Review of Scale of Biological World

The cell at 1,000,000 X

Review of Protein Structure / Folding

Goals for this review unit:

Protein Structure

- **Definitions of primary, secondary, tertiary and quaternary structures**
- **Common secondary structures**
- **Phi, Psi (ϕ / ψ) angles / How to read a Ramachandran Plot**
- **Common terms used to describe protein structure**
  motifs / domains - some examples

Protein Folding / Unfolding (denaturation)

- **Energetics / Intra and Intermolecular forces**

Prediction of Protein Structure

Early method – Chou / Fasman

Size scale of the biochemical world.

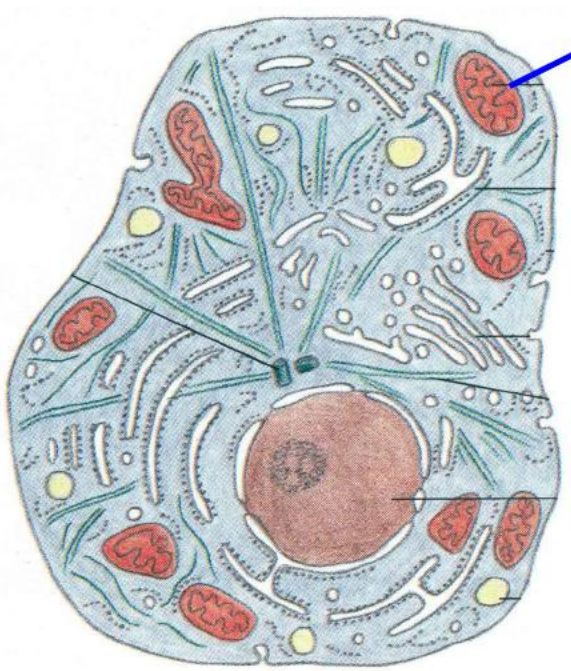
How big is an atom? Length of chemical bond?

Size of a typical bacterial cell? Eukaryotic cell?

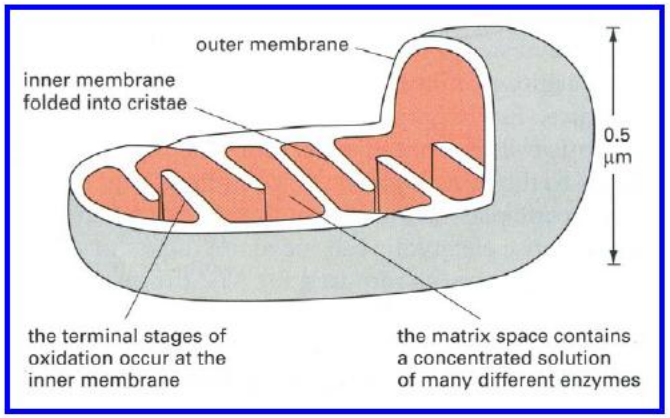
Prokaryotic cells

Eukaryotic cells have cell nuclei, other intracellular compartments, many other biological differences from prokaryotes.

Cell components



Mitochondria



- power plants of all eucaryotic cells
- converting food and O_2 into ATP

iClicker: Review Questions

1. What is the approximate diameter of an alpha helix?

A) 0.01 nm B) 0.1 nm C) 1 nm D) 2 nm E) 10 nm

2. Which amino acid has a "phenol" group in its side chain group

A) Trp B) Phe C) Arg D) Lys E) Tyr

3. The Keq for the reaction A → B is 10. Under conditions such that the concentration of B is ten times that of A, $\Delta G^{o'}$ would be expected to be:

A) negative B) zero C) positive

Proteins: Biological function depends on conformation

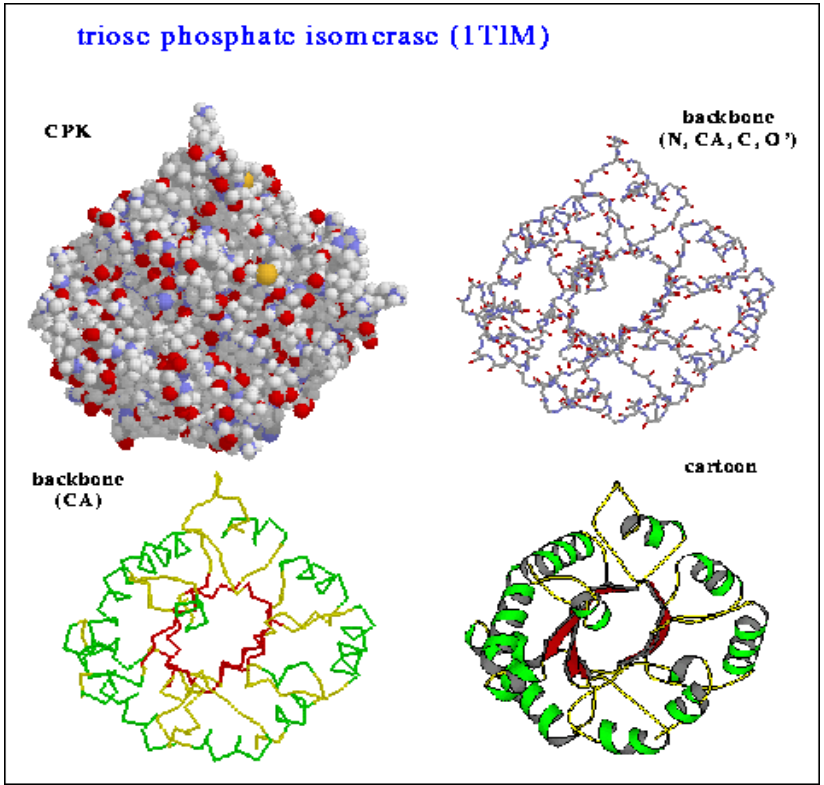
Unique Primary Structure = Unique 3D Structure ??
(Covalent bonds) (Noncovalent Interactions)

Globular Proteins: water soluble, compact, hydrophobic interior / hydrophilic surface
enzymes, receptors, carriers, hormones, etc. (***dynamic*** *agents*)

Fibrous Proteins: water insoluble, structural roles, extended structure
collagen (tendons, bone), a-keratin (hair, nails), etc. (~***static*** *agents*)

APRKFFVGGNWKMNGDKKSLGELIHTL
NGAKLSADTEVVCGAPSIYLDFARQKL
DAKIGVAAQNCYKVPKGAFTGEISPAM
IKDIGAAWVILGHSERRHVFGESDELI
GQKVAHALAEGLGVIACIGEKLDEREA
GITEKVVFEQTKAIADNVKDWSKVVLA
YEPVWAIGTGKTATPQQAQEVHEKLRG
WLKSHVSDAVAQSTRIIYGGSVTGGNC
KELASQHDVDGFLVGGASLKPEFVDII
NAKH

=

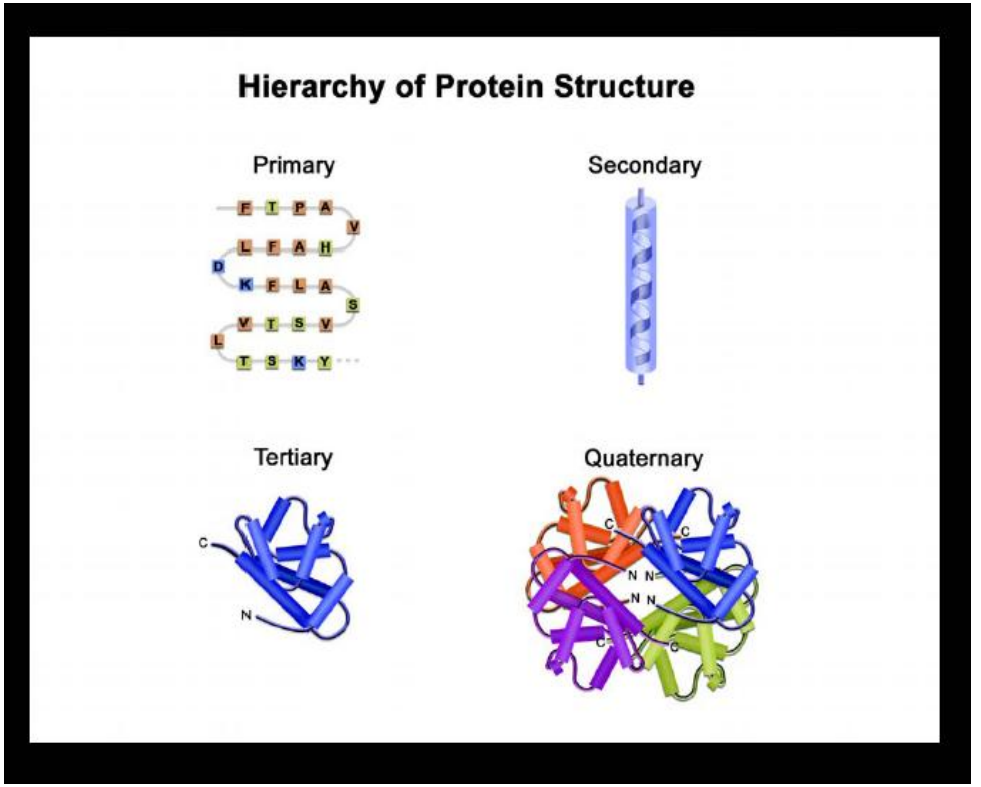


Four Levels of Description of (Native) Protein Structure

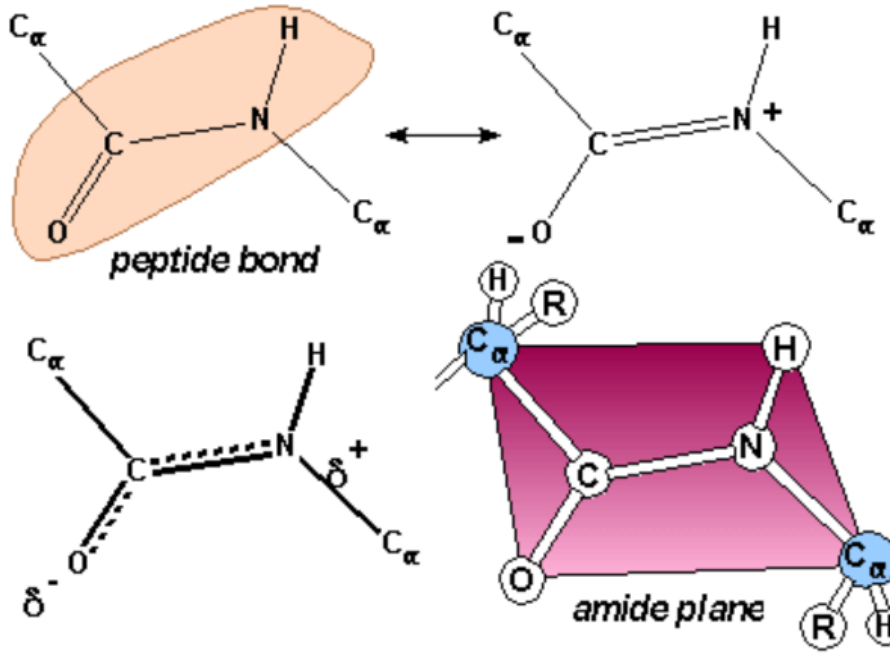
- **Primary Structure**: (~60-1000 amino acid residues) linear seq. of amino acid residues, covalent bonding including -SS- (also called "covalent structure“)
  (*the **primary structure** of a biological molecule is the exact specification of its atomic composition and the chemical bonds connecting those atoms (including stereochemistry). In general, polypeptides are unbranched polymers. However, proteins can become cross-linked, most commonly by disulfide bonds, and the primary structure also requires specifying the cross-linking atoms, e.g., specifying the cysteines involved in the protein's disulfide or other covalent bonds.)*
- **Secondary Structure**:
  Local conformations of backbone, maintained by hydrogen bonds
- **Tertiary Structure**:
  3D structure of a subunit (one polypeptide chain) in its native state
- **Quaternary Structure**:
  Spatial arrangement of subunits in oligomeric proteins

****

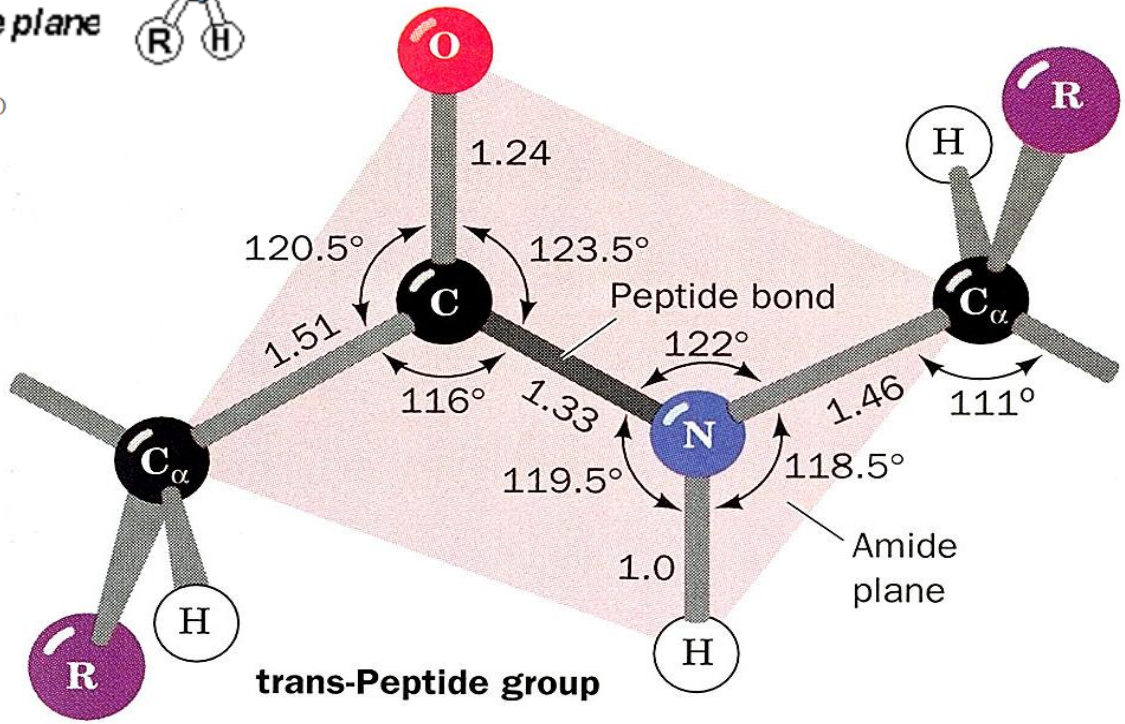
- ***Denaturation**: Partial to complete unfolding*
  Denatured Protein: Protein that has lost its native conformation



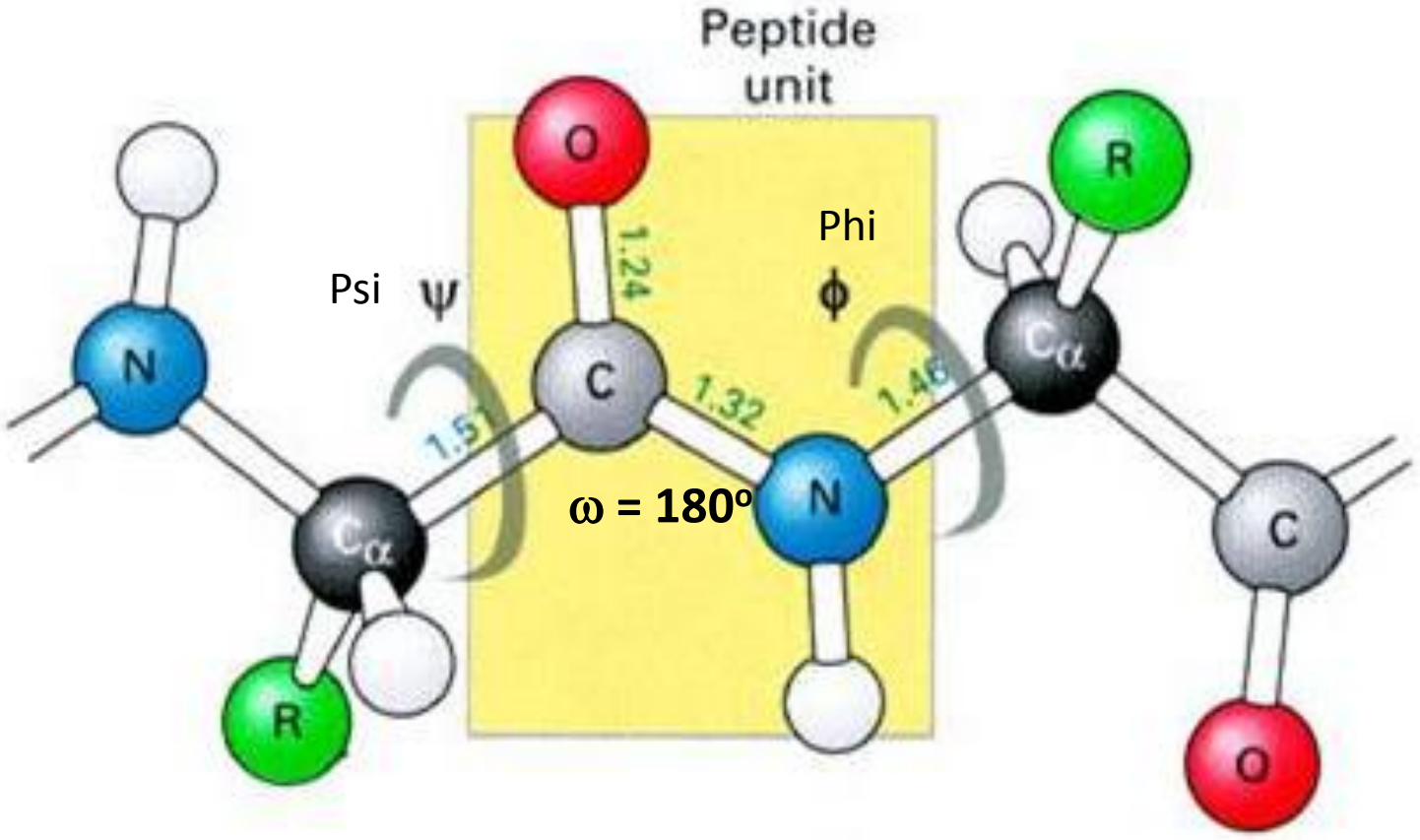
Chemistry of the peptide bond



This image was created by Dr. George Helmkamp, Jr. (UKMC)

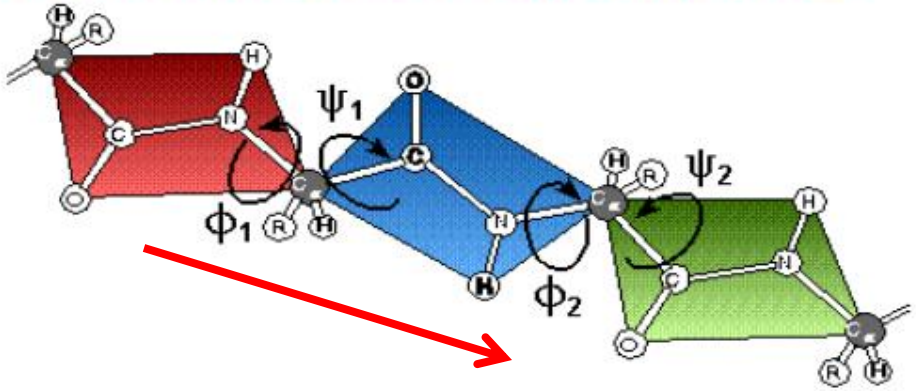


Phi-Psi angles



A peptide has partial double bond character (**ω = 180°**), thus only two angles (phi and psi) will determine the backbone for trans peptides.

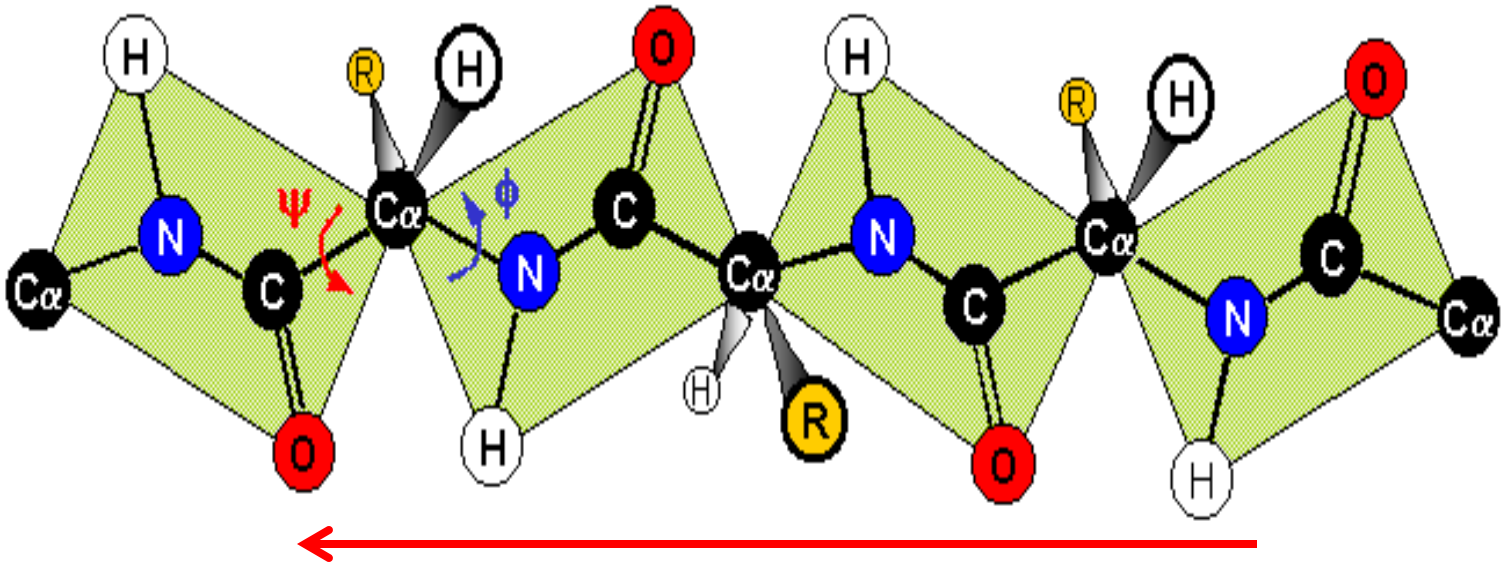
Conformation of a polypeptide



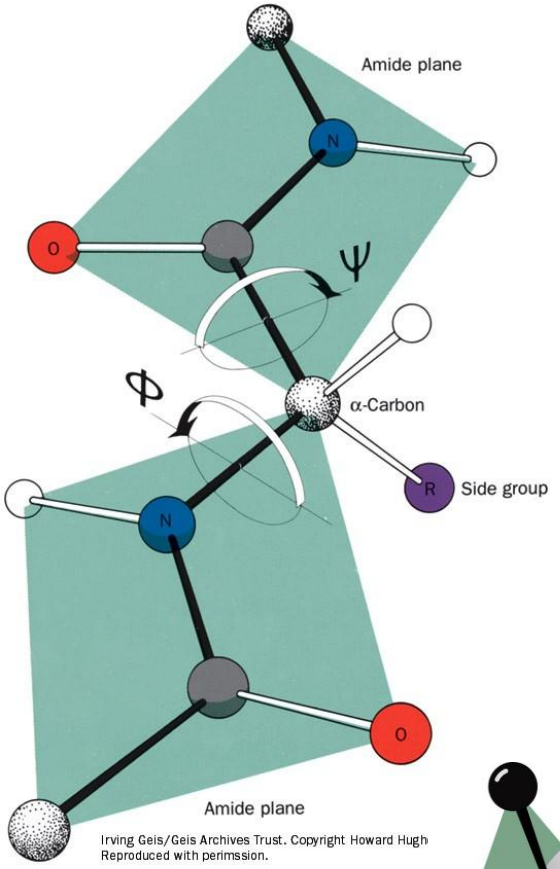
ϕ - rotation around the N-C_α bond

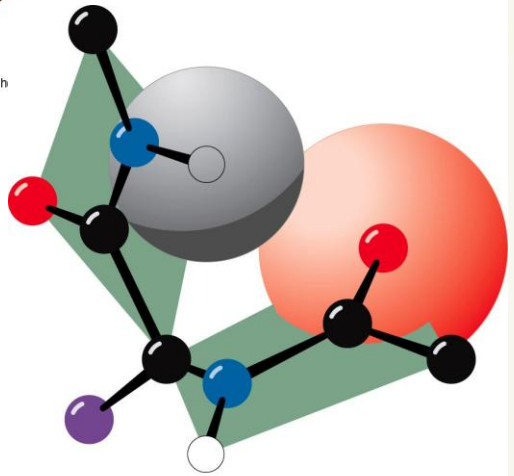
ψ - rotation around the C_α-C bond

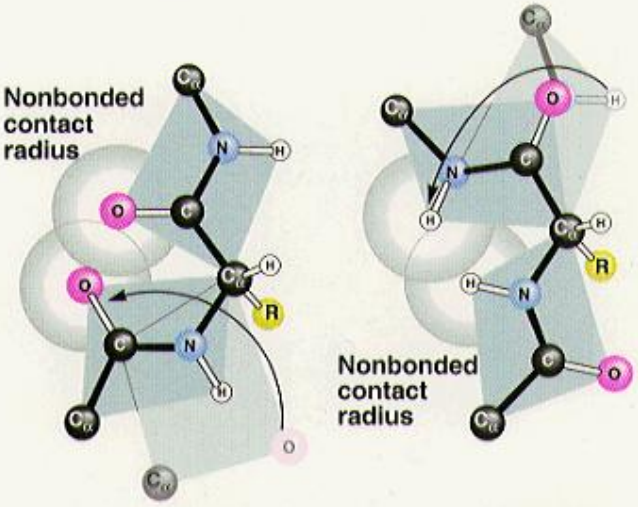
FULLY EXTENDED POLYPEPTIDE CHAIN



Torsion angles / steric restrictions

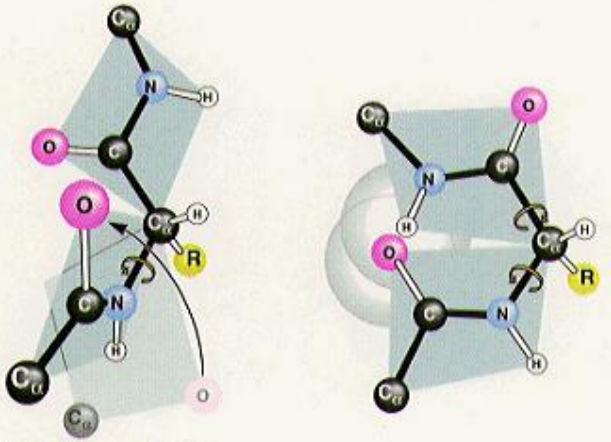






$\phi = 0°, \psi = 180°$

$\phi = 180°, \psi = 0°$

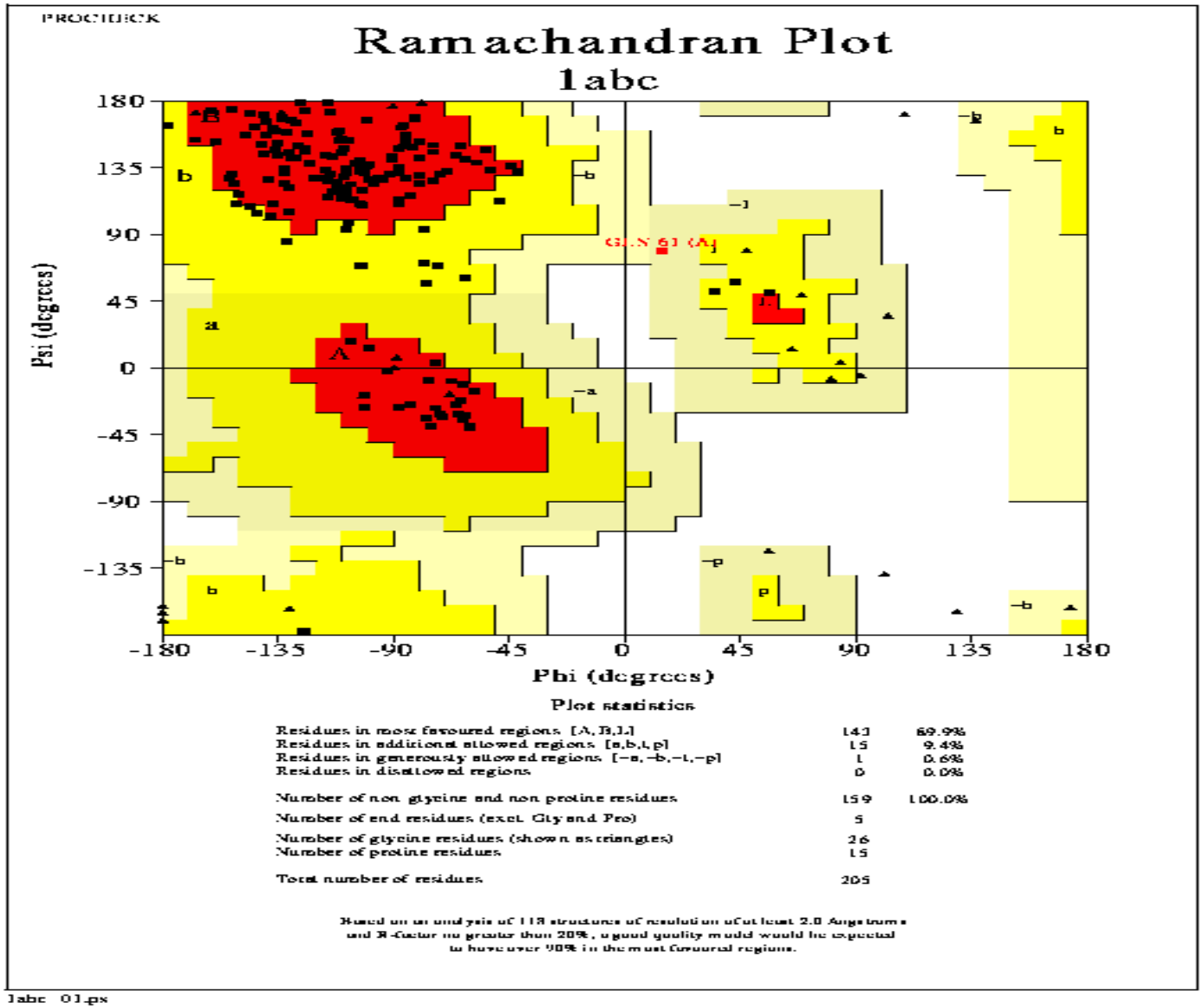


$\phi = -60°, \psi = 180°$

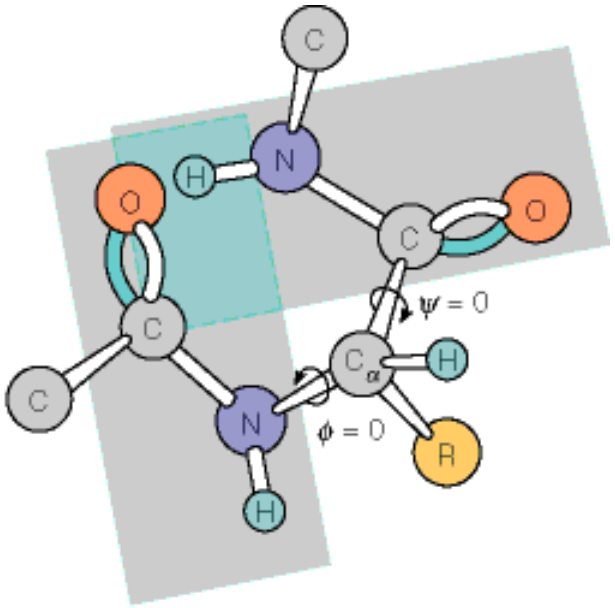
A further ϕ rotation of 120° removes the bulky carbonyl group as far as possible from the side chain

$\phi = 0°, \psi = 0°$

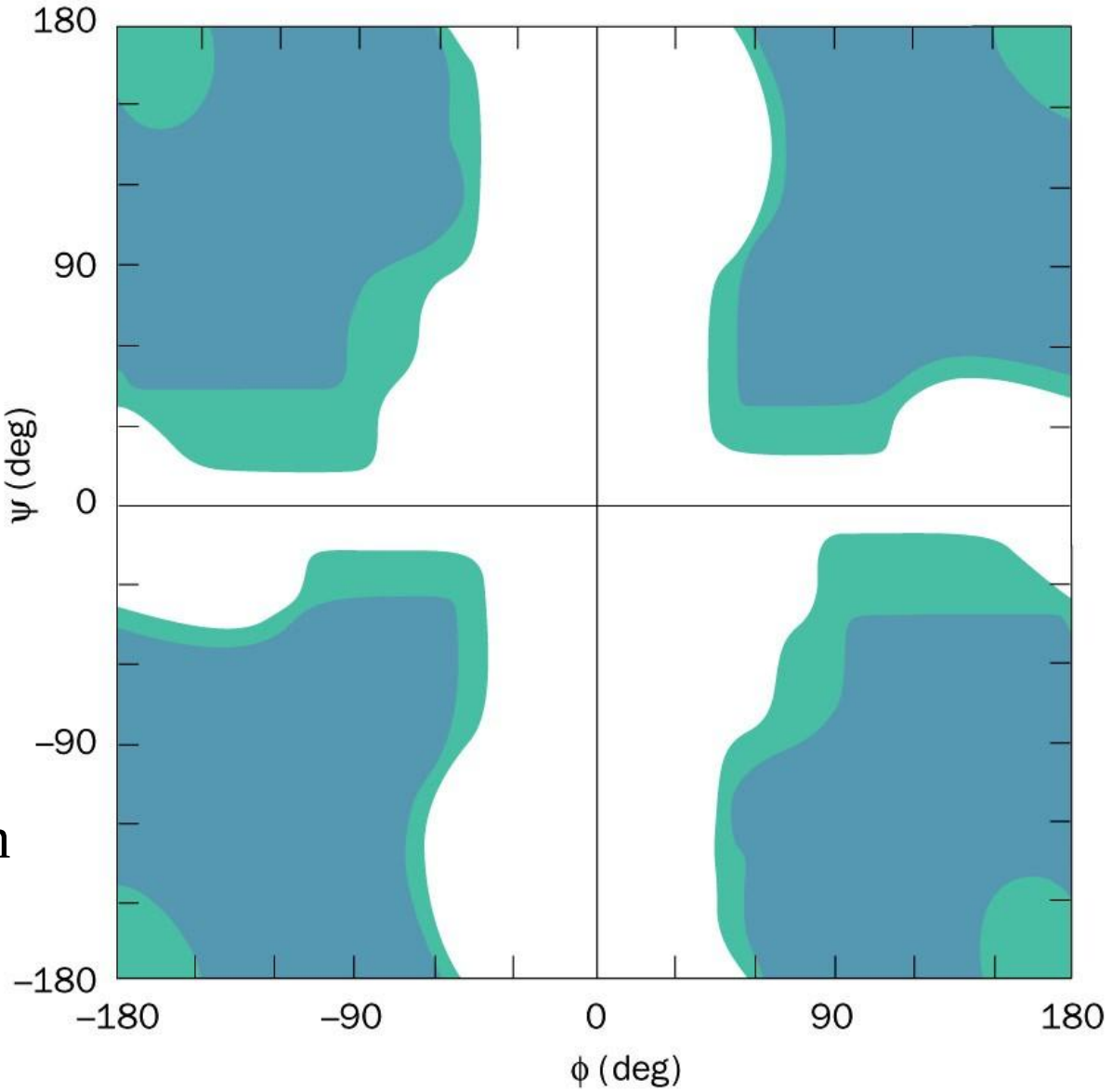
Ramachandran Plot



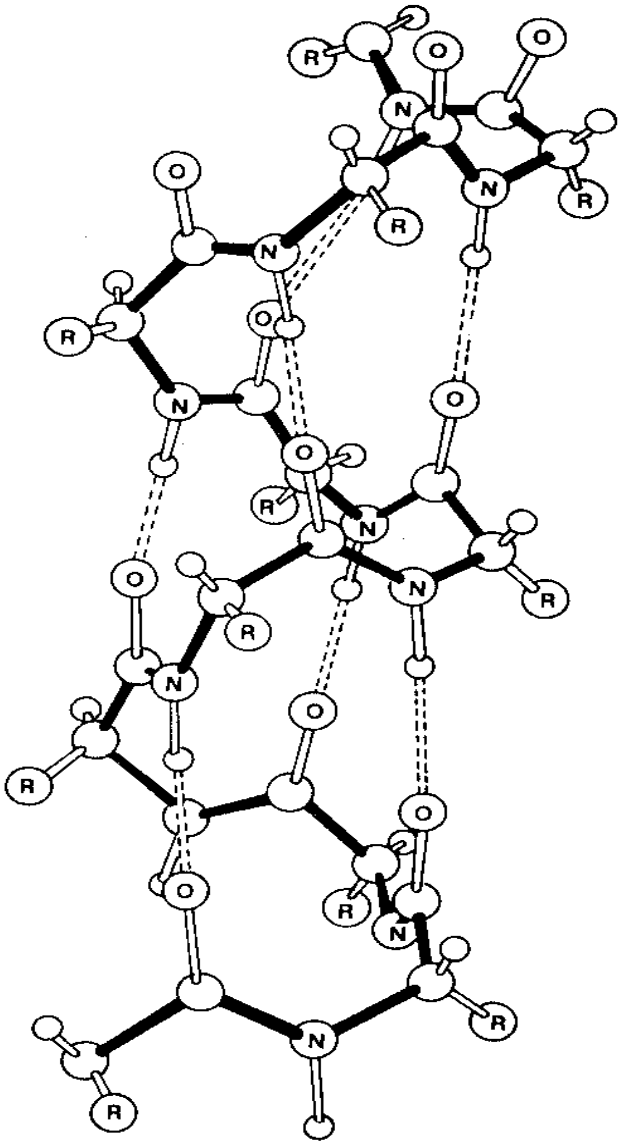
Allowed torsion angles for Gly residues: Restrictions only from the peptide units



Forbidden conformation in the center of plot



Protein Secondary Structure: Helices

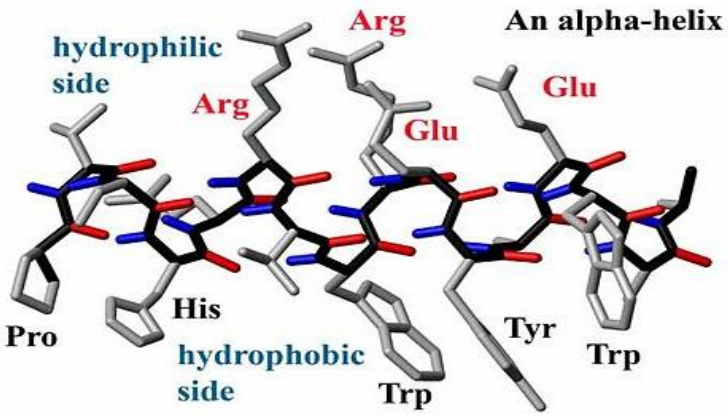


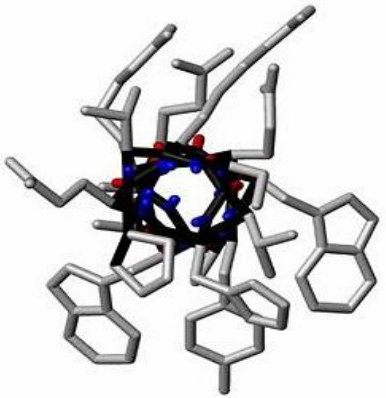
Alpha-helix:

- Right-handed helix
- 3.6 residues per helix turn
- Hydrogen bond between n and n+4
- ϕ = -57° ; ψ = -47° (right handed α helix);
- Linus Pauling & Robert Corey - 1951

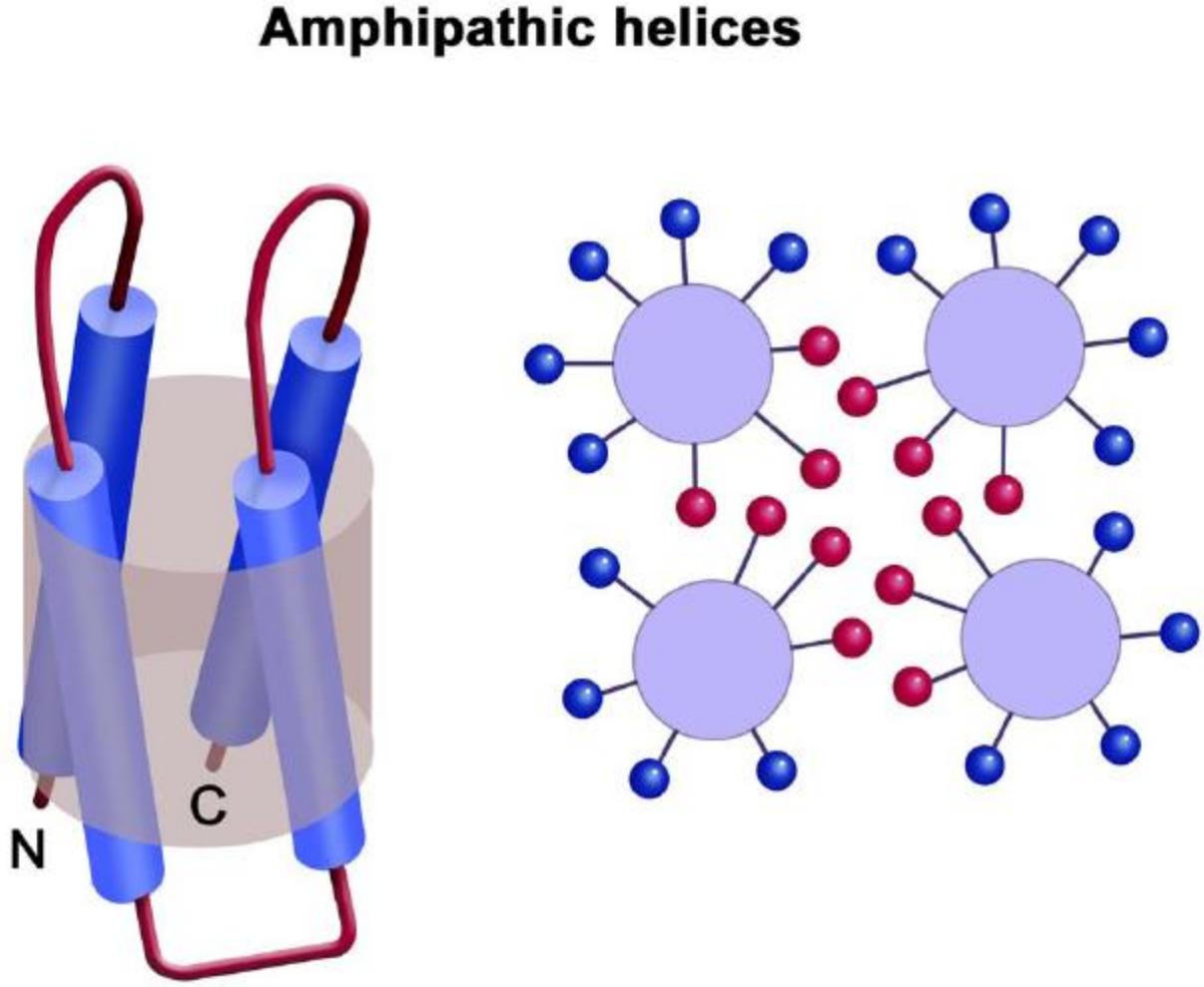
3_{10} helix

- Carbonyl (i) hydrogen bonds to amide (i+3)



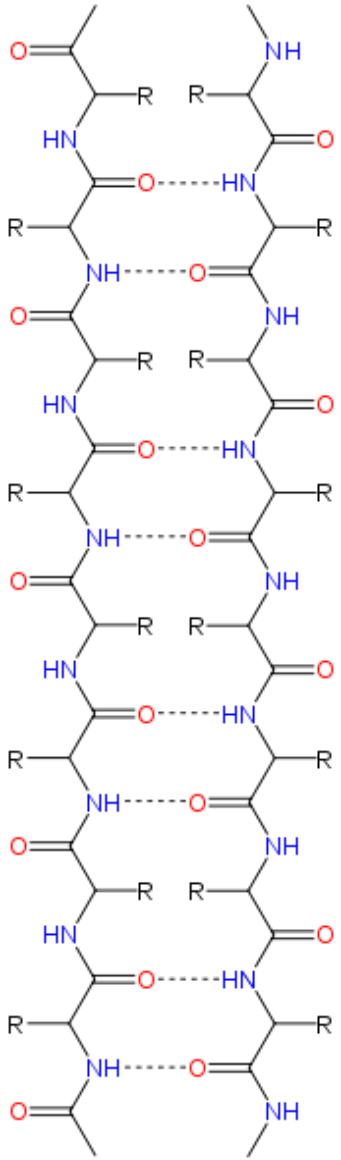


Amphipathic helices

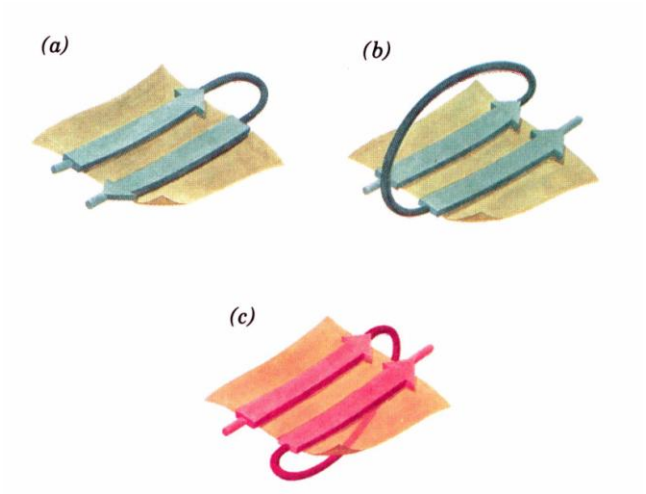


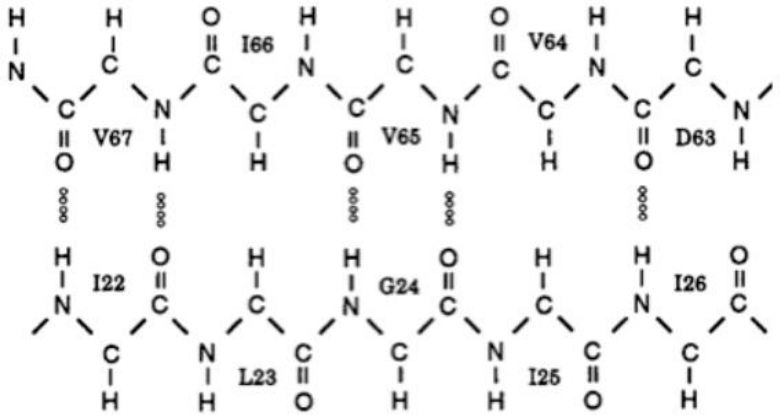
Michael Yaffe

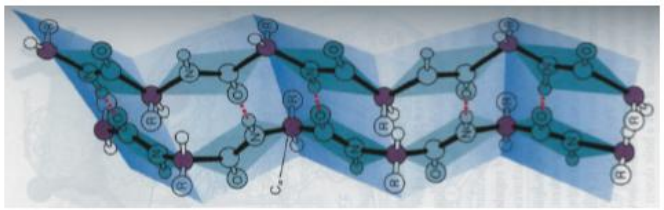
Protein Secondary Structure: Sheet

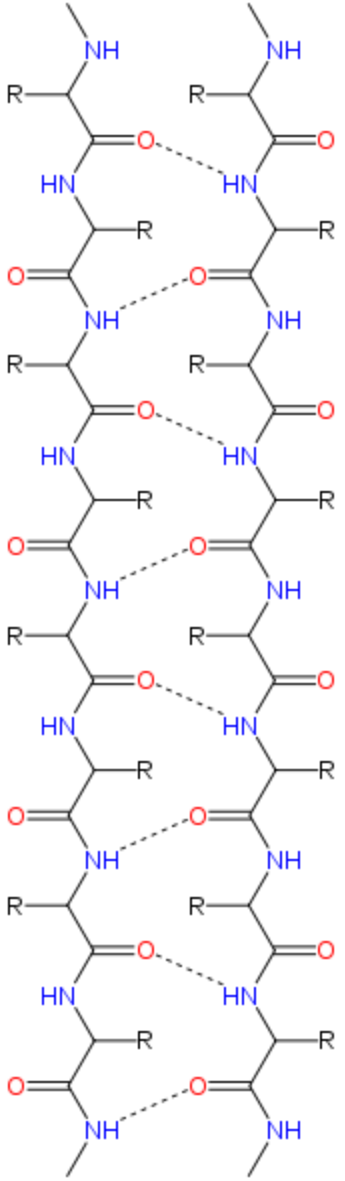


Anti-parallel





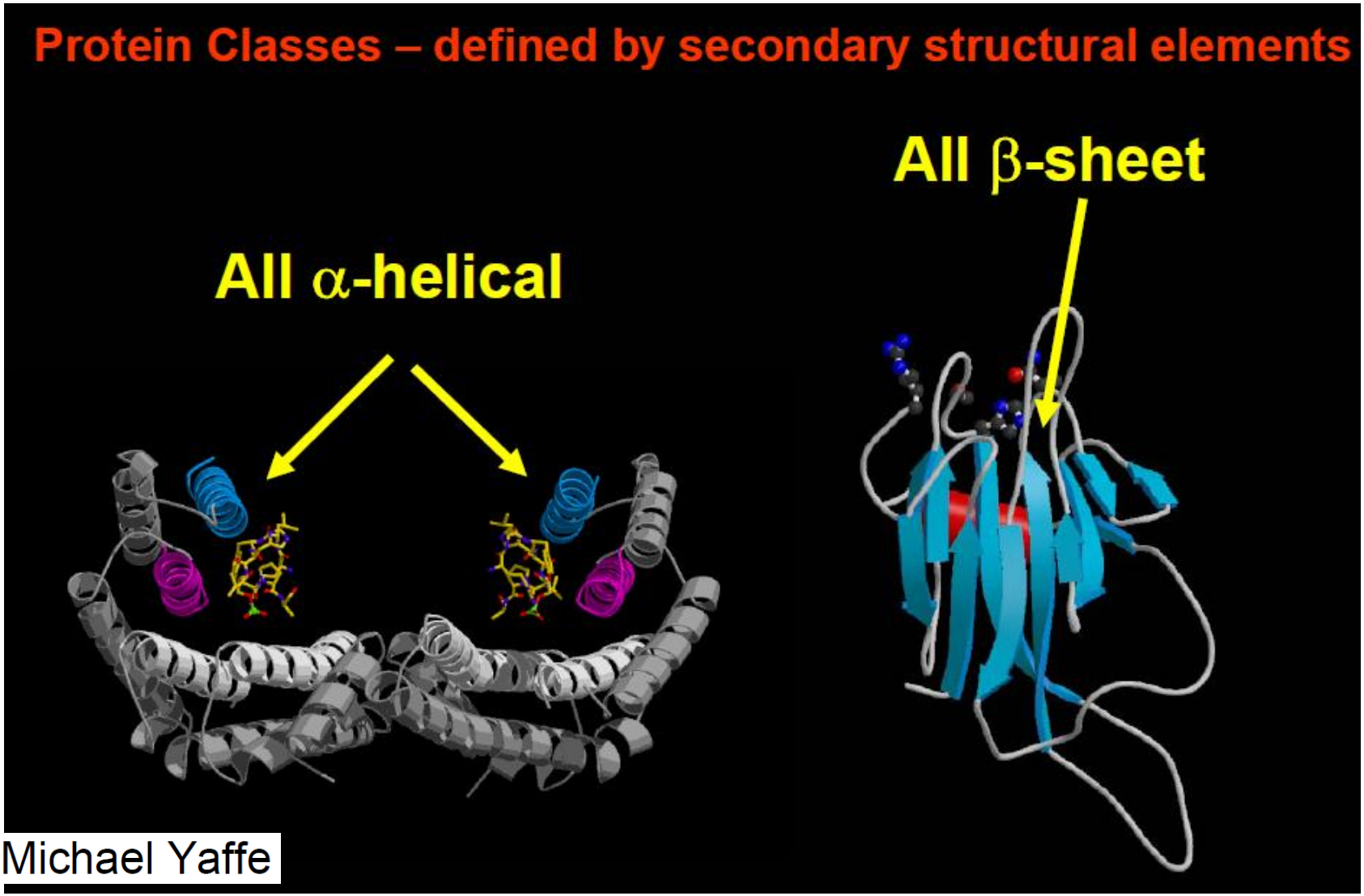




Parallel

Motifs and Domains: **Rossmann Fold / Zn finger / Leucine zipper**

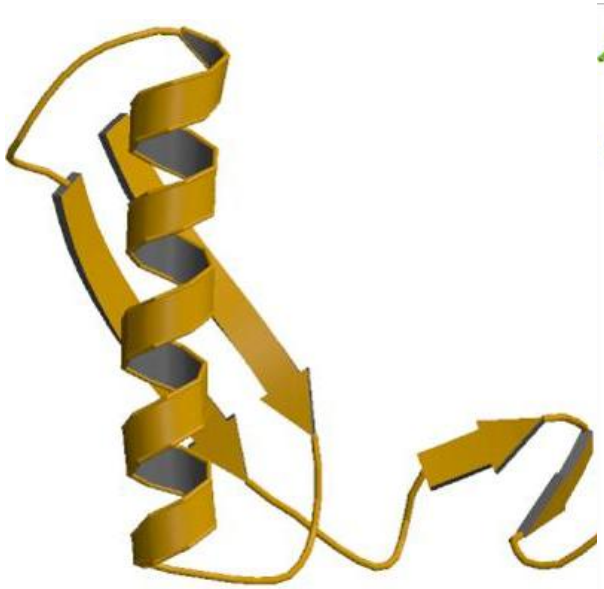
Tertiary Structure: 3D structure



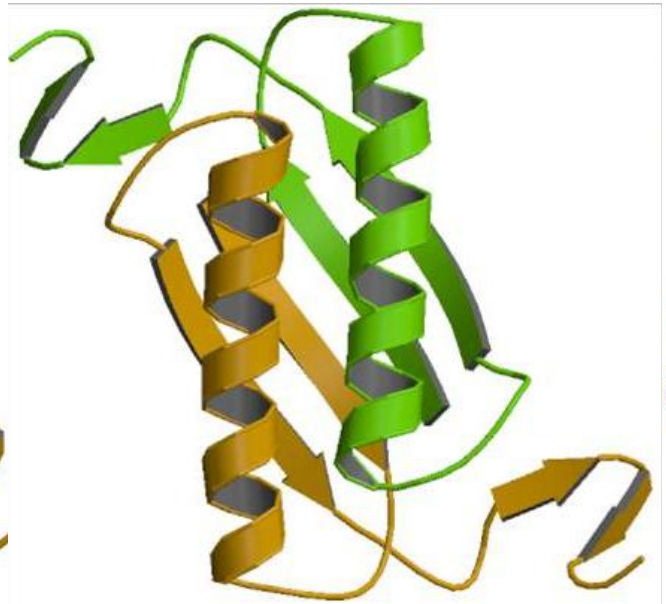
Quaternary Structure:

Arrangements of subunits in oligomers

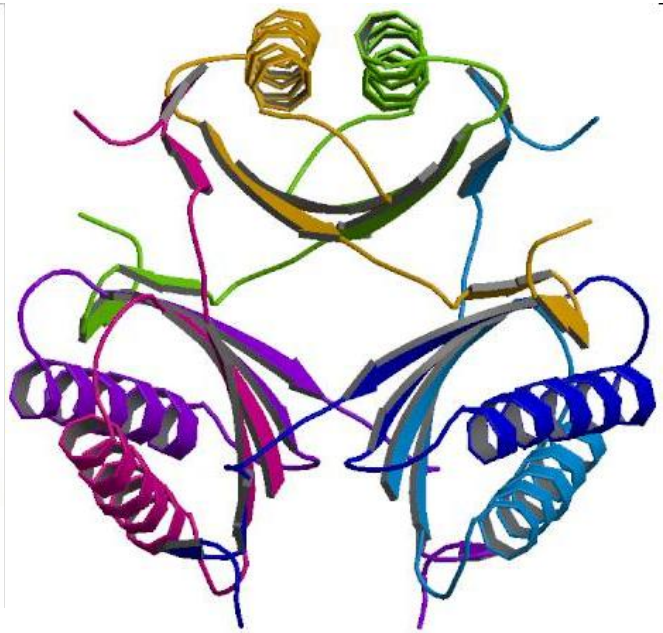
α_4 ; α_{12} ; $(\alpha\beta)_2$; $(\alpha\beta)_6$



monomer



dimer



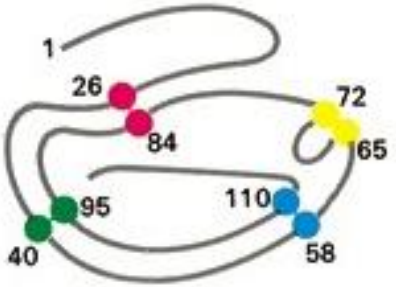
Hexamer

4-oxalocrotonate tautomerase

Denaturation (Non-native state): There are many denatured states of macromolecules. Denaturation can occur from many causes:

Denaturation : heat, high salt, hi & lo pH, organic solv., mechanical

- Tm (melting temperature)
- 8M Urea ; 5M quanidinium chloride ; 1% SDS
- Anions : sulfate > phosphate > Cl- > Br- > SCN-
- Cations: ammonium > Cs+ > K+ > Na+ > Li+ > Mg2+ > Ca2+ > Ba2+



Native ribonuclease

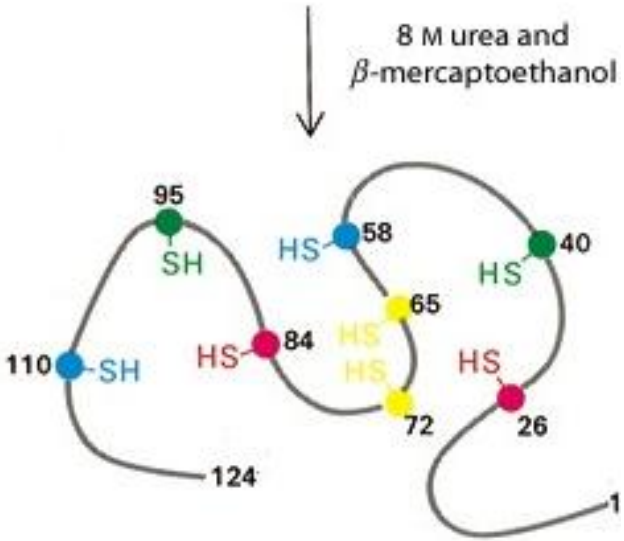
Renaturation :

Chris Anfinsen - Folding of Ribonuclease

124 a.a. + 4 disulfides

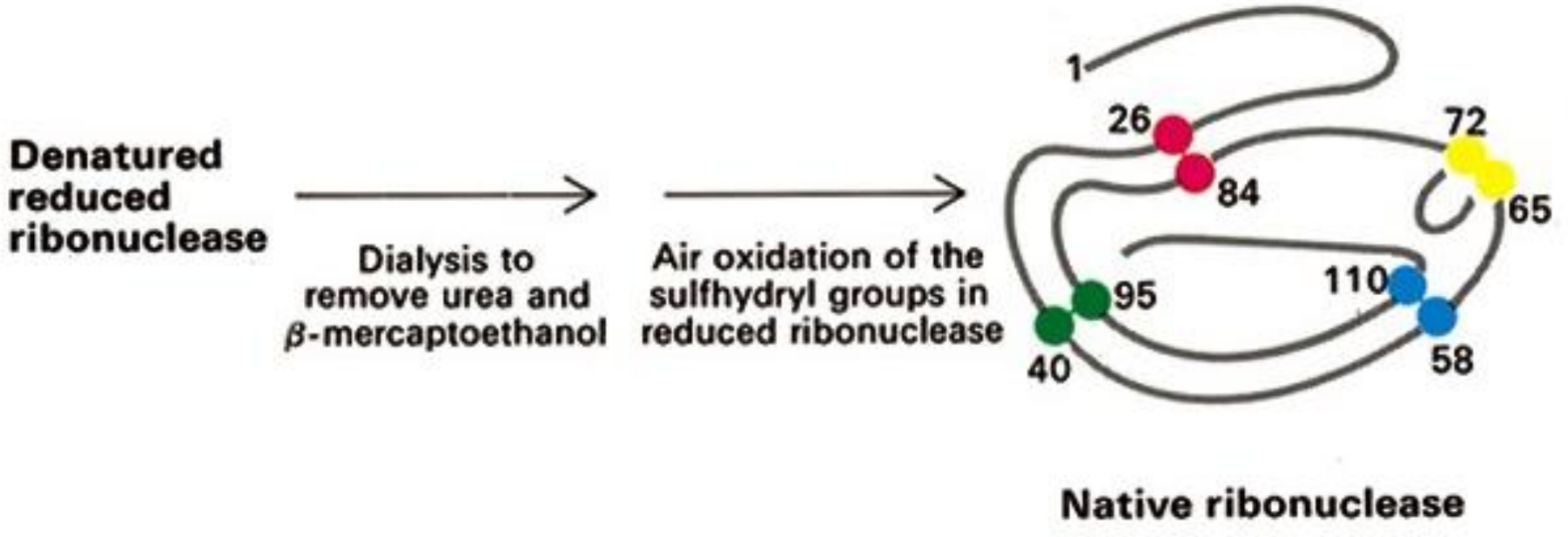
(26 → 84; 40 → 95; 58 → 110; 65 → 72)

(7 x 5 x 3 x 1 = 105 four disulfide combinations)



Denatured reduced ribonuclease

Chris Anfinsen - Folding of Ribonuclease (4 disulfides)

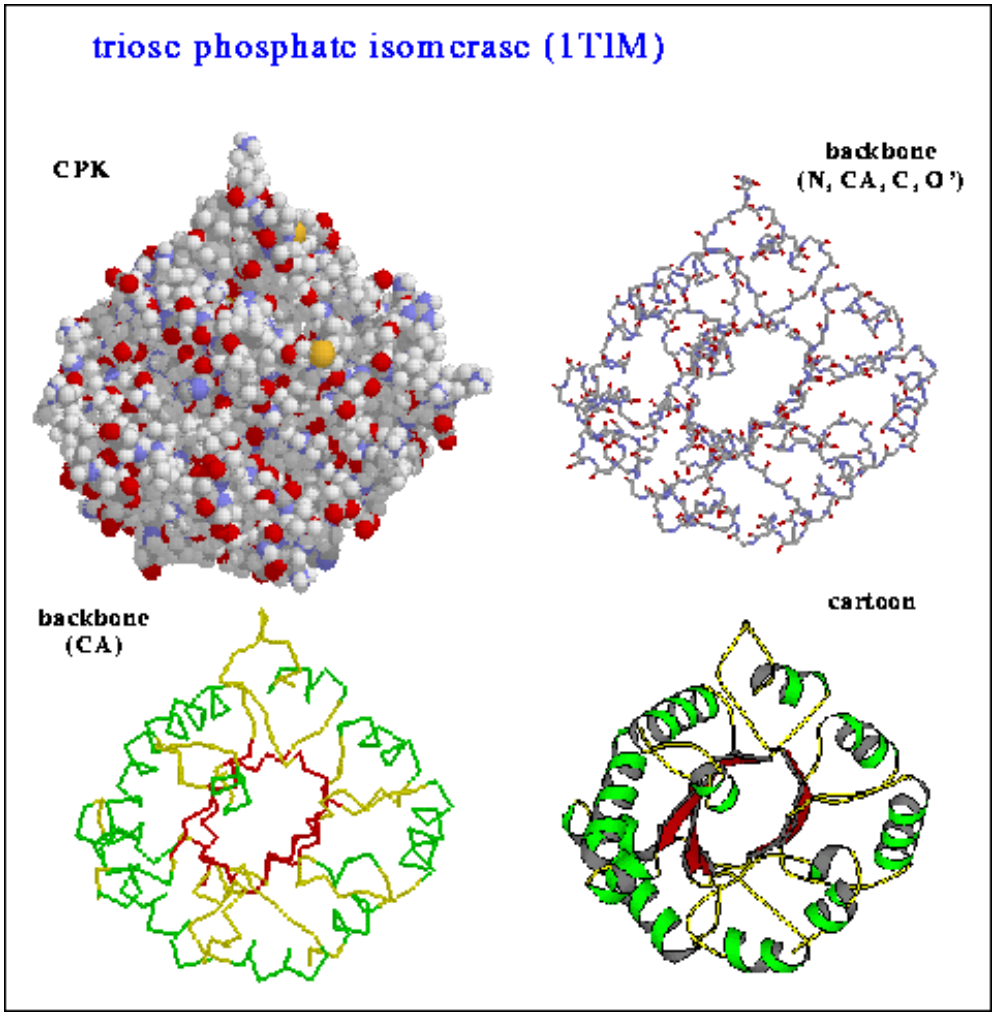


Conclusion: All the information necessary for folding the peptide chain into its native structure is contained in the primary amino acid sequence of the peptide.

Protein Structure

APRKFFVGGNWKMNGDKKSLG
ELIHTLNGAKLSADTEVVCGA
PSIYLDFARQKLDAKIGVAAQ
NCYKVPKGAFTGEISPAMIKD
IGAAWVILGHSERRHVFGESD
ELIGQKVAHALAEGLGVIACI
GEKLDEREAGITEKVVFEQTK
AIADNVKDWSKVVLAYEPVWA
IGTGKTATPQQAQEVHEKLRG
WLKSHVSDAVAQSTRIIYGGS
VTGGNCKELASQHDVDGFLVG
GASLKPEFVDIINAKH

=



Protein Folding: Stability / Denaturation and Renaturation

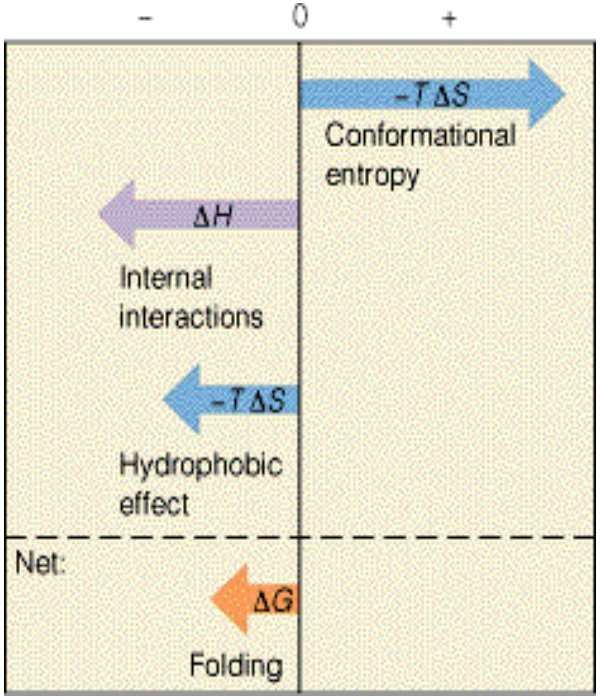
Protein folding (tertiary structure) is determined by weak interactions

H-bonds Dipole interactions Hydrophobic interactions Vander Waals forces Salt bridges

$$\Delta G = G_f - G_u = \Delta H_{prot} + \Delta H_{solv} - T\Delta S_{prot} - T\Delta S_{solv} \text{ (largest - } T\Delta S_{solv} \text{ for nonpolar R)}$$

Folding as a cooperative, sequential process : Local sec. st. / Domains / Molten globules

Molecular chaperones : (GroEL , GroES) assist with folding of some proteins



Protein	ΔG (kJ/mol)	ΔH (kJ/mol)	ΔS (J/K·mol)
Ribonuclease	−46	−280	−790
Chymotrypsin	−55	−270	−720
Lysozyme	−62	−220	−530
Cytochrome *c*	−44	−52	−27
Myoglobin	−50	0	+170

Note: Data adapted from P. L. Privalov and N. N. Khechinashvili, *J. Mol. Biol.* (1974) 86:665–684. Each data set has been taken at the pH value where the protein is maximally stable; all are near physiological pH. Data are for the folding reaction: Denatured $\rightleftharpoons$ native.

Force that destabilizes protein: Entropy

A folded protein is limited to a much smaller conformation space than an unfolded protein.

Consider backbone only

For an unfolded protein of 100 residues

Each residue: three possible ψ and three possible ϕ

$$S = R\ln W = R\ln 9^{99} = 1.8\text{kJ/mol}\cdot\text{K}$$

For a folded protein of 100 residues

Each residue: one possible ψ and one possible ϕ

$$S = R\ln W = R\ln 1^{99} = 0$$

$$\Delta G_{\text{conformation}} = -T\Delta S = 540\text{kJ/mol}$$

Forces that stabilize proteins:

H-bond

Ion-ion interaction

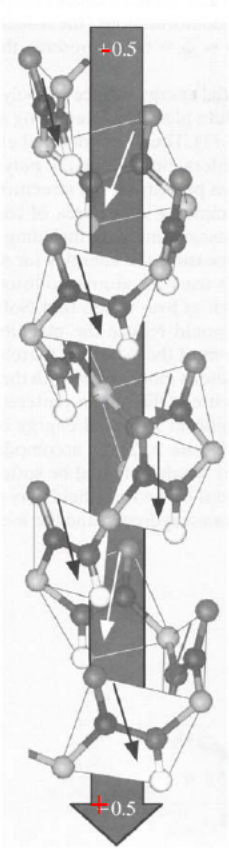
Dipole-related interaction

van de Waals interaction

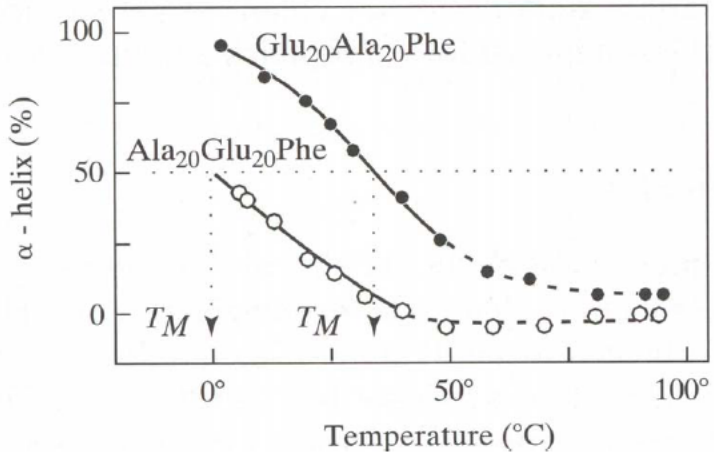
Hydrophobic interaction

Disulfide-bond

Dipole-related interaction



α-helix has a large dipole moment.
Direction of the dipole: C-terminus to N-terminus

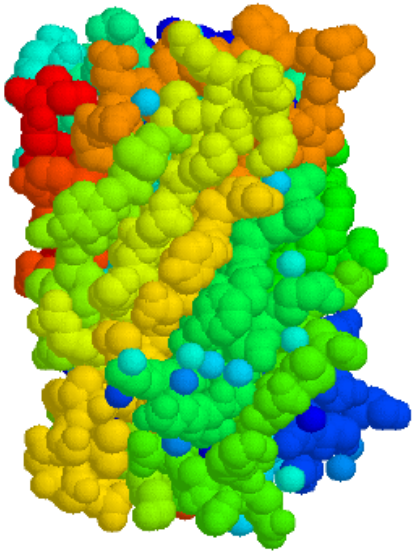


Ala: form a helix
Glu: negatively charge and disordered

$Glu_{20}Ala_{20}$ is stabilized by the dipole-charge interaction.

Protein folding

sequence



structure

- Proteins assume specific 3D structures.
- Protein structures are essential for their function.
- The protein structures are determined by their sequences.

The Levinthal Paradox (1969)

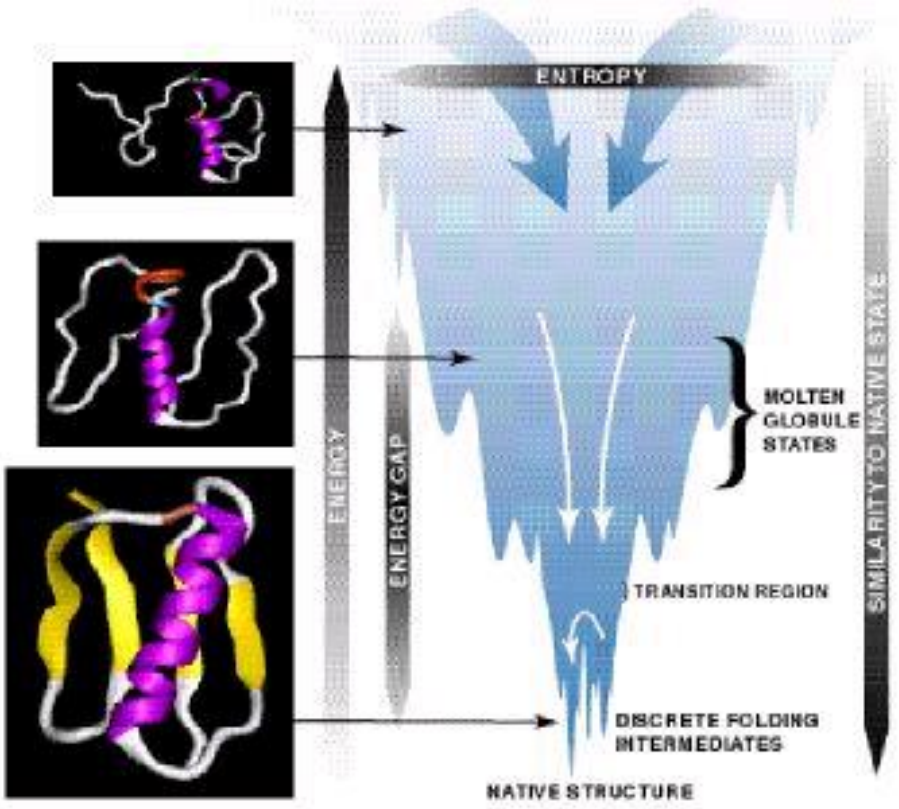
- There are too many possible conformations for a protein to fold by a random search.
- Consider just for the peptide backbone, there are at least 3 conformations per amino acid in the unfolded state, For a 100 a.a. protein we have 3^{100} conformations.
- If the chain can sample 10^{12} conformations/sec, it takes 5×10^{35} sec (2×10^{28} year)
- Conclusion: Protein folding is not random, must have pathways.

Protein Folding Landscape Theory

(Wolynes, Onuchic, Dill, Chan, Sali, Karplus, Brooks etc)

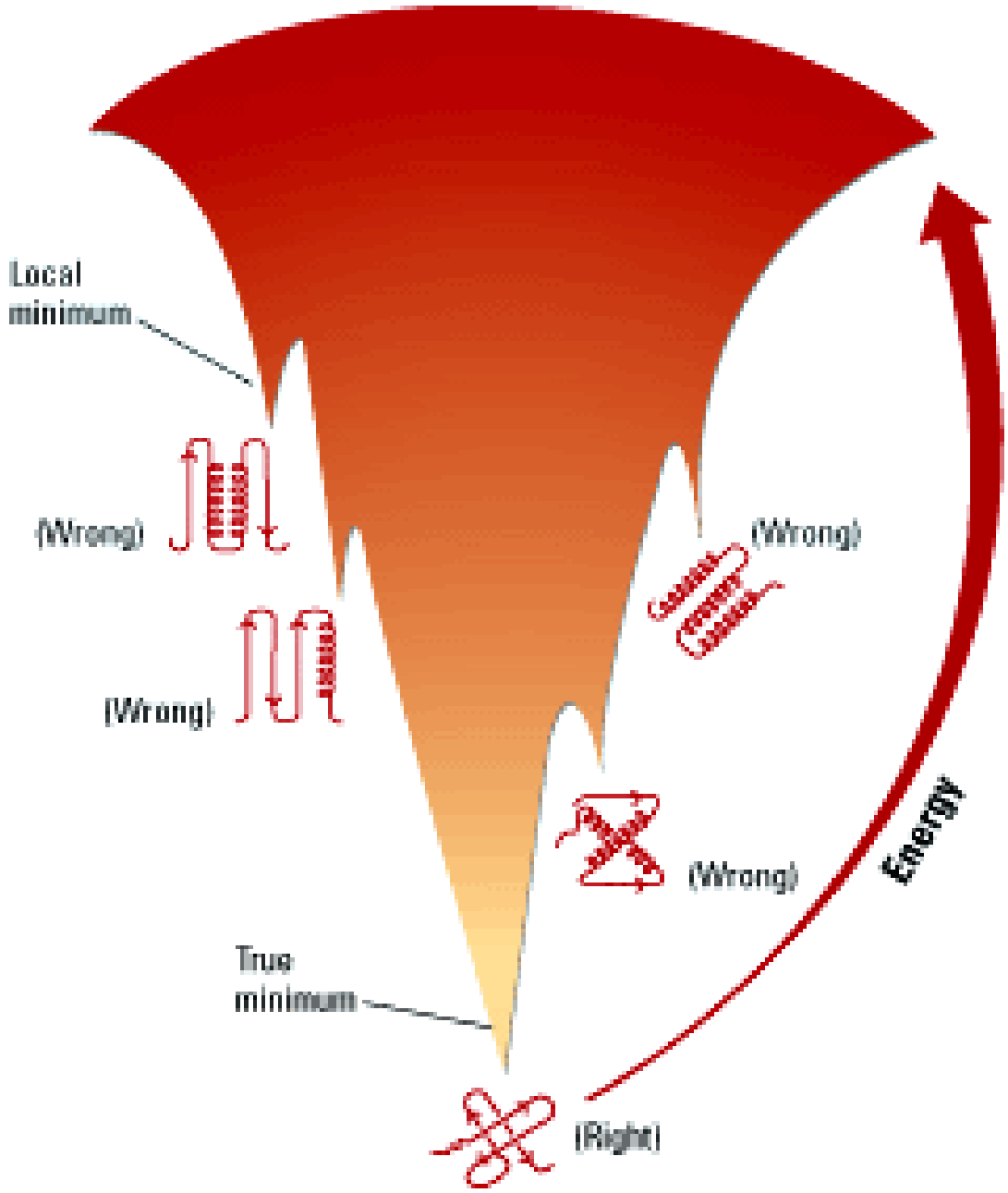
Proteins fold on timescales ranging from a microsecond to a few minutes, so they obviously drive or are driven quickly toward the native state.

- Folding can be described as the descent of the folding chain down a 'folding funnel,' with local roughness of the funnel reflecting the potential for transient trapping in local minima and the overall slope of the funnel representing the thermodynamic drive to the native state.
- A key notion is, in all but the final stages of folding, there exists an ensemble of structures (**molten globules**)--protein folding consequently occurs via **multiple pathways**.



So theoretically, if we have the protein sequence, we can know its structure and its function.

The transition state is composed of a broad ensemble of structures rather than one particular structure.



Chou-Fasman

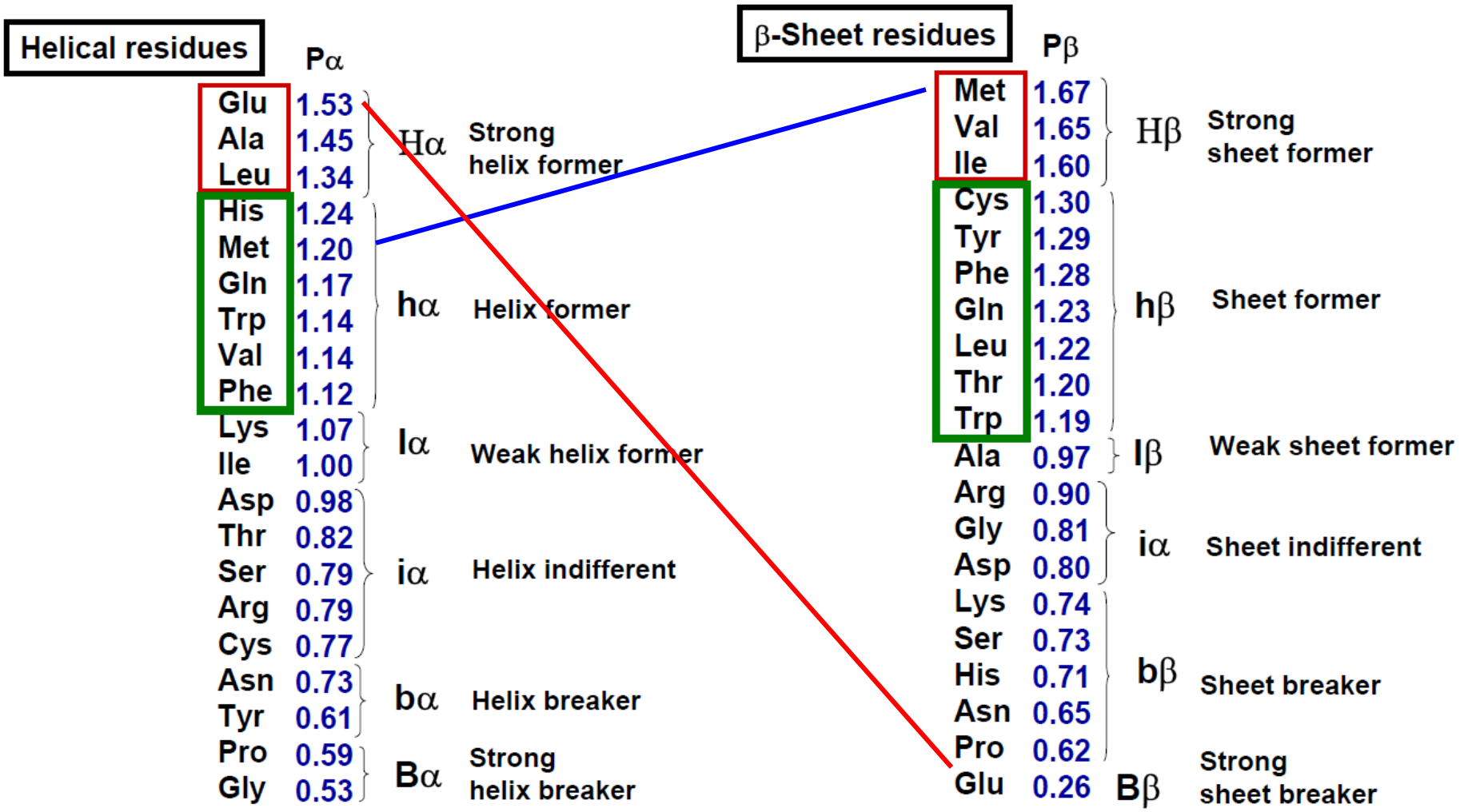
Biochemistry, *13*: 222-245, 1974

•Statistical Method

- Based on 15 proteins of known conformation, 2473 total amino acids

- Determined "protein conformational parameters" $P\alpha$, $P\beta$, based on $f_i^s / (\Sigma f_j^s / 20) \rightarrow$ 0.5-1.5

Michael Yaffe



Michael Yaffe

Chou-Fasman

Empirical rule set for secondary structure nucleation using <Pα>, <Pβ>

- Search for helical nuclei: locate clusters of 4 (Hα or hα) out of 6 residues. Unfavorable if > 1/3 (bα or Bα).
- Extend helical segments in both directions until tetrminated by tetrapeptides with <Pα><1.0. Helix breakers include b4, b3i, etc. Some of the tetrapeptide residues can be in the helical ends (except Pro).
- Refine boundaries: Pro, Asp, Glu prefer N-terminal end, His Lys, Arg prefer C-terminal end.
- **Rule #1 – Any segment ≥ 6 residues with <Pα>≥1.03 and <Pα>><Pβ>, satisfying above conditions is predicted as helical.**

Michael Yaffe

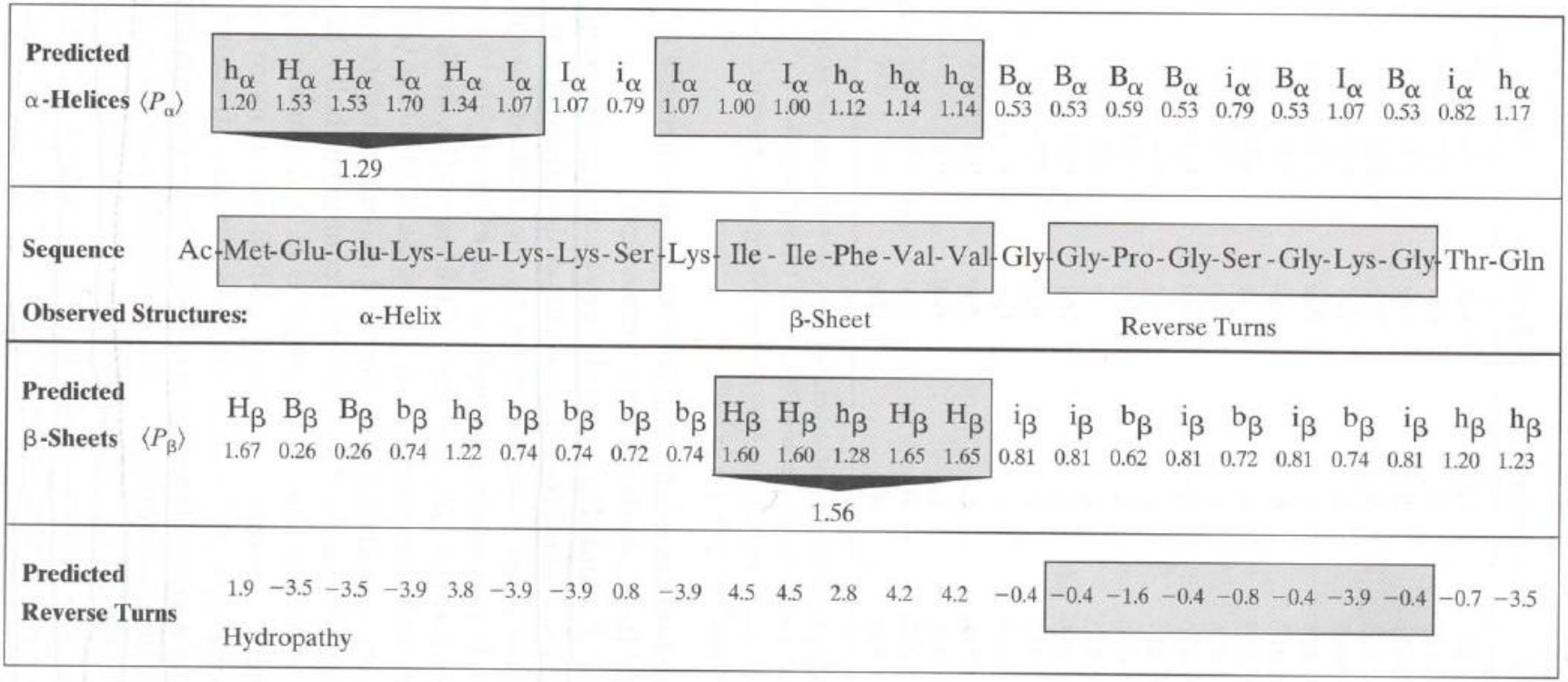
Chou-Fasman

Empirical rule set for secondary structure nucleation using <Pα>, <Pβ>

- Search for β-sheet nuclei: locate clusters of 3 β residues (Hβ or hβ) out of 5 residues. Unfavorable if > 1/3 β breakers (bβ or Bβ).
- Extend β-sheet segments in both directions until tetrminated by tetrapeptides with <Pβ><1.0. β-sheet breakers include b4, b3i, etc.
- Refine boundaries: Glu occurs rarely in β-region and Pro equally uncommon within inner β-sheets. Charged residues rare at either end. Trp most frequently at N-terminal end
- **Rule #2 – Any segment ≥ 5 residues with <Pβ>≥1.05 and <Pβ>><Pα>, satisfying above conditions is predicted as β-sheet.**

Michael Yaffe

Predict the secondary structure



The reliability of this prediction approach is only 70% because the tertiary structures are not taken into account.

http://www.courses.fas.harvard.edu/~chem163/

BIOINFORMATICS APPLICATIONS NOTE Vol. 14 no. 10 1998
Pages 892–893

JPred: a consensus secondary structure prediction server

James A. Cuff[1,2], Michele E. Clamp[2], Asim S. Siddiqui[1], Matt Finlay[1] and Geoffrey J. Barton[1,2]

Uses 6 different prediction methods: DSC, PHD, NNSSP, PREDATOR, MULPRED and ZPRED. Each method is run and the results are combined into a single, consensus structure prediction.

Predictions for request hmgl

```
                : 1---------11--------21--------31--------41--------51--------61--------71--------81--------91-
OrigSeq         : MAAMRKALPRRLVGLASLRAVSTSSMGTLPKRVKIVEVGPRDGLQNEKNIVSTPVKIKLIDMLSEAGLSVIETTSFVSPKWVPQMGDHTEVLK

dsc             : --------------------------------EEEEEHHHHHHHH---EEHHHHHHHHHHHHHHHHHHH-----------HHHHHHH--HH
jalign          : --------------------------------EEEEEEE-----------HHHHHHHHHHHHHH-----EEEE--------HHHHHHHHHHH
jfreq           : -HHHHHH---HHHHHHHH-------------EEEEEE---------------HHHHHHHHHHHH----EEEE--------------HHHHHH
jhmm            : --------------------------------EEEEE---------------HHHHHHHHHHHHH---EEEE-------HHHHHHHHHHHH
jnet            : -HHHHHH---HHHHH----------------EEEEEEE-------------HHHHHHHHHHHH----EEEE-------------HHHHHHH
jpssm           : -HHHHHHHHHHHH----EEEEHH-------EEEEEEE-------------HHHHHHHHHHHH---EEEE--------------HHHHHH
mul             : --------------------------------EEEEE------H-------HHHHHHHHHHHHH-HEEEE-----------HHHHHHH-HH
phd             : --------------------------------EEEEEE--------------HHHHHHHHHHHH---EEEEE--------HHHHHHHHHEE-
pred            : ----------HHHHHH---------------EEEEEE-------------HHHHHHHHHHHHH----EEEE-----------HHHHHHHHH

Jpred           : --------------------------------EEEEEE--------------HHHHHHHHHHHH----EEEE-----------HHHHHHHHH


PHDHtm          : ------------------------------------------------------------------------------------------
MCoil           : ------------------------------------------------------------------------------------------
MCoilDI         : ------------------------------------------------------------------------------------------
MCoilTRI        : ------------------------------------------------------------------------------------------
Lupas 21        : ------------------------------------------------------------------------------------------
Lupas 14        : ------------------------------------------------------------------------------------------
Lupas 28        : ------------------------------------------------------------------------------------------

PHDacc          : -----BBBBBB-B-U-BB-BBBBBBBB-B--BBBBB-BBB-----B--BBB-B---B-BB-BB--BBB-BB-BBB-BBB--BB-BB--BBB--
Jnet_25         : ---B---B---BBBB-BB--------------B-BBBBBBBBB-BBB-B-B-B--BB-BB-BBB-BBB-BBBBBBBBBB-BBB-BB---BBB-
Jnet_5          : --------------------------------B-B-BBB---B------------B--BB--B---BB--BBB-BBB----B-----------
Jnet_0          : ----------------------------------B-------B----------------B--B-------------------------------

PHD Rel         : 998998888777777778778876666778727998514567766664323456689999999855982799615557852234565432102
Pred Rel        : 00707706705886705657550790078986908966667877777687556889999999988699688658789899667867888888
Jnet Rel        : 878988614213310011124534156888747999704688886778736588999999998782798186357765433101221589999
```

Ab initio Prediction of Protein Structure

- Need to find a potential function where

  $E(S, C_{native}) < E(S, C_{non\text{-}native})$.

- Need to construct an algorithm to find the global minimum of this function.

  Still an unsolved, computationally demanding problem

******

→ **Homology Modeling**

→ **BLAST / PDB**

(find related proteins whose structures are known)

CASP 10

Author	Message
TJ Forum moderator Project administrator Project developer Project scientist Send message Joined: Oct 22 10 Posts: 4 ID: 398718 Credit: 216,670 RAC: 0	Message 72941 - Posted 30 Apr 2012 19:37:35 UTC Hello everyone ! CASP 10, a community wide experiment in structure prediction starts tomorrow on May 1st and runs to August 1st. During this time we will be using BOINC heavily for structure prediction. If your work unit starts with the label rb you're running a CASP 10 target! rb is short for Robetta which is our publicly available server for structure prediction. CASP CASP is an international experiment to assess the state-of-the-art of the protein structure prediction field. Sequences, whose structures have been solved but which have not yet been published are sent out to participating teams and we have a 3 days to send back predictions. The whole thing is conducted in a double-blind fashion ensuring fair assessment and truly blind prediction. Robetta Structure prediction for the community, by the community. Robetta is a server for protein structure prediction that shares Rosetta's structure prediction capabilities to the scientific community (and to the public). The computation for this will be conducted on BOINC meaning that you guys will be crunching

Protein Structure Prediction Center

Menu

- Home
- FORCASP Forum
- PC Login
- PC Registration
- CASP Experiments
  - CASP ROLL
    - Home
    - My CASP ROLL profile
    - Targets
      - Target List
      - Target Submission
  - CASP10 (2012)
    - Home
    - My CASP10 profile
    - Targets
    - Meeting registration
    - CASP10 in numbers

Welcome to the Protein Structure Prediction Center!

Our goal is to help advance the methods of identifying protein structure from sequence. The Center has been organized to provide the means of objective testing of these methods via the process of blind prediction. The Critical Assessment of protein Structure Prediction (CASP) experiments aim at establishing the current state of the art in protein structure prediction, identifying what progress has been made, and highlighting where future effort may be most productively focused.

There have been nine previous CASP experiments. The tenth experiment is planned to start in April 2012. Description of these experiments and the full data (targets, predictions, interactive tables with numerical evaluation results, dynamic graphs and prediction visualization tools) can be accessed following the links:

CASP1 (1994) | CASP2 (1996) | CASP3 (1998) | CASP4 (2000) | CASP5 (2002) | CASP6 (2004) | CASP7 (2006) | CASP8 (2008) | CASP9 (2010) | CASP10 (2012)

Raw data for the experiments held so far are archived and stored at our data archive.

Starting November 2011, we are opening a new rolling CASP experiment for all-year-round testing of ab initio modeling methods:

CASP ROLL

Details of the experiments have been published in a scientific journal *Proteins: Structure, Function and Bioinformatics*. CASP proceedings include papers describing the structure and conduct of the experiments, the numerical evaluation measures, reports from the assessment teams highlighting state of the art in different prediction categories, methods from some of the most successful prediction teams, and progress in various aspects of the modeling.

Message Board

Sep.6 - early bird registration deadline; CASP fellowships letters

Dear CASP participants, All recipients of CASP student fellowships have been identified and notified. Please proceed with the registrations according to the instructions provided in the award let ...

Abstract collection; meeting fellowships; early bird registration

1. Just three days ago we received the last CASP10 prediction, and today we start collecting methods abstracts. The Abstract Submission web page is available through a link from the CASP10 main page. ...

End of CASP10 regular