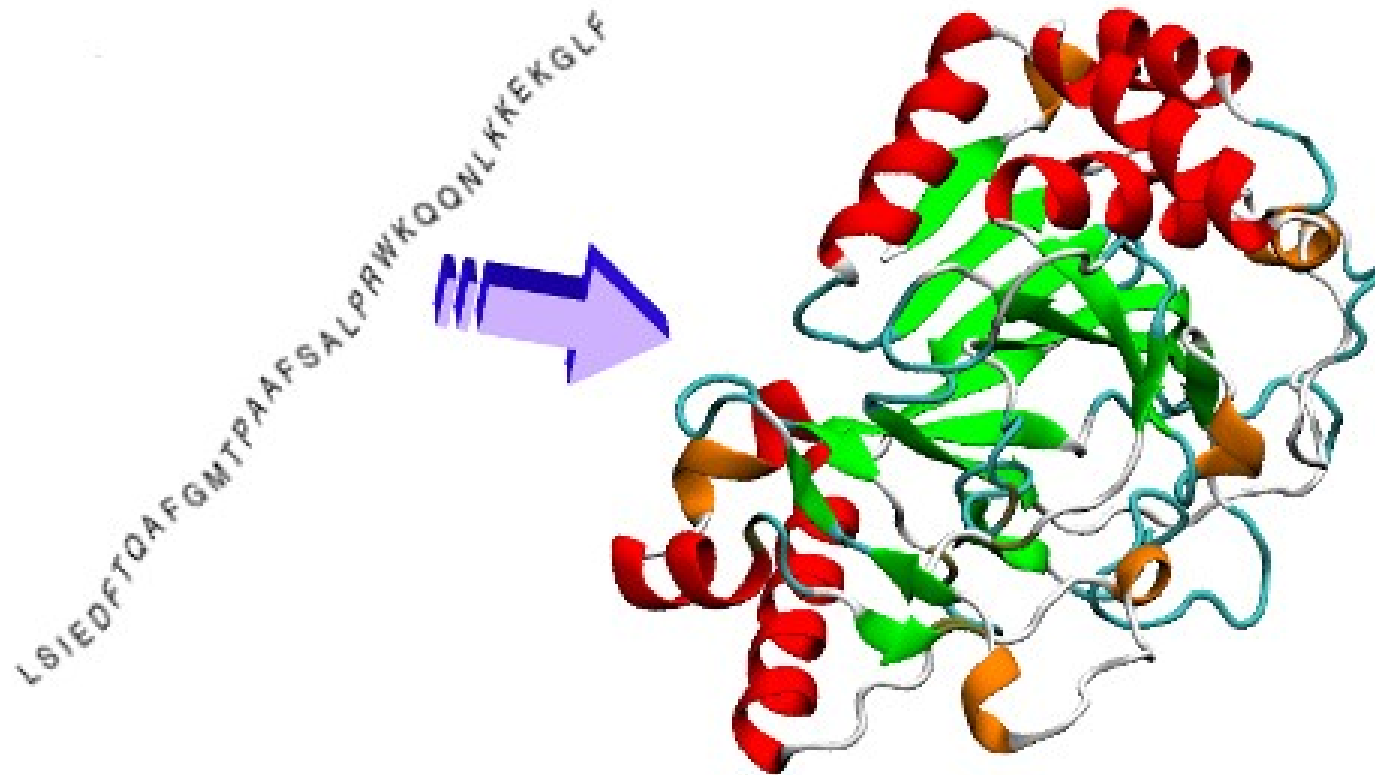


**A cool living room
made from Lego pieces!**

Properties From Folding



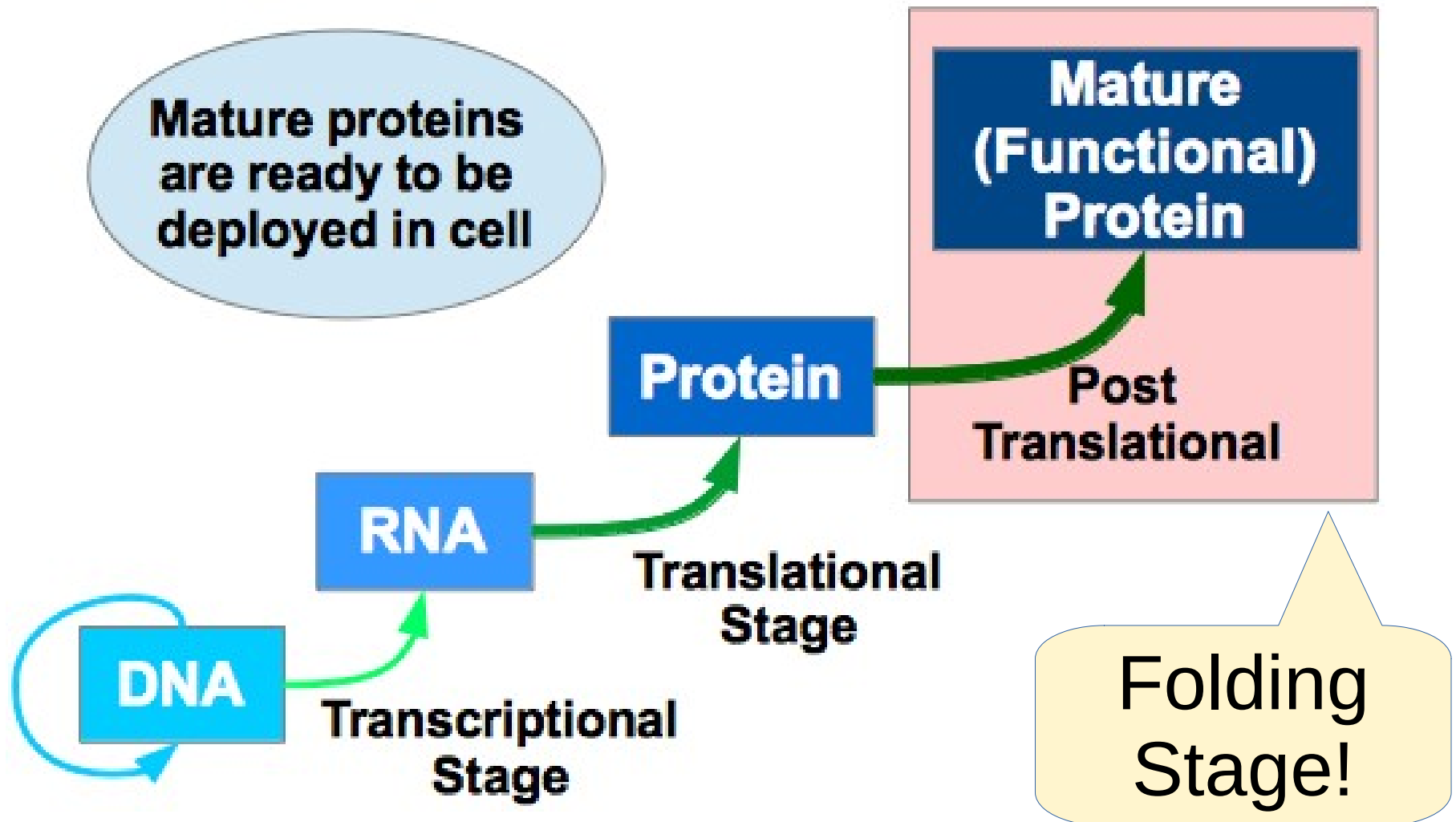
Protein Folding



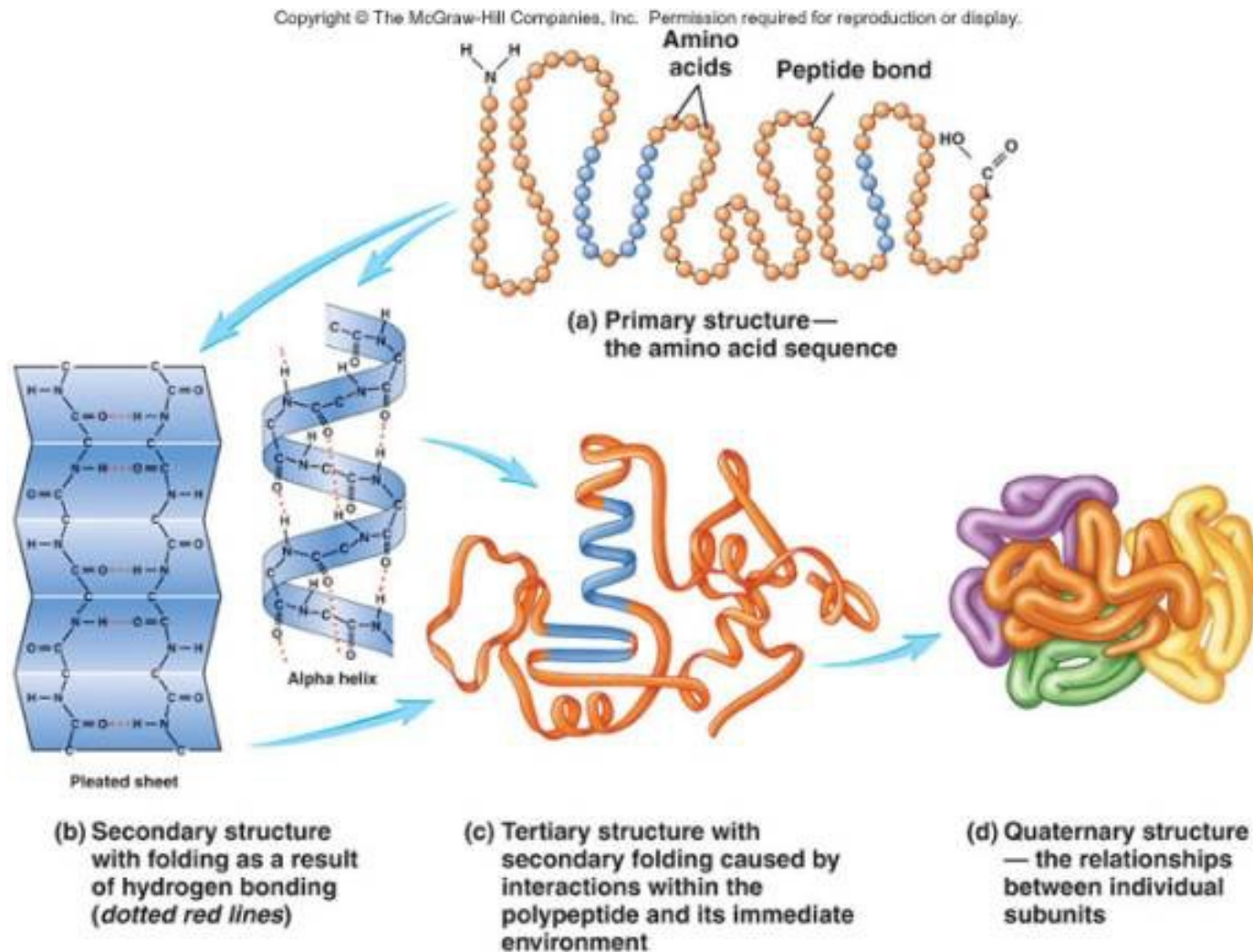
- linear chain of amino acids produced by ribosome during translation
- folds into functional, 3D state based on properties of amino acids and structure



Protein Folding and the Central Dogma of Biology

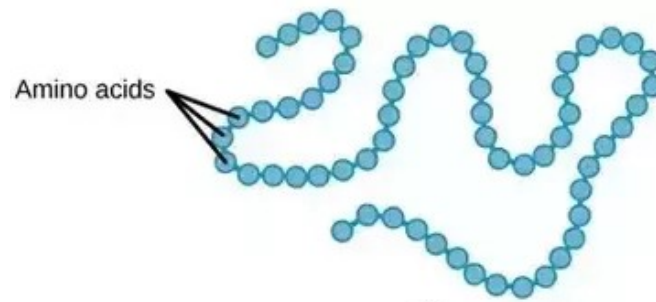


Protein Folding: Four Stages

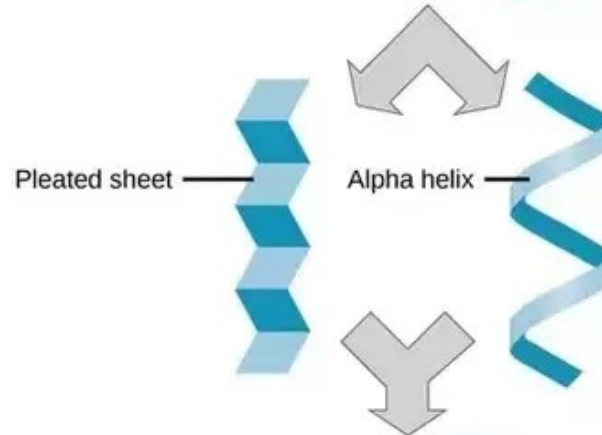


- Linear chain of amino acids produced by ribosome during translation
- Folds into functional, 3D state based on properties of amino acids and structure

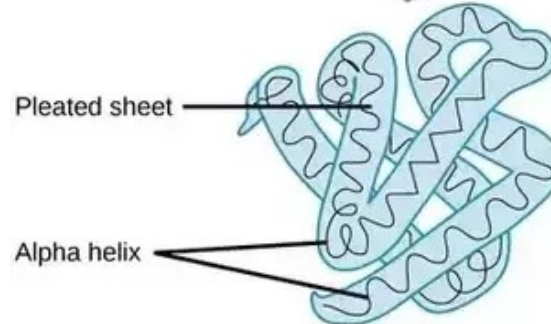
Protein Folding: Another View



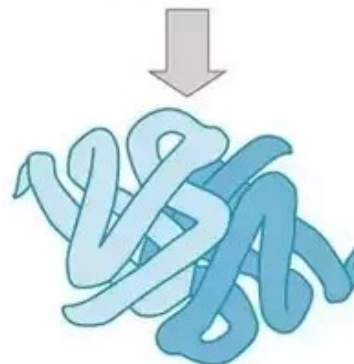
Primary Protein structure
sequence of a chain of amino acids



Secondary Protein structure
hydrogen bonding of the peptide backbone causes the amino acids to fold into a repeating pattern



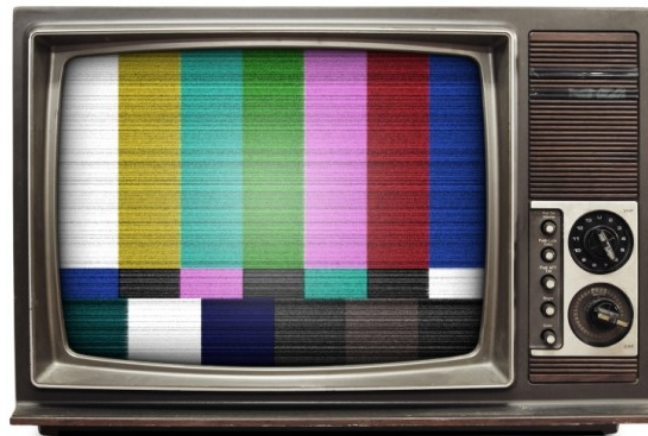
Tertiary protein structure
three-dimensional folding pattern of a protein due to side chain interactions



Quaternary protein structure
protein consisting of more than one amino acid chain

Supporting Videos

- **Protein Folding (3 mins)**
 - <https://www.youtube.com/watch?v=yZ2aY5lxEGE>
- **What is a protein? (3D shape and function, 3 mins)**
 - <https://www.youtube.com/watch?v=qBRFIMcxZNM>
- **Protein folding simulation (3 mins)**
 - <https://www.youtube.com/watch?v=meNEUTn9Atg>



Protein Folding - Applications

- **Protein must fold correctly to function**
- Misfolded proteins
 - Accumulation – Huntington's and Parkinson's disease
 - Tagged for degradation – emphysema, cystic fibrosis
 - Pharmaceutical chaperones – fold mutated proteins to render them functional
- Antiviral drug development
 - Antibiotics vs antivirals
 - Bacteria – cells
 - Viruses – invade host's cells



Bacteria

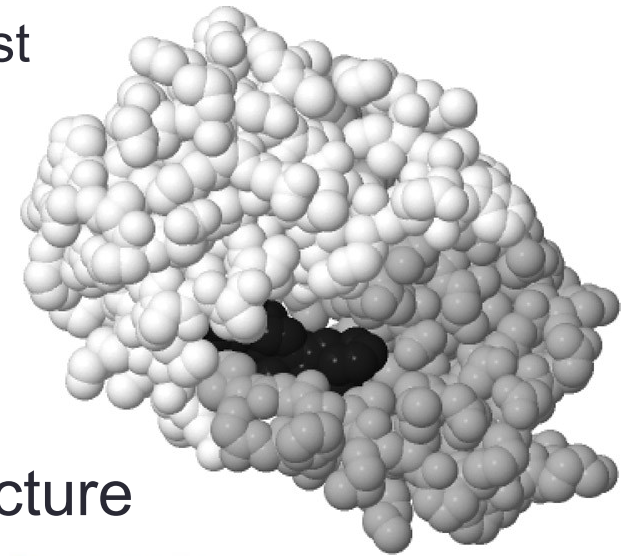
VS.



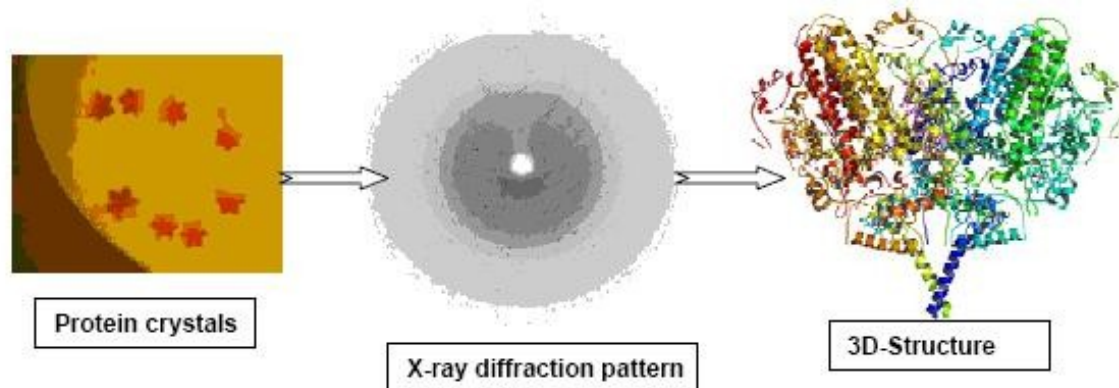
Virus

Protein Folding - Applications

- Antimicrobial Drugs: helps to...
 1. Be effective against the disease-causing agent
 2. Be selectively toxic
 - kill or inhibit the microbe without harming the host
- Rational Drug Design: helps to...
 - Understand 3-D structure of viral proteins
 - Design drugs to fit protein and block function



Laboratory – challenging to predict 3-D structure





Genomics & Computational Structural Biology

Genomics

- Determines the ordered sequence of nucleotides in a genome
- Assigns predicted functions to regions of nucleotides via annotation

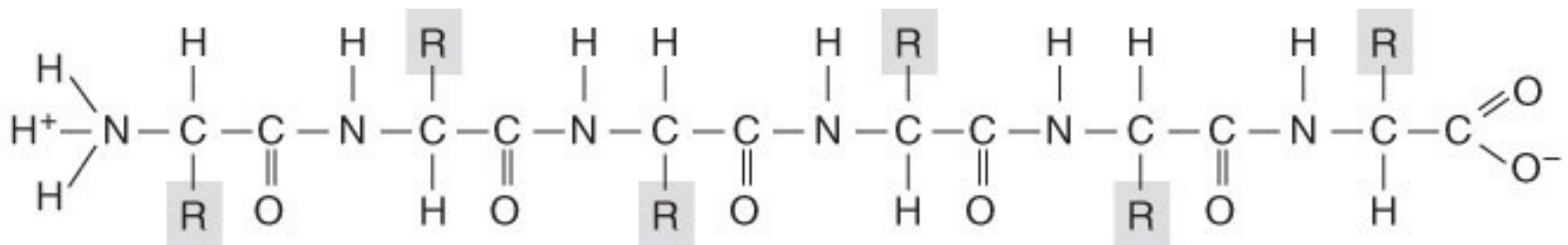
Computational Structural Biology

- Takes predicted gene sequence for translation into primary amino acid sequence
- Predicts 3-D protein structure based on primary amino acid sequence
- Note: this step is very difficult because the number of possible outcomes to process and consider is enormous
- The study of structural rules and their contribution to the final mature protein.

Structural Rules for Protein Folding

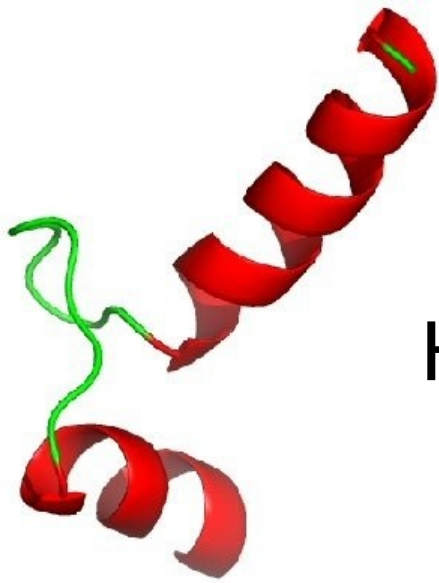
- Linus Pauling – Studied the limitations on protein folding
 - Nature of chemical bonds between amino acids
 - Bond angles
 - Rotation of atoms
 - Flexibility of side chains

Christian B. Anfisen – Studied the influence of thermodynamics of cellular environment

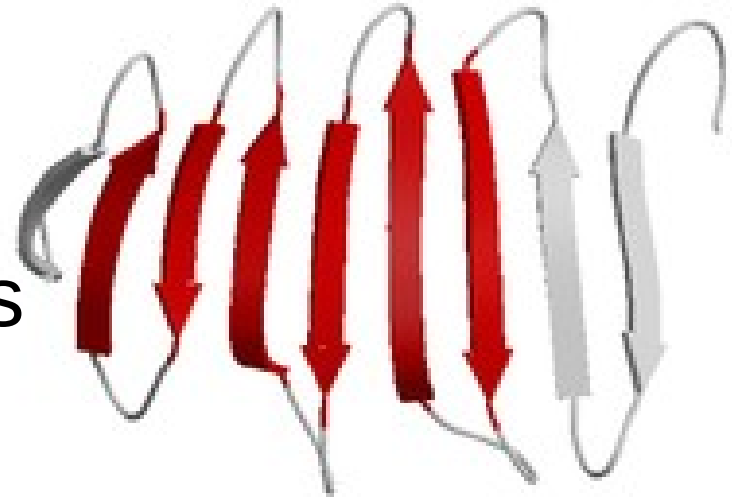


(A) Primary (1°) structure

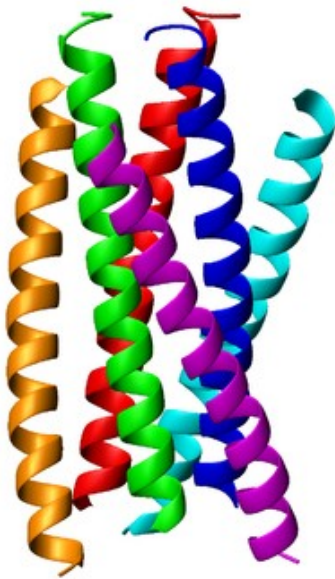
Parts of Protein (Structures)



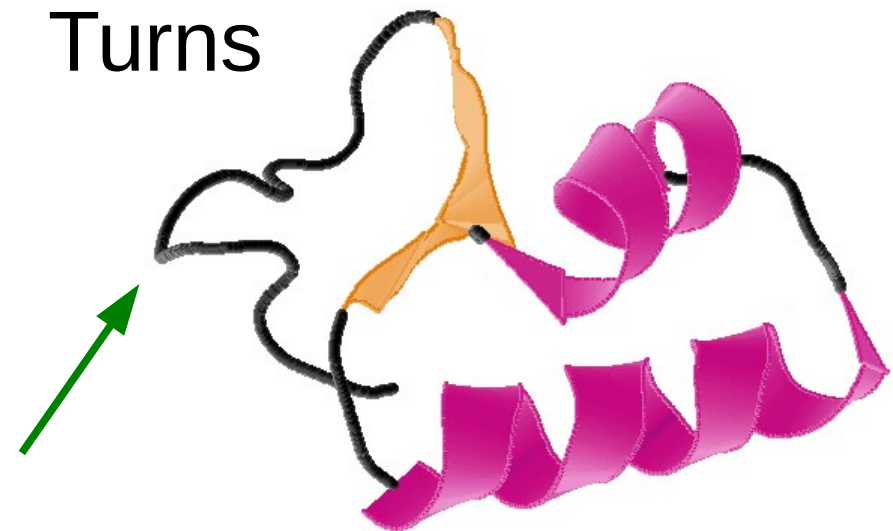
Helices



Sheets



Coils



Turns



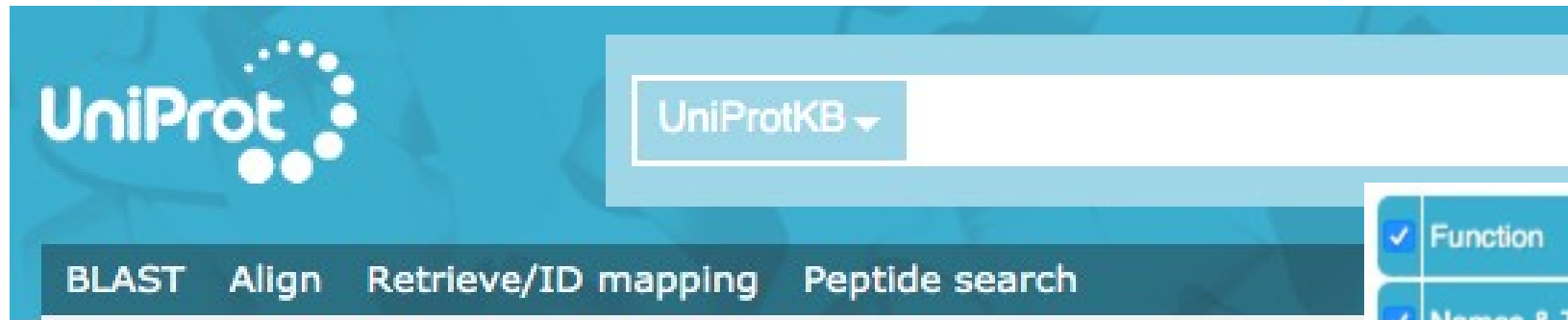
Protein Folding: An Idea of Structure

- **Garnier**: a text-based, command-line tool from EMBOSS
 - Input: protein sequence in fasta format
 - Output: a model of folding in text base
 - Usage: **garnier file.fasta**

```
      .   10   .   20   .   30   .   40   .   50
MQIFVKLTGTITLEVEPSDTIENVKAKIQDKEGIPPDQQLIFAGKQL
helix  HH                      HHHHHHHHHHHH      H
sheet  EEEE          EEEEE
turns   T                      TTTT  TT
coil    CCC CC          CCC          CC          CCCCC
      .   60   .   70
EDGRTLSDYNIQKESVNHLVLRRLGG
helix                      HHH HHH
sheet                      EEEEE
turns  TTTT  TT          TTT
coil    CCC  CC  C

#-----
#
#  Residue totals: H: 20   E: 19   T: 16   C: 21
#                   percent: H: 33.3 E: 31.7 T: 26.7 C: 35.0
#
#-----
```

H: Helices, E: Sheets
T: Turns, C: Coils



Protein Information: The mission of UniProt is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.

- Some protein study databases require information from UniProt to find protein samples.
- <https://www.uniprot.org/>

✓	Function
✓	Names & Taxonomy
✓	Subcellular location
✓	Pathology & Biotech
✓	PTM / Processing
✓	Expression
✓	Interaction
✓	Structure
✓	Family & Domains
✓	Sequences (2)
✓	Similar proteins
✓	Cross-references
✓	Entry information
✓	Miscellaneous

PREDICT PROTEIN

Dashboard Overview for CD44_HUMAN



Structural Annotations of protein: prediction of protein function, e.g. assisting in the annotation of subcellular localization (LocTree, LocTree2, NLSpred), identifying protein-protein interaction sites (PPSites) and protein-DNA binding sites, and more.

- <https://www.predictprotein.org/>
- <https://open.predictprotein.org/>



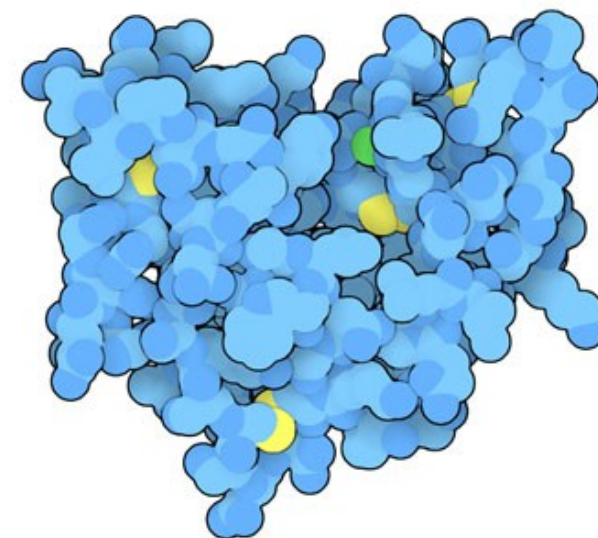
RCSB PDB

PROTEIN DATA BANK

Protein archives: This resource is powered by the Protein Data Bank archive-information about the 3D shapes of proteins, nucleic acids, and complex assemblies that helps students and researchers understand all aspects of biomedicine and agriculture, from protein synthesis to health and disease.

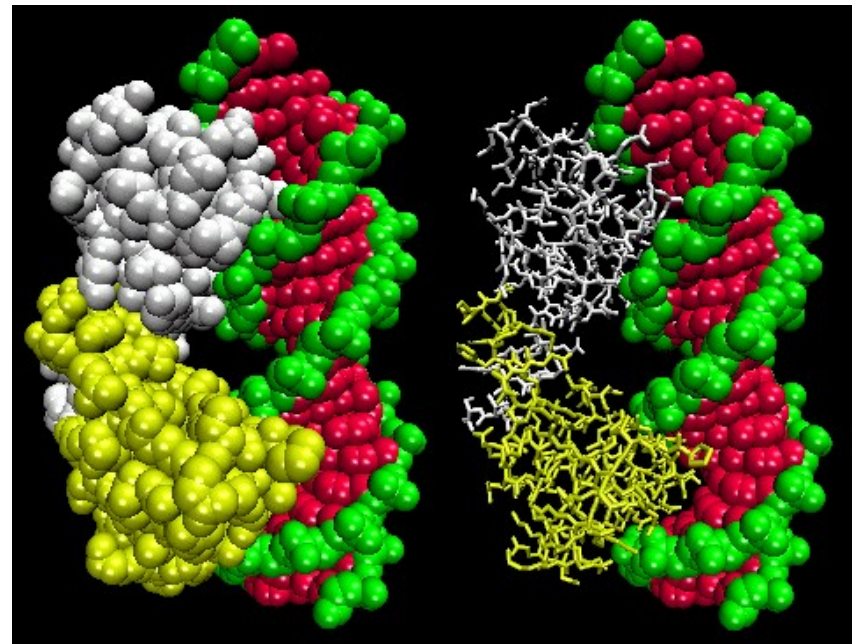
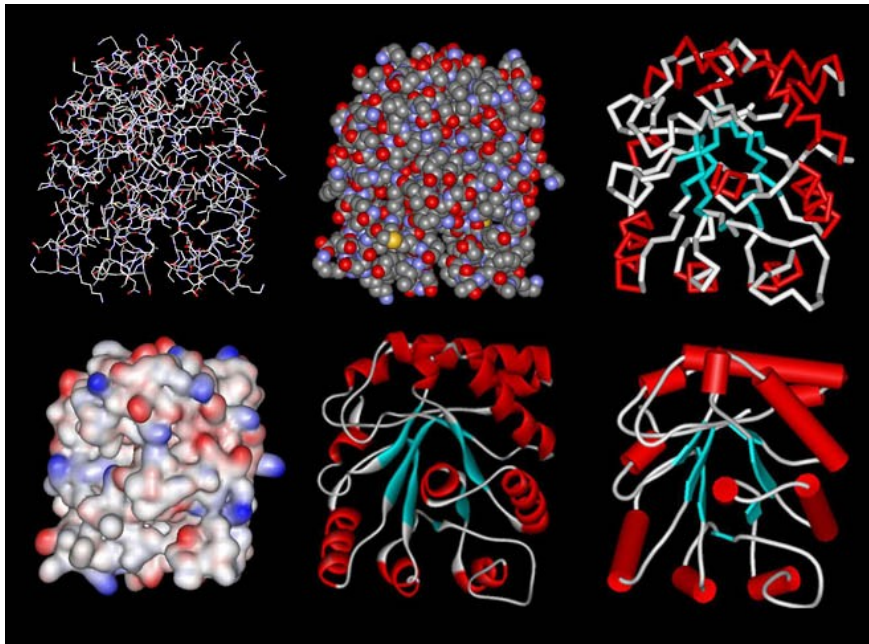


- <http://www.rcsb.org/>

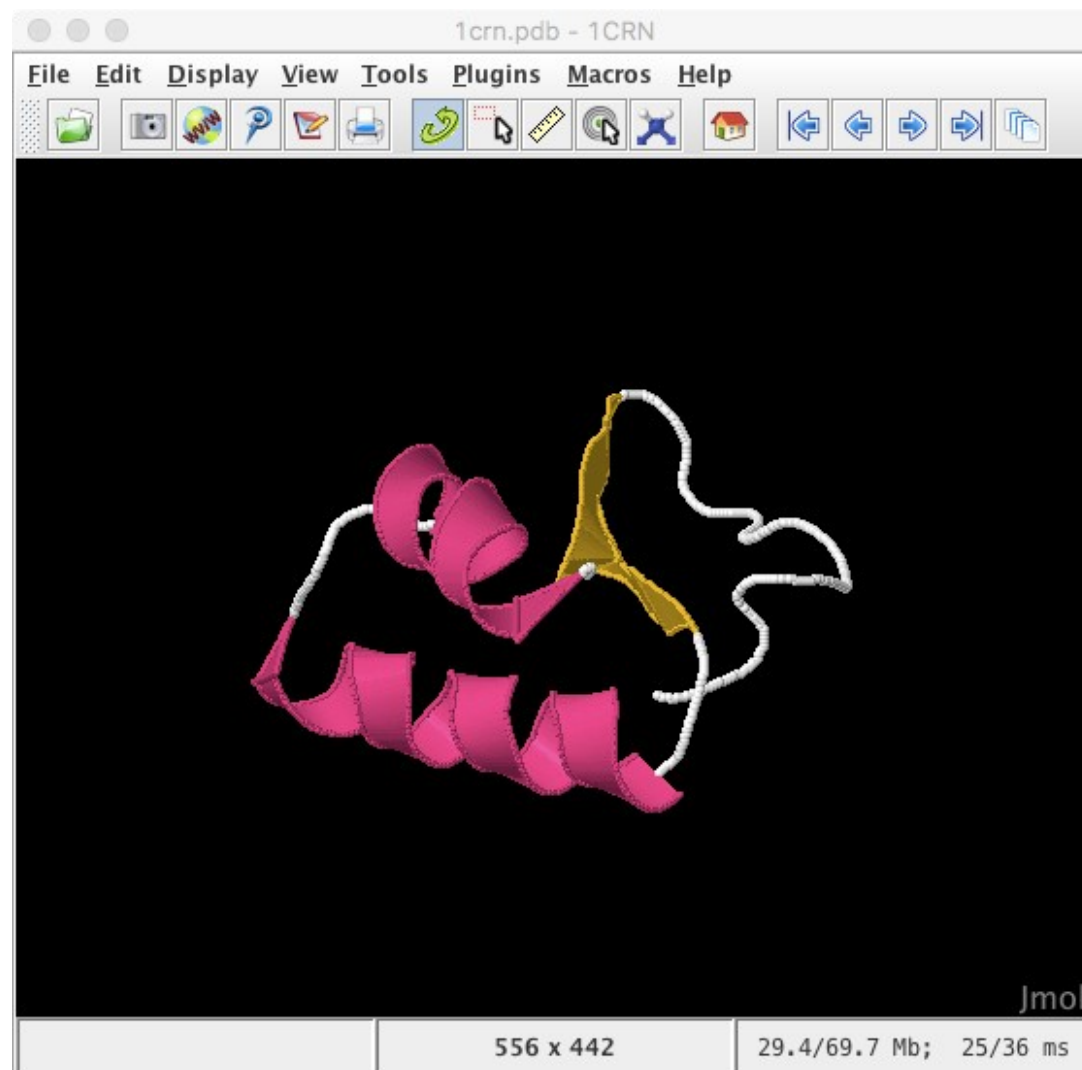


Protein DataBase (PDB)

- Database for 3-D structural data of large biological molecules
- <https://www.rcsb.org/>
- Data is viewable using jmol.



Jmol: A Graphical Viewer For Protein Sequences



- Download:
 - <http://jmol.sourceforge.net/>
- Wiki:
 - http://wiki.jmol.org/index.php/Jmol_Application#Installing_Jmol_Application