Post-translational Modification

Biological diversity / the Proteome

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

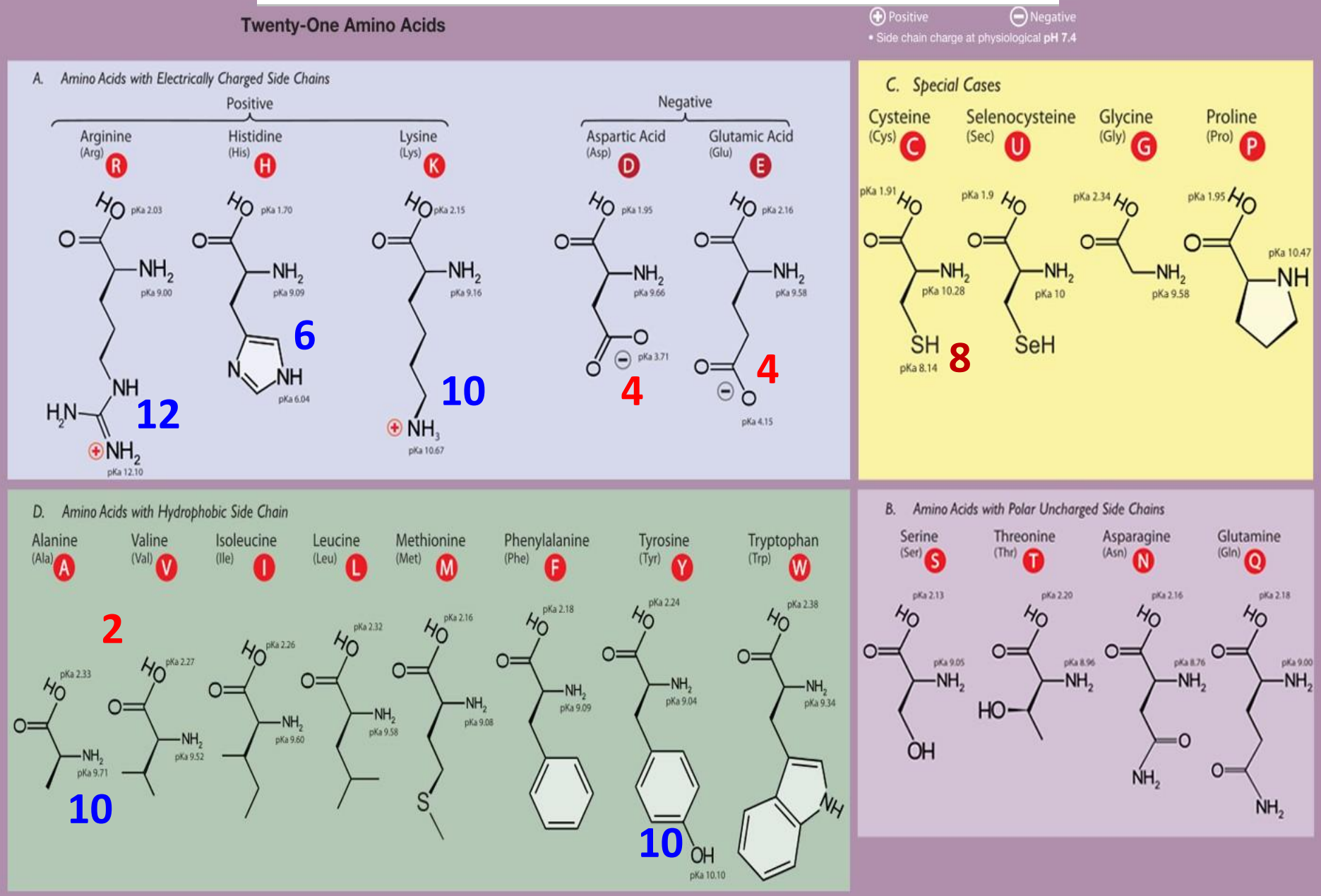
Protein Folding / Unfolding (denaturation)

- **Energetics / Intra and Intermolecular forces**

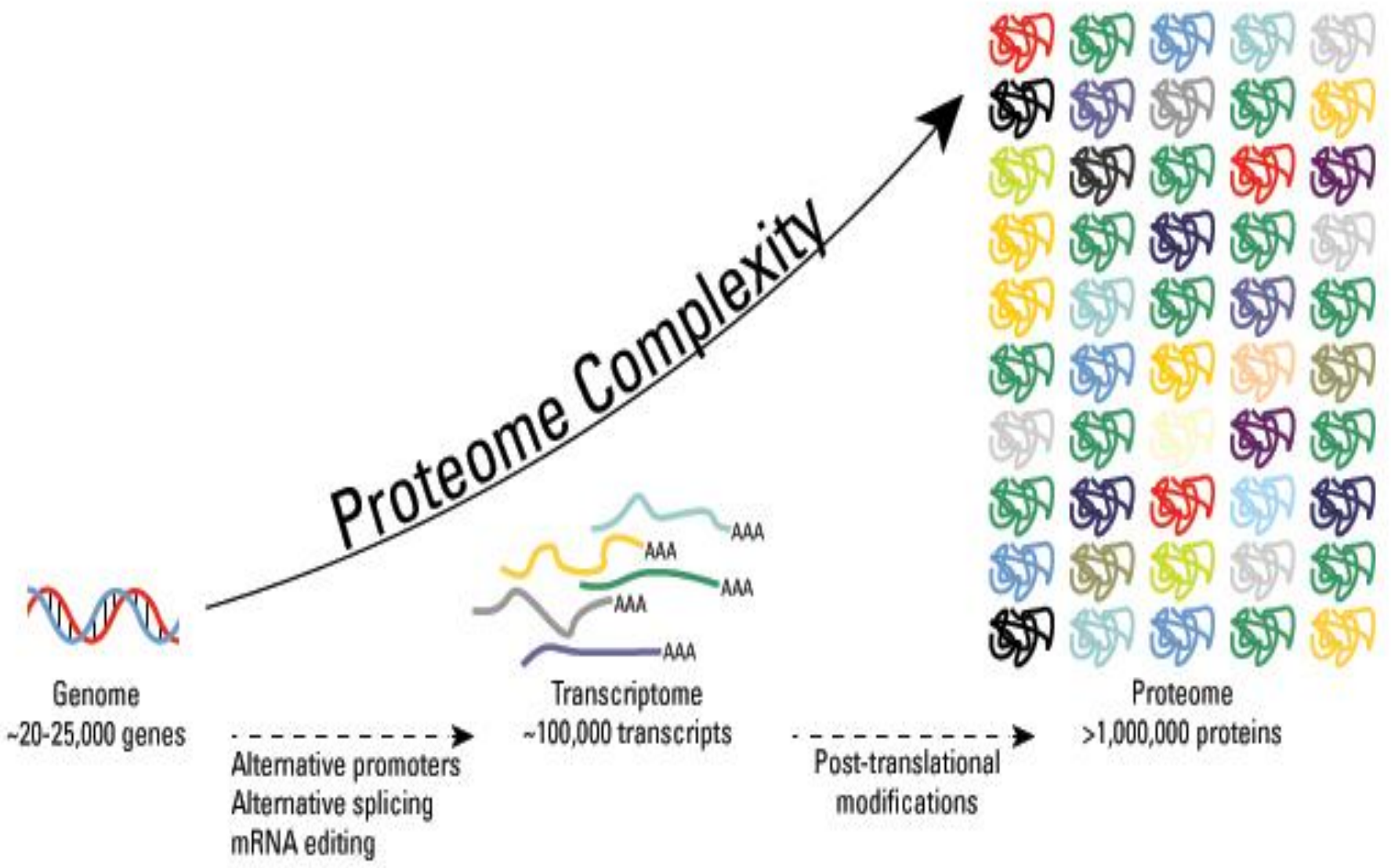
Prediction of Protein Structure

- **Early method – Chou / Fasman**
- **CASP / Rosetta**

Amino Acid Abbreviations



Biological Complexity: Genome → Transcriptone → Proteome

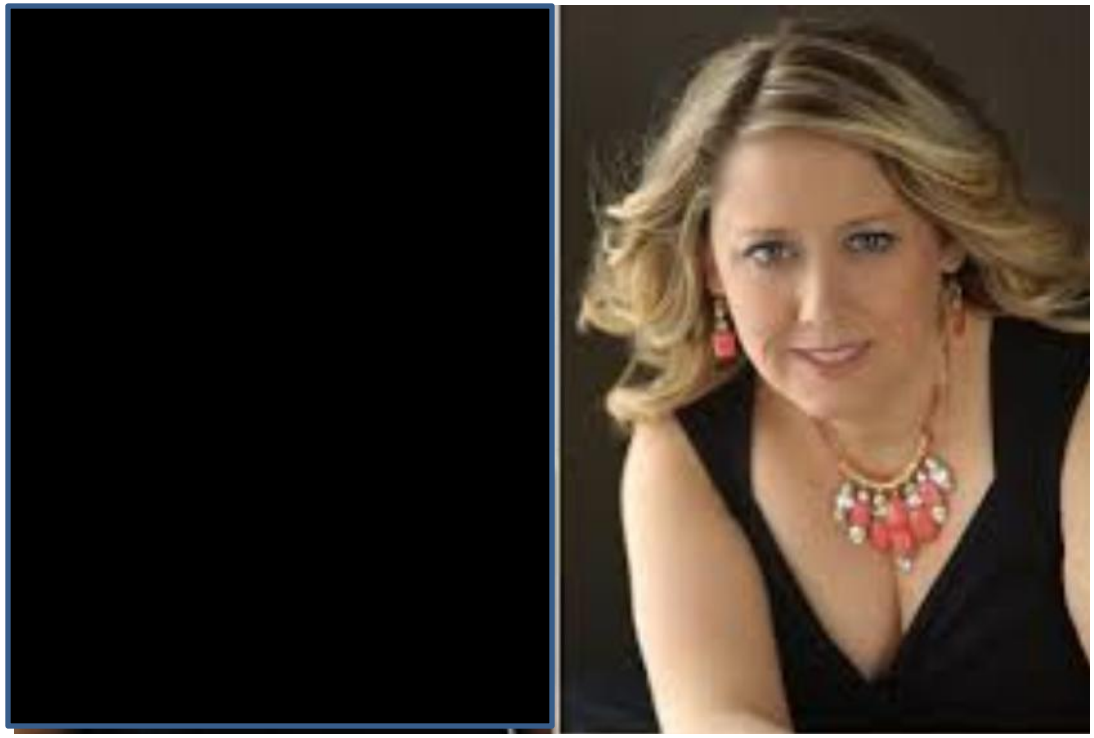


Post-translational Modification

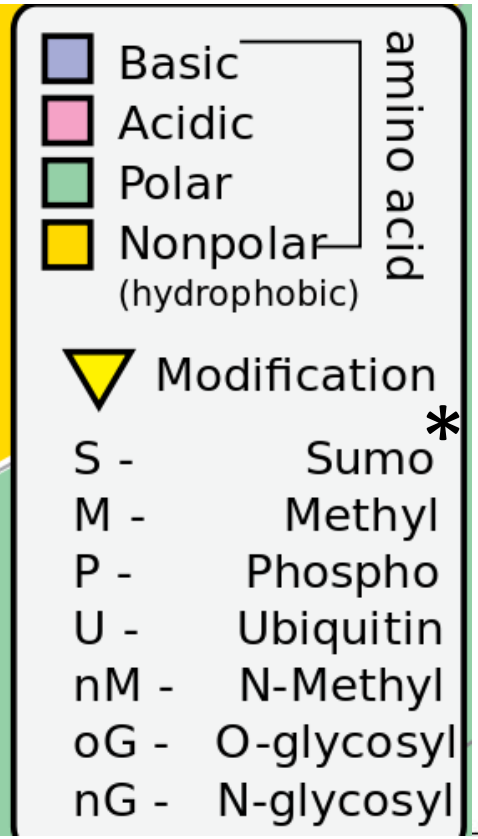
a.k.a. "Glamour Shots"

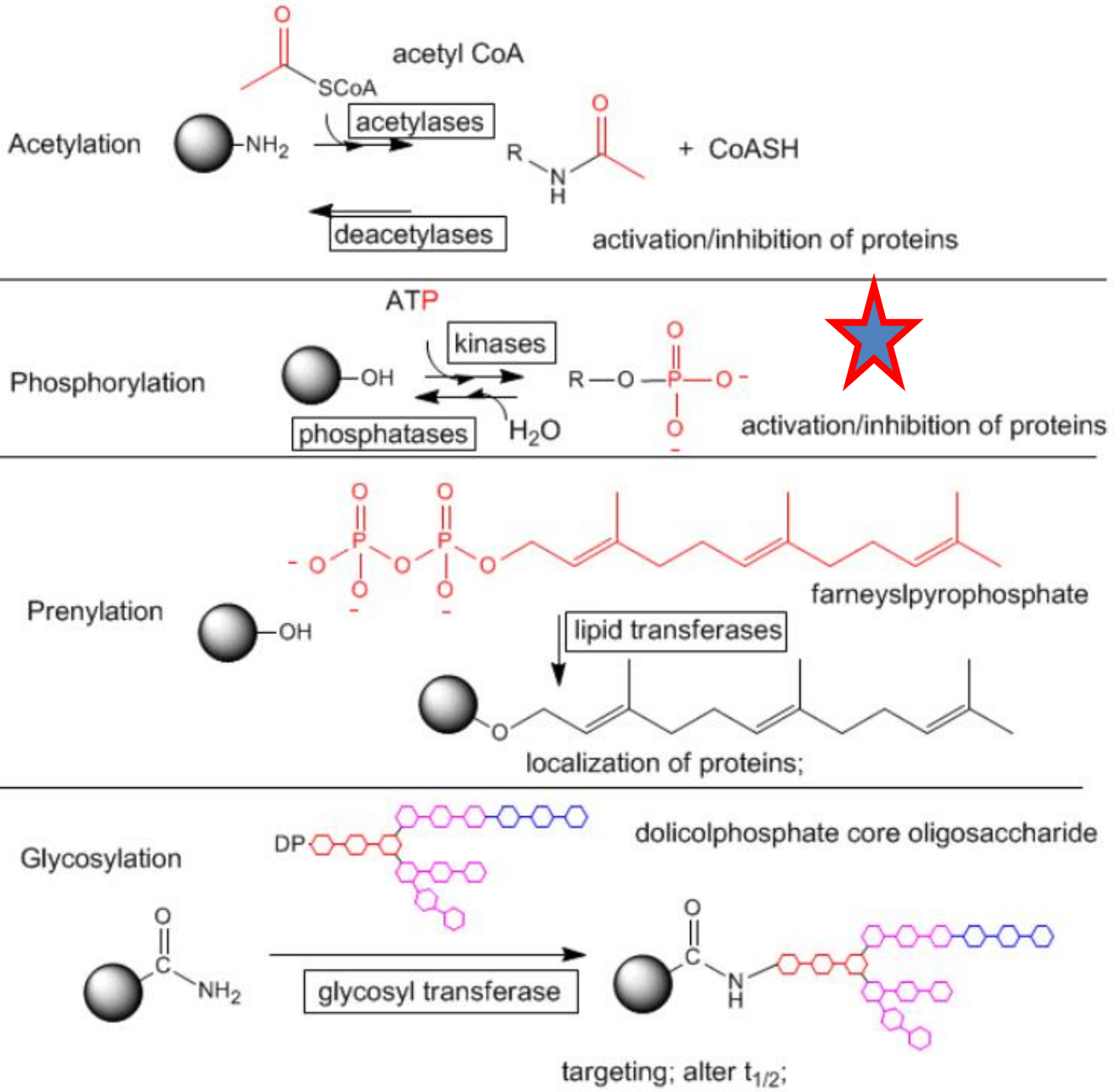
Post-translational Modification

a.k.a. "Glamour Shots"

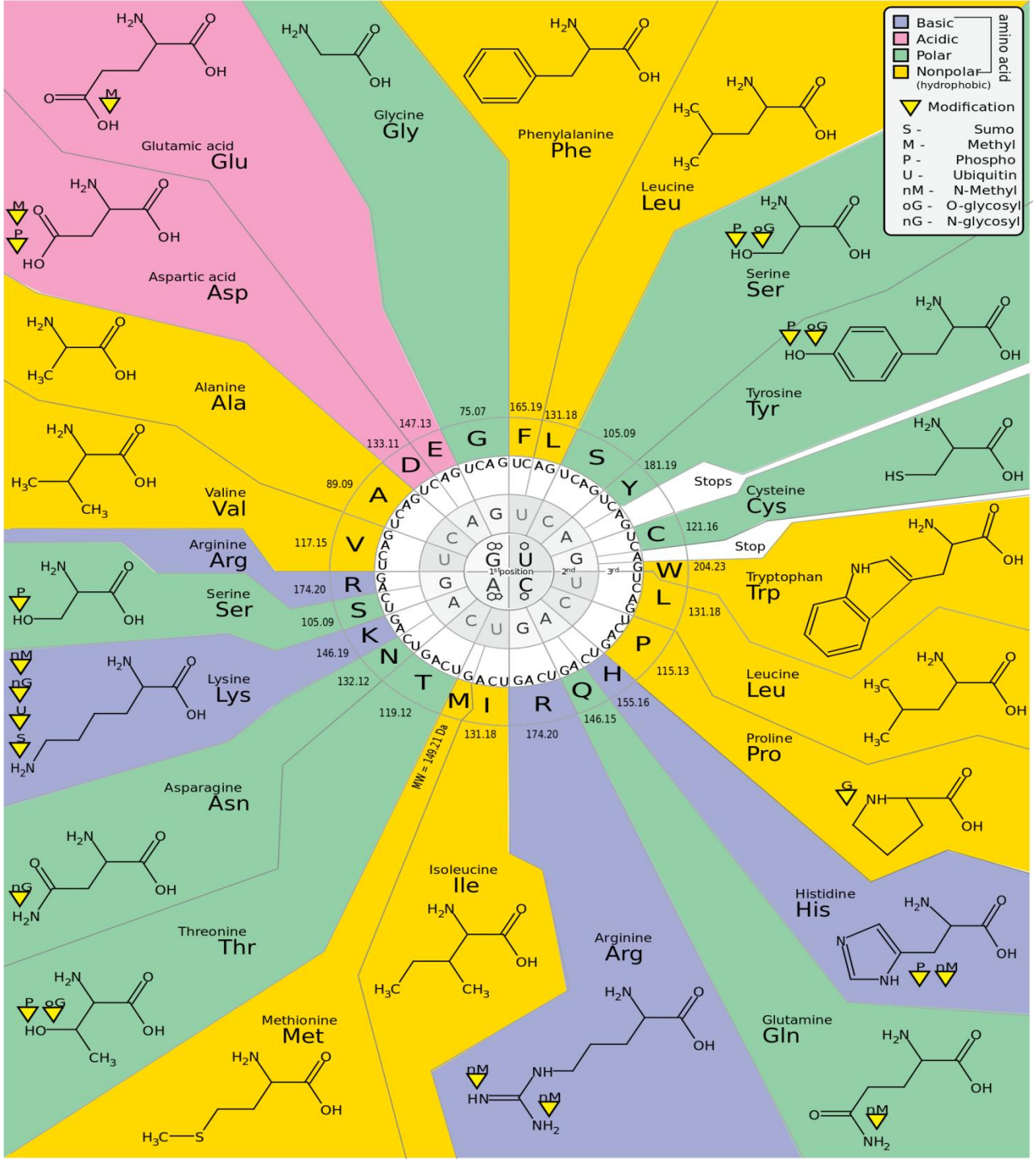


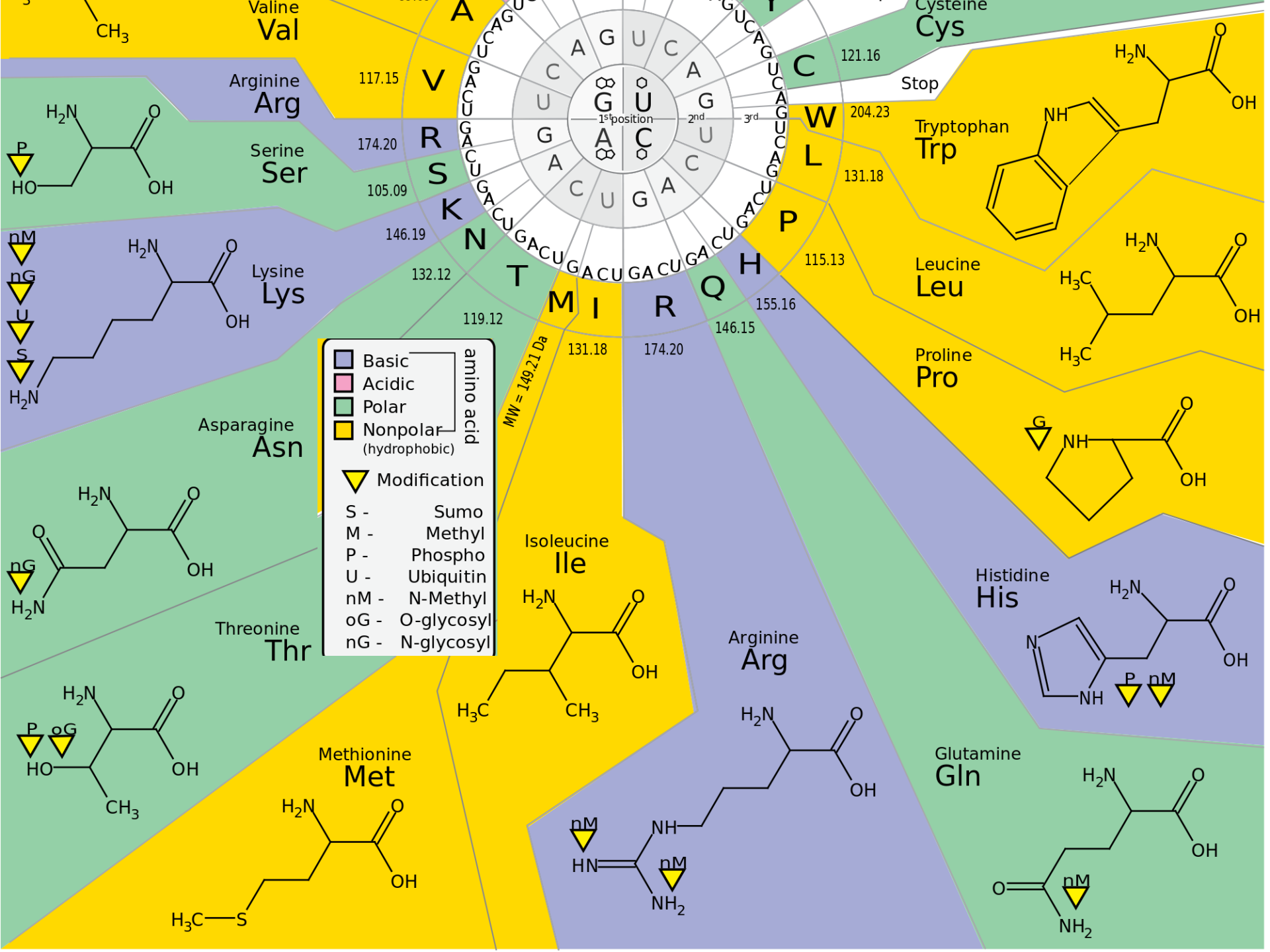
Post-translational Modification





* SUMO = *small ubiquitin-related modifier protein*


amino acid
Basic
Acidic
Polar
Nonpolar
(hydrophobic)
Modification
S - Sumo
M - Methyl
P - Phospho
U - Ubiquitin
nM - N-Methyl
oG - O-glycosyl
nG - N-glycosyl
Glutamic acid
Glu
Glycine
Gly
Phenylalanine
Phe
Leucine
Leu
Aspartic acid
Asp
Serine
Ser
Alanine
Ala
Tyrosine
Tyr
Valine
Val
Stops
Cysteine
Cys
Arginine
Arg
Stop
Tryptophan
Trp
Serine
Ser
Lysine
Lys
Leucine
Leu
Proline
Pro
Asparagine
Asn
Histidine
His
Isoleucine
Ile
Threonine
Thr
Arginine
Arg
Glutamine
Gln
Methionine
Met
1st position
2nd
3rd
MW = 149.21 Da
147.13
75.07
165.19
131.18
133.11
105.09
181.19
89.09
121.16
117.15
204.23
174.20
131.18
105.09
146.19
115.13
132.12
155.16
119.12
146.15
131.18
174.20


Valine
Val
Arginine
Arg
Serine
Ser
Lysine
Lys
Asparagine
Asn
Threonine
Thr
Methionine
Met
Isoleucine
Ile
Arginine
Arg
Histidine
His
Glutamine
Gln
Proline
Pro
Leucine
Leu
Tryptophan
Trp
Cysteine
Cys
Stop
1st position
2nd
3rd
117.15
174.20
105.09
146.19
132.12
119.12
131.18
174.20
155.16
146.15
115.13
131.18
204.23
121.16
MW = 149.21 Da
Basic
Acidic
Polar
Nonpolar
(hydrophobic)
amino acid
Modification
S - Sumo
M - Methyl
P - Phospho
U - Ubiquitin
nM - N-Methyl
oG - O-glycosyl
nG - N-glycosyl

Proteins: Biological function depends on conformation

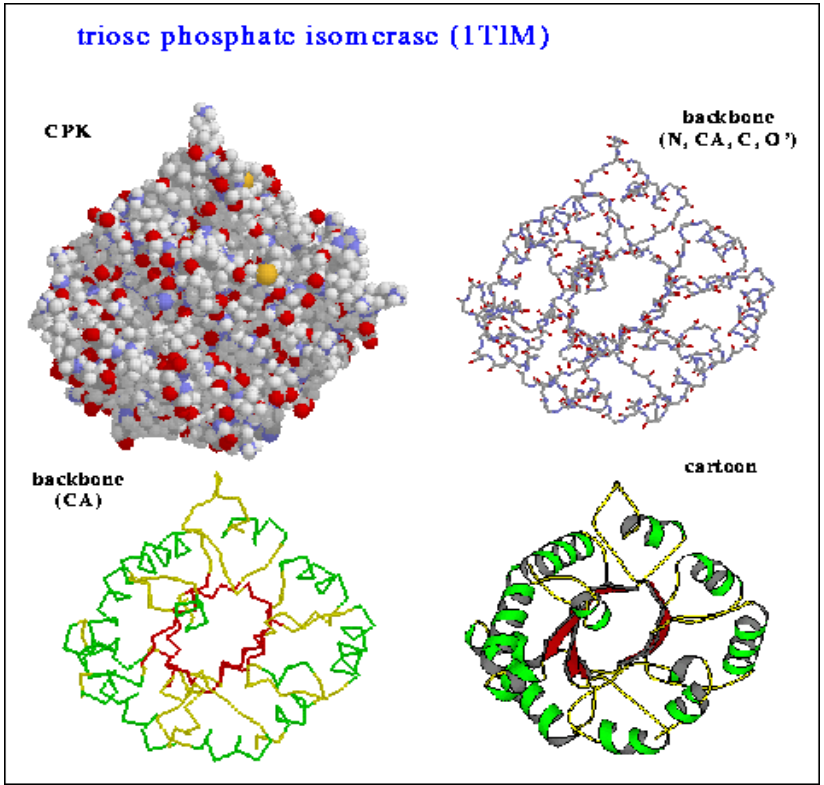
Unique Primary Structure = Unique 3D Structure ??
(Covalent bonds) (Noncovalent Interactions)

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

=

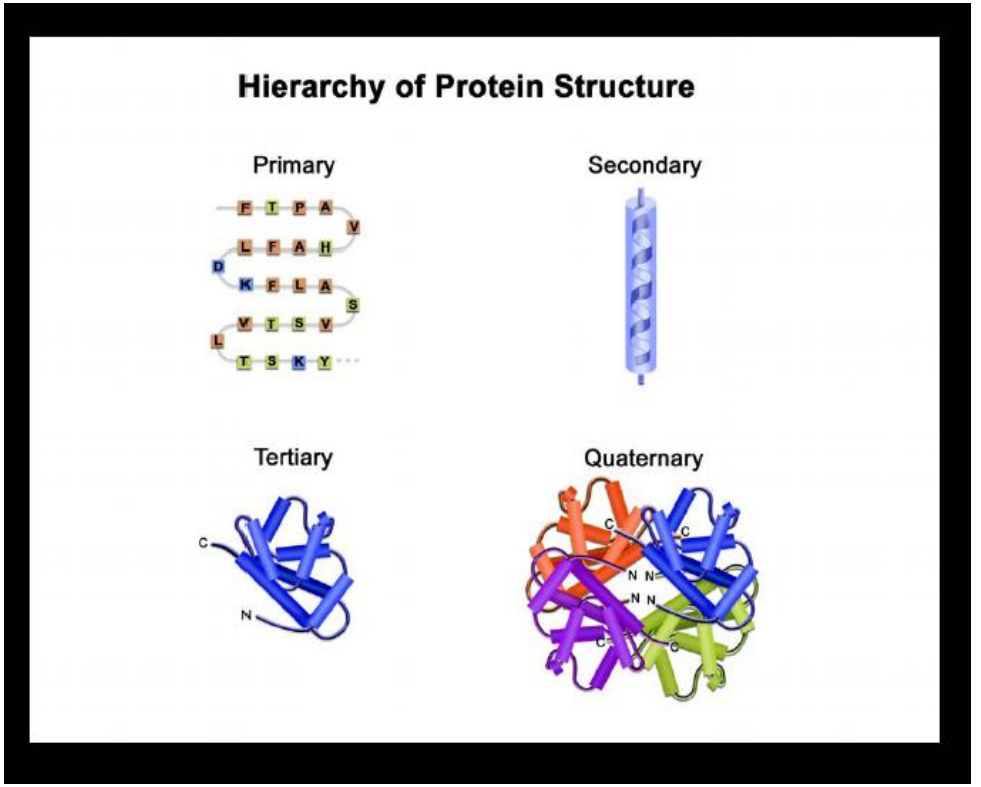


Four Levels of Description of (Native) Protein Structure

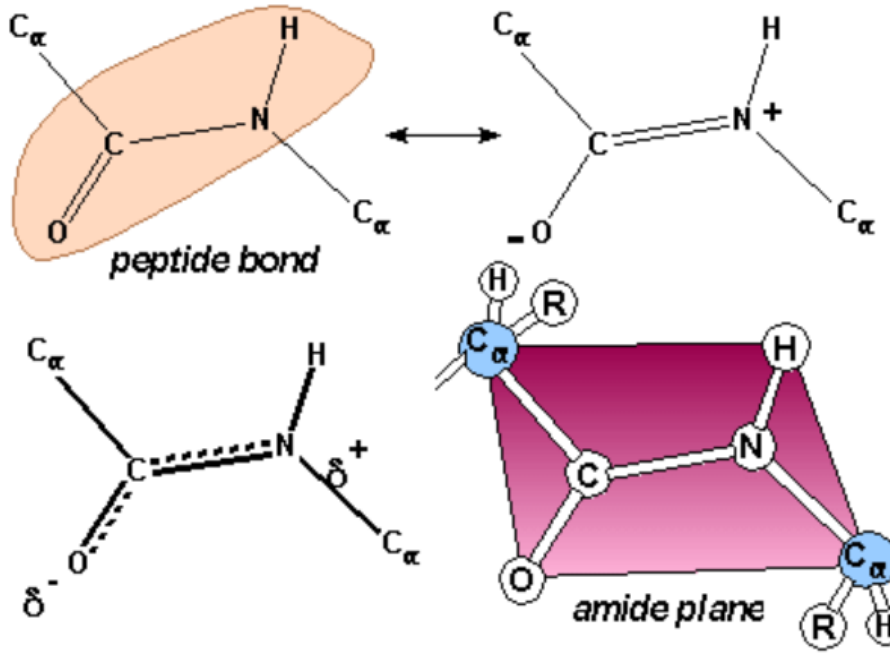
- **Primary Structure**: (~60-1000 amino acid residues) linear seq. of amino acid residues, covalent bonding including -SS- (also called "covalent structure“)
  (*the **primary structure** of a biological molecule is the exact specification of its atomic composition and the chemical bonds connecting 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

****

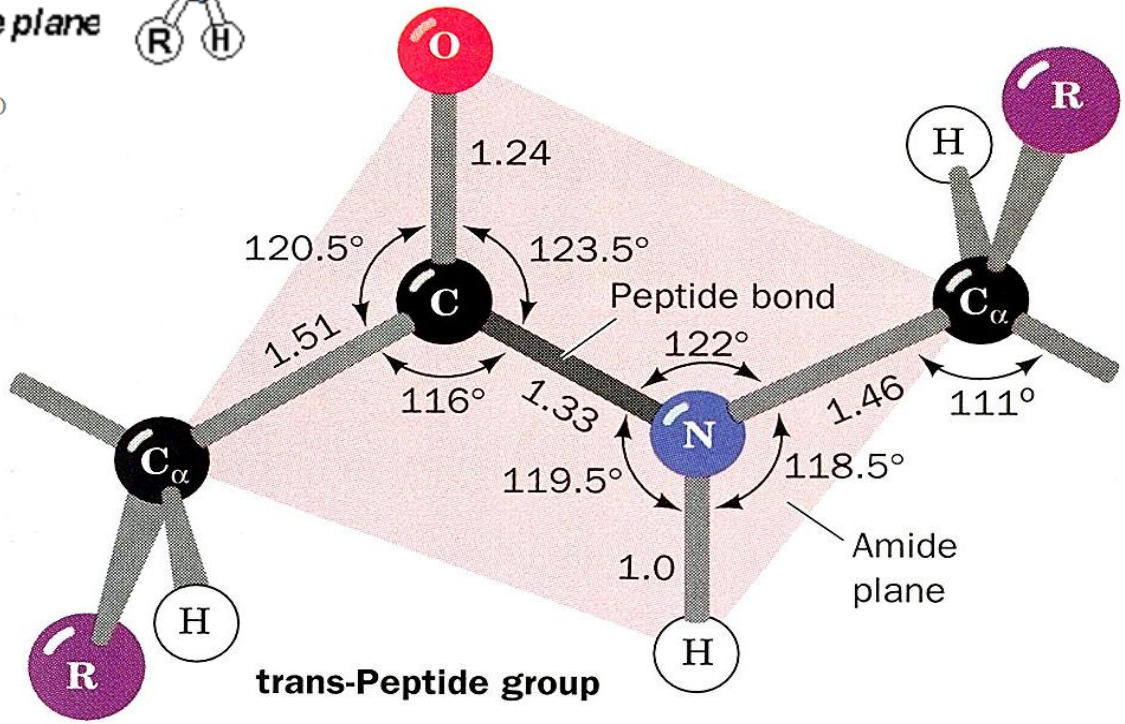
- ***Denaturation**: Partial to complete unfolding*
  Denatured Protein: Protein that has lost its native conformation



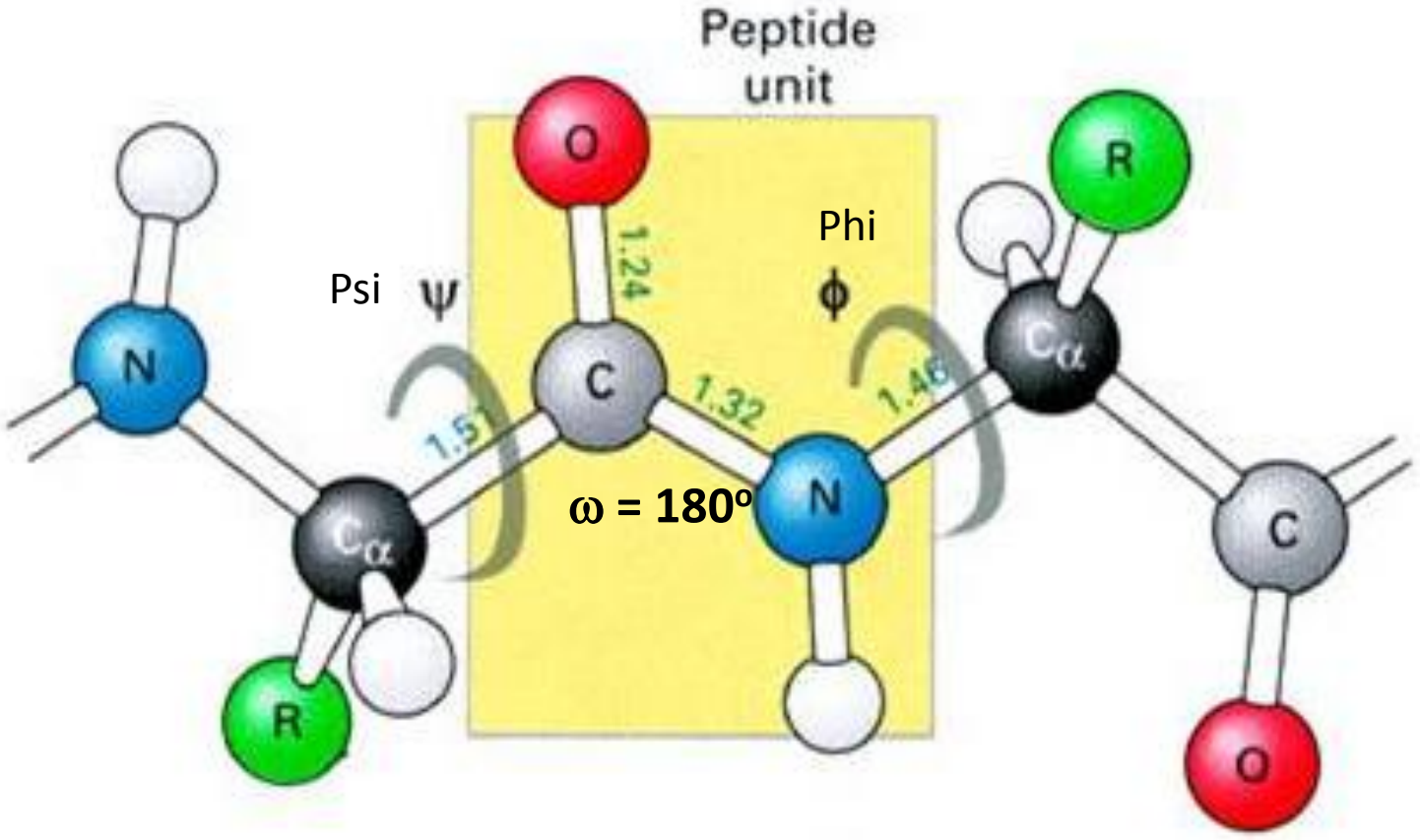
Chemistry of the peptide bond



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

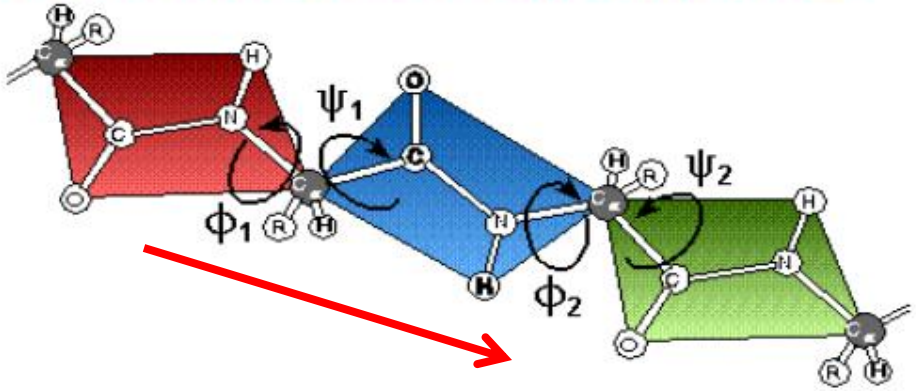


Phi-Psi angles



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

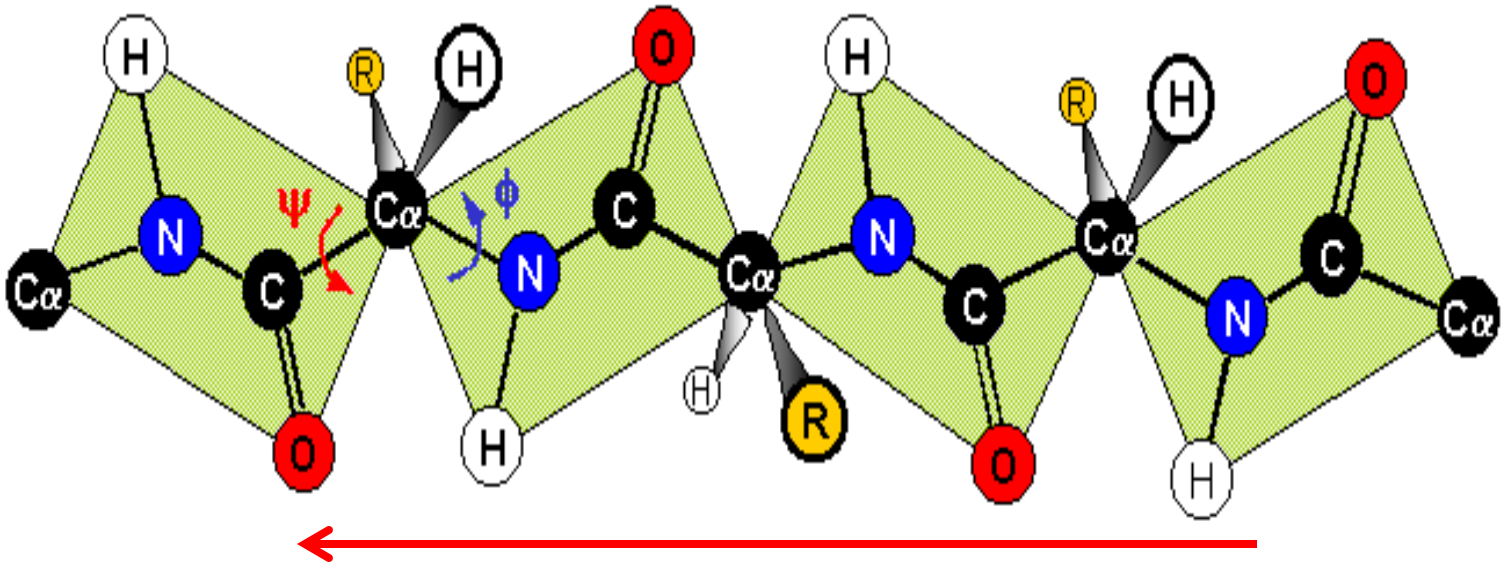
Conformation of a polypeptide



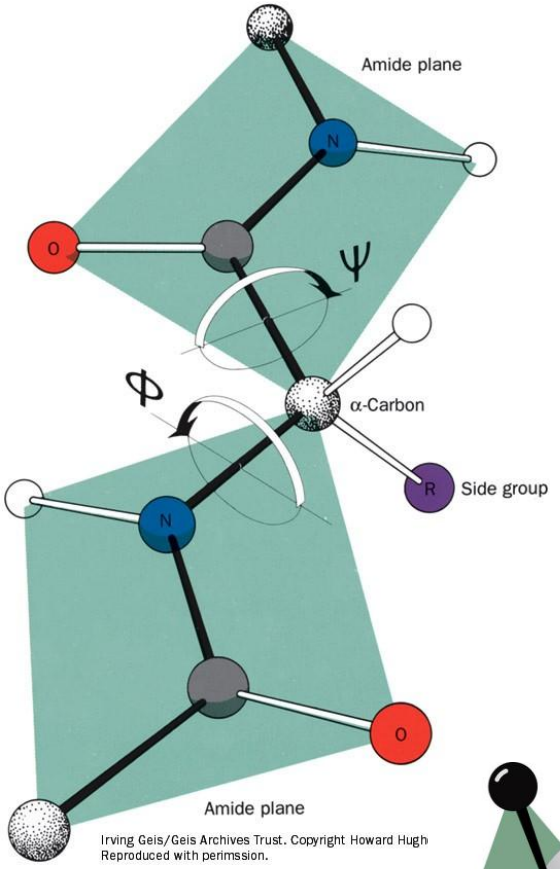
ϕ - rotation around the N-C_α bond

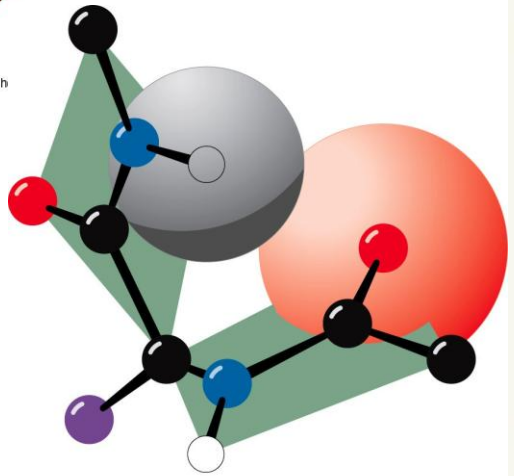
ψ - rotation around the C_α-C bond

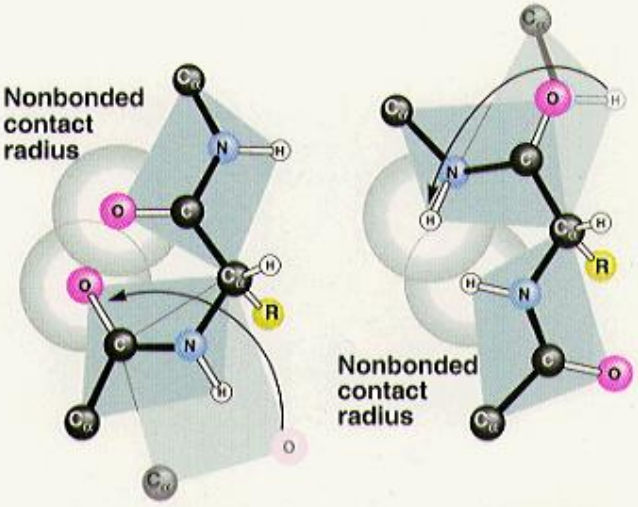
FULLY EXTENDED POLYPEPTIDE CHAIN



Torsion angles / steric restrictions

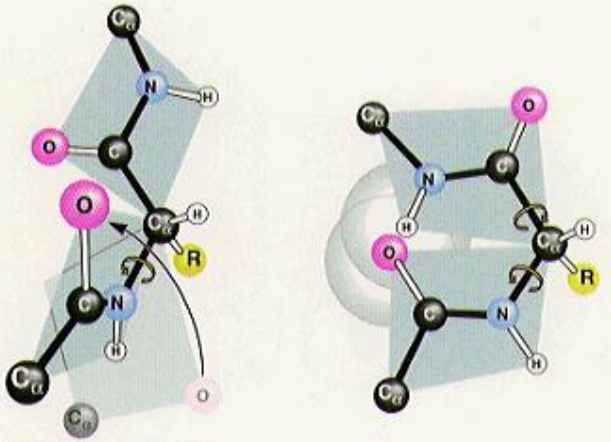






$\phi = 0°, \psi = 180°$

$\phi = 180°, \psi = 0°$

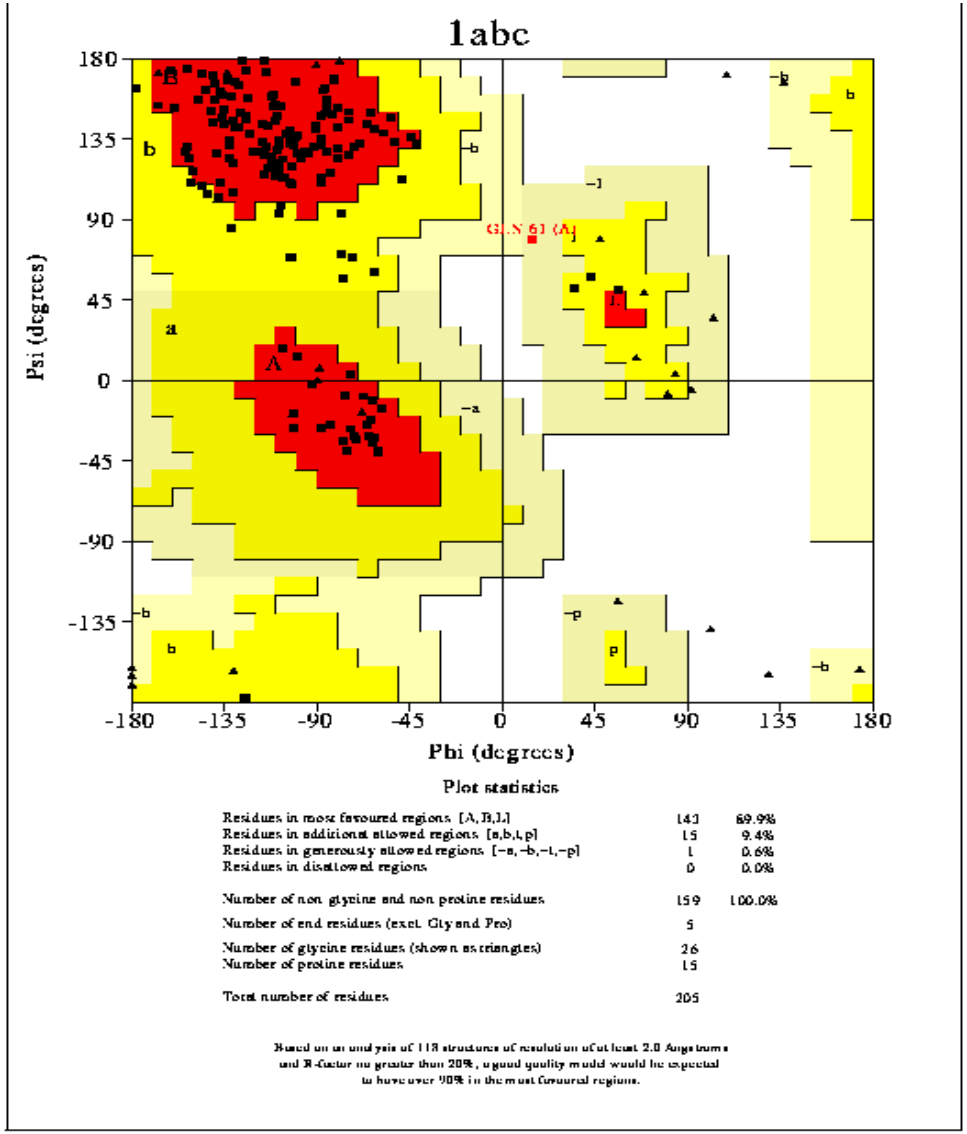


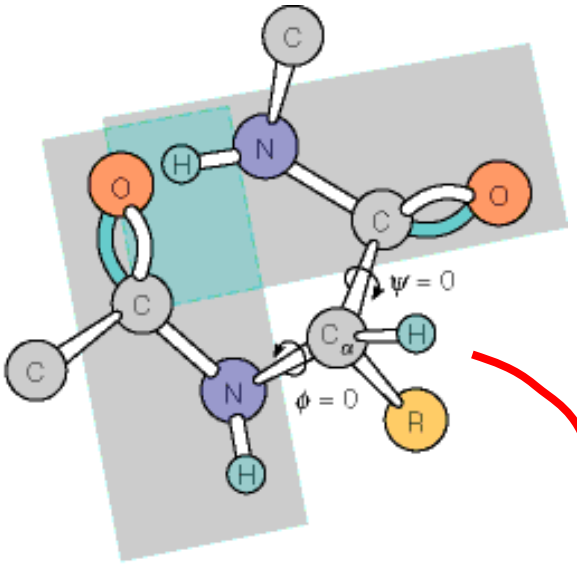
$\phi = -60°, \psi = 180°$

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

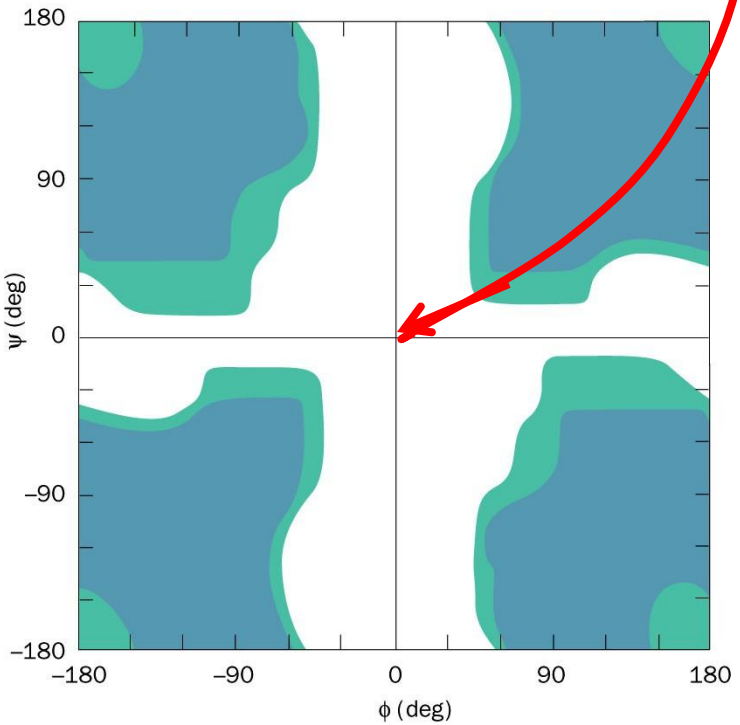
$\phi = 0°, \psi = 0°$

Ramachandran Plot

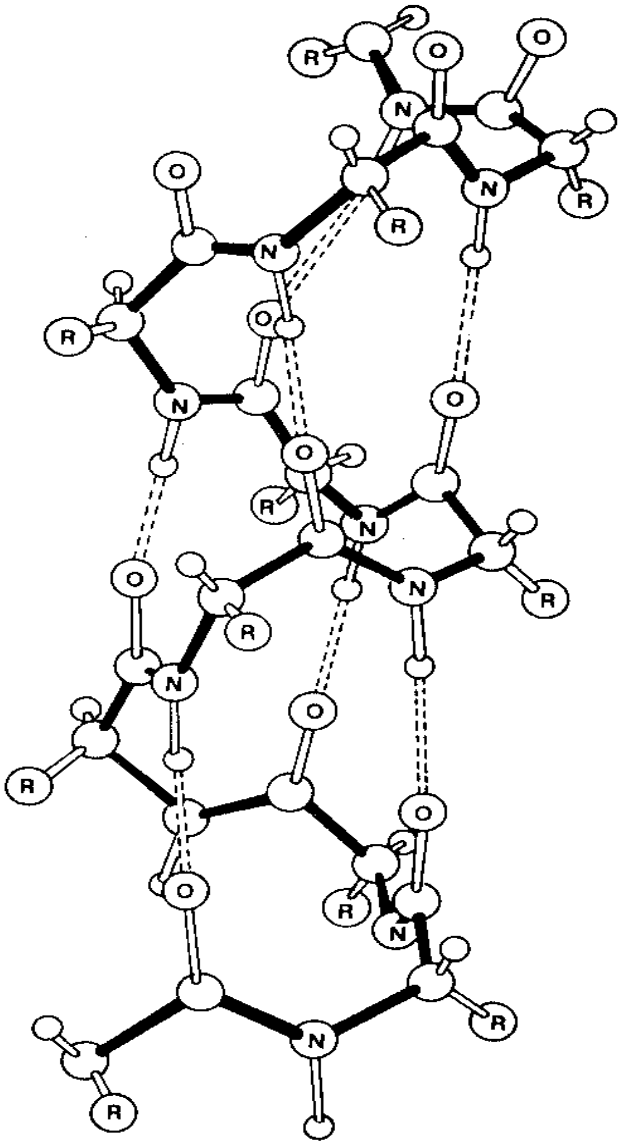




Allowed torsion angles for Gly residues:
Restrictions only from the peptide units



Protein Secondary Structure: Helices

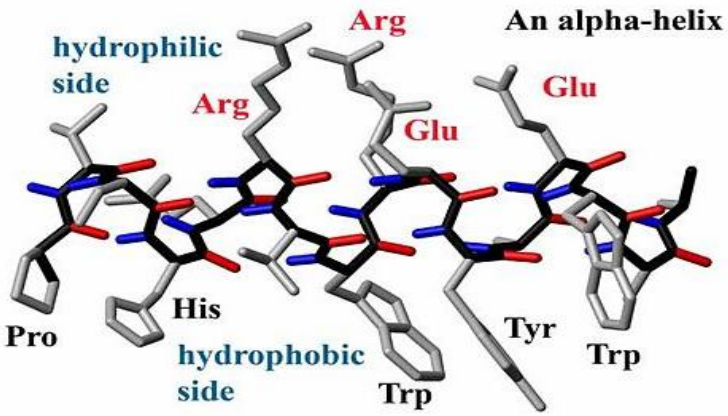


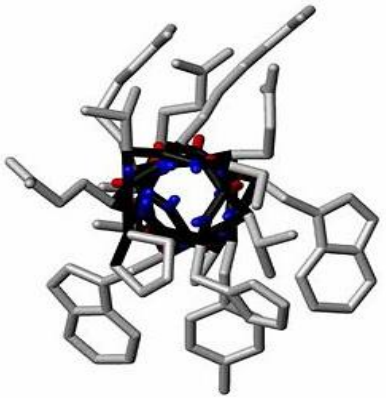
Alpha-helix:

- 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

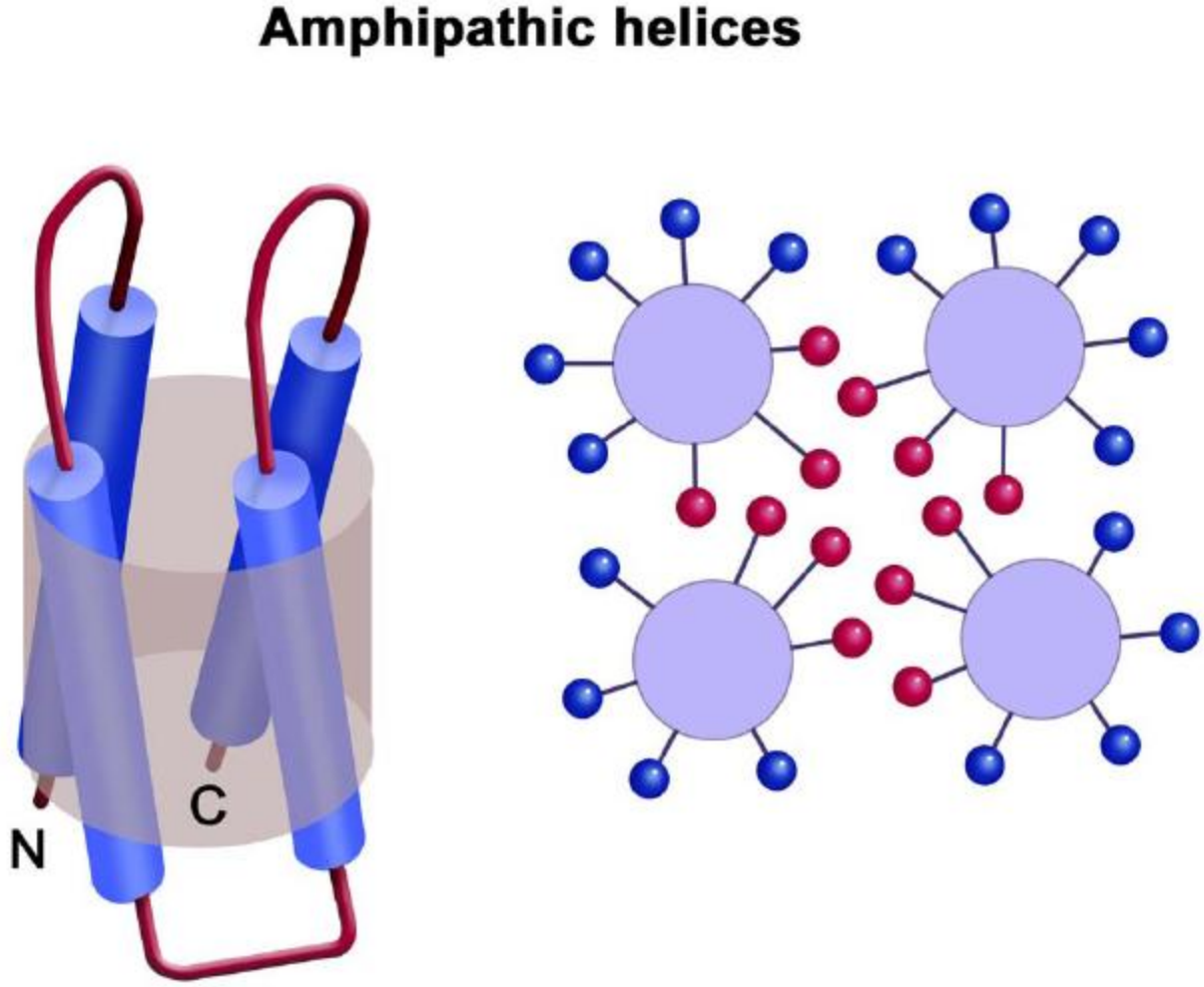
3_{10} helix

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



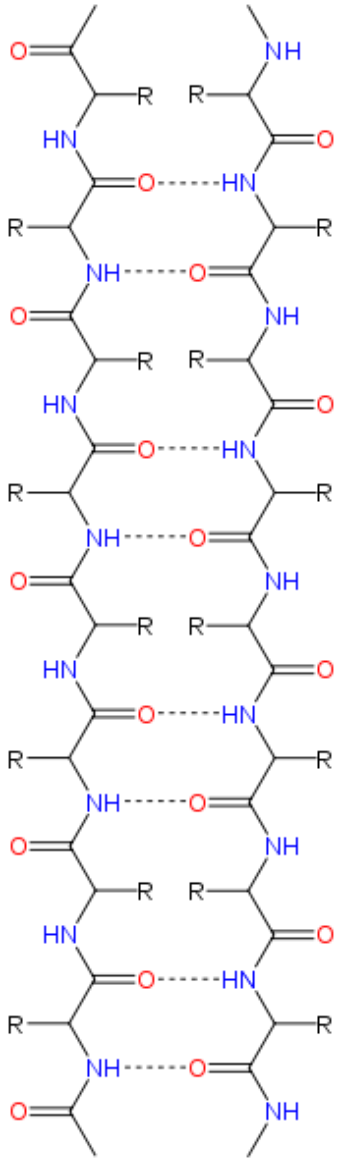


Amphipathic helices

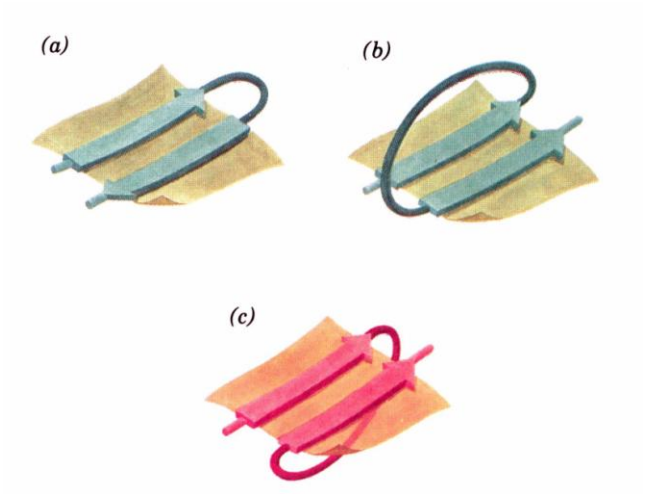


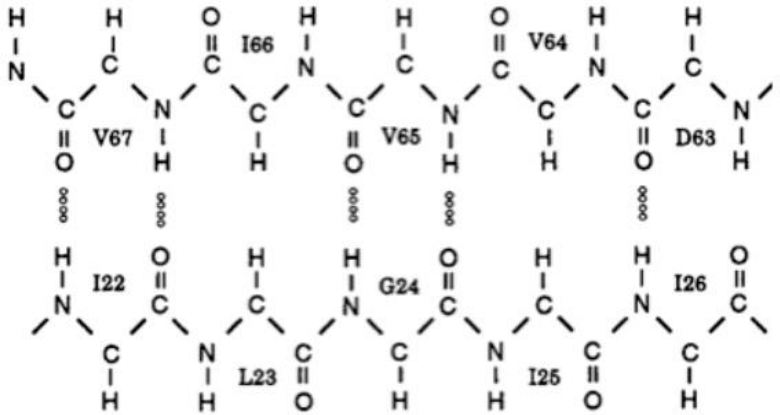
Michael Yaffe

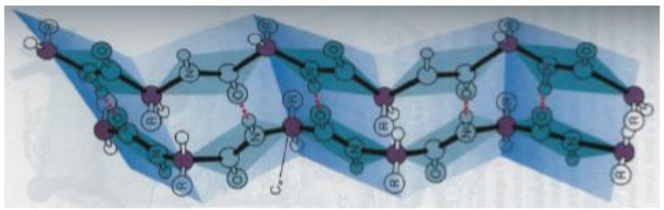
Protein Secondary Structure: Sheet

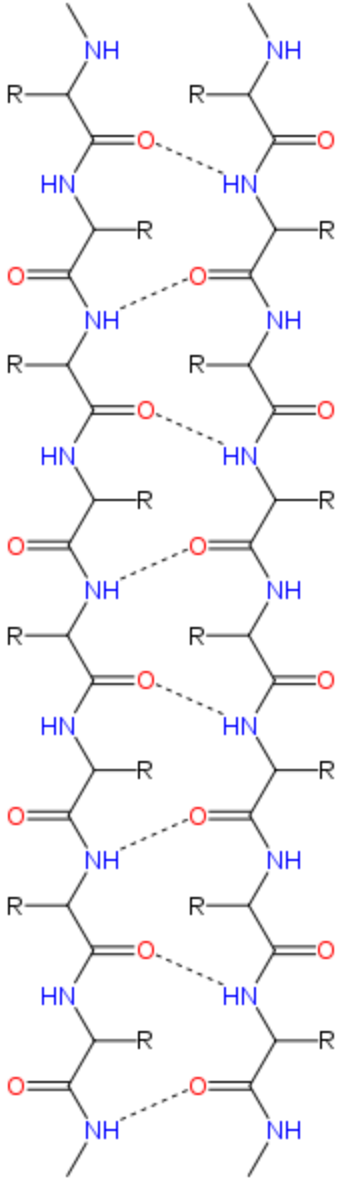


Anti-parallel





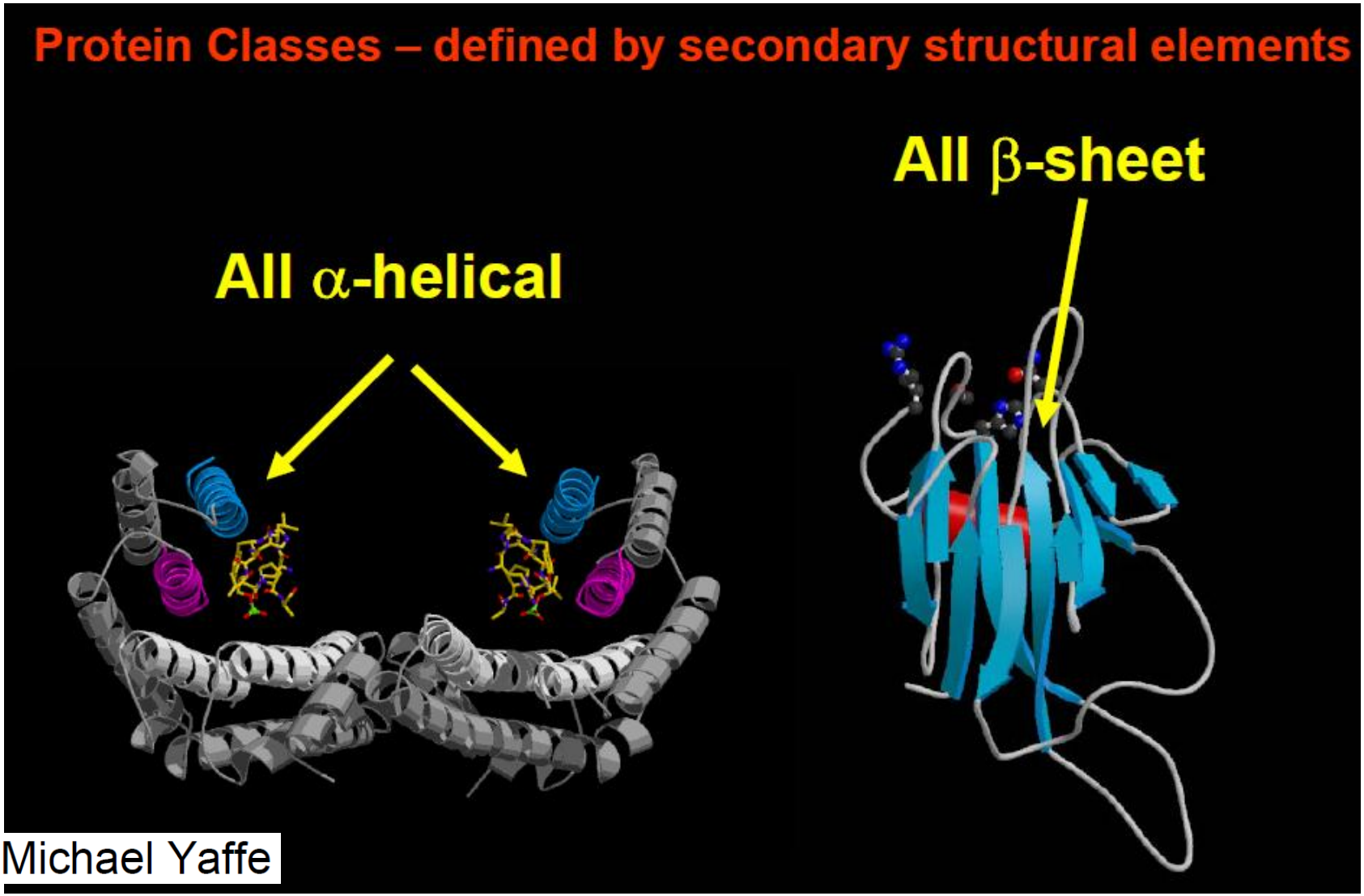




Parallel

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

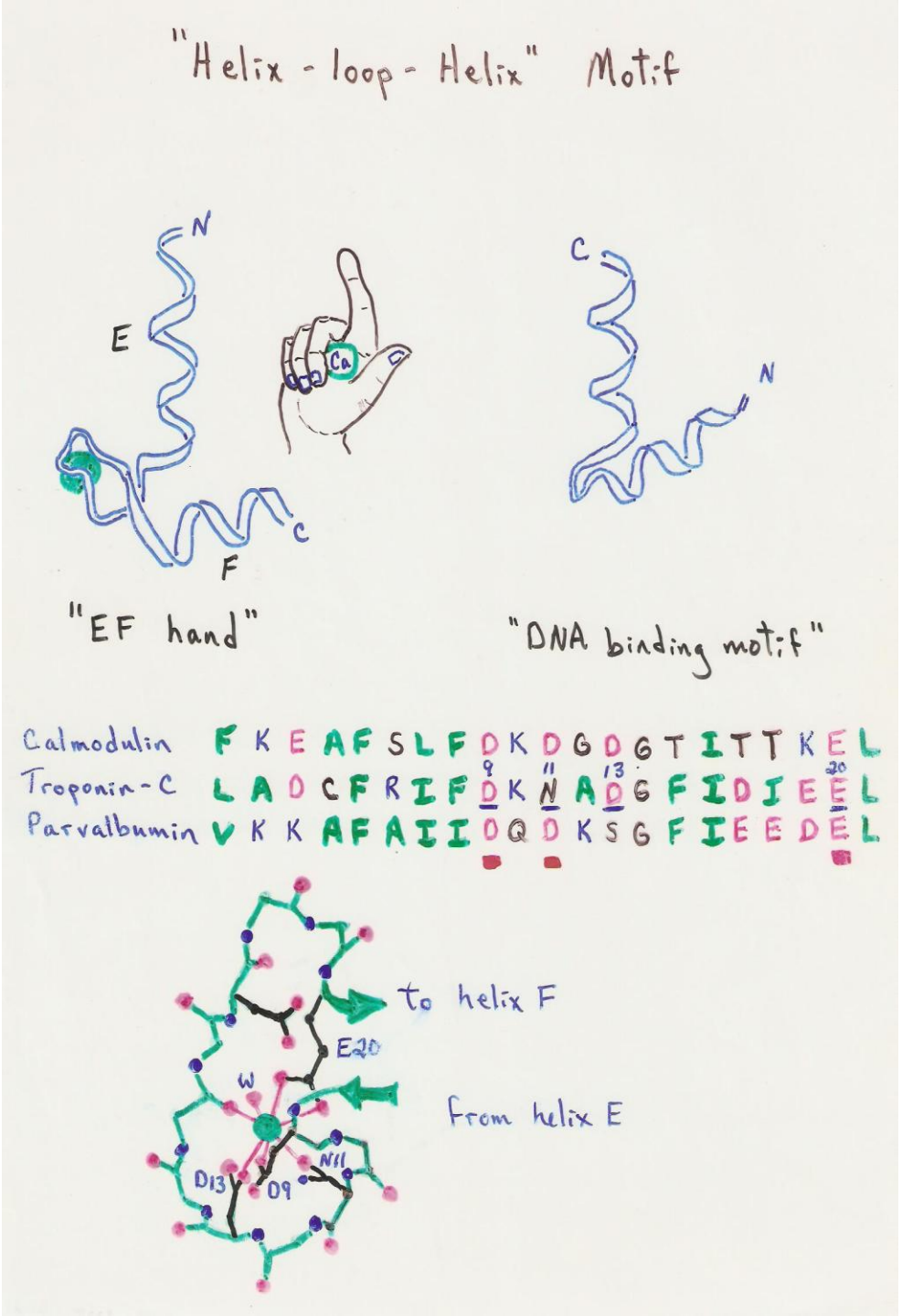
Tertiary Structure: 3D structure



Motif

1) A **motif is a sequence**, which is predictive of which sequences belong to the defined group.
For example**, sequence motifs** can characterize which proteins (protein sequences) belong to a given protein family. A simple motif could be, for example, some pattern which is strictly shared by all members of the group, e.g. WTRXEKXXY (where X stands for any amino acid).

2) A **structural motif** (also called super-secondary structure) refers to a set of contiguous secondary structure elements that either have a particular functional significance or define a portion of an independently folded domain.

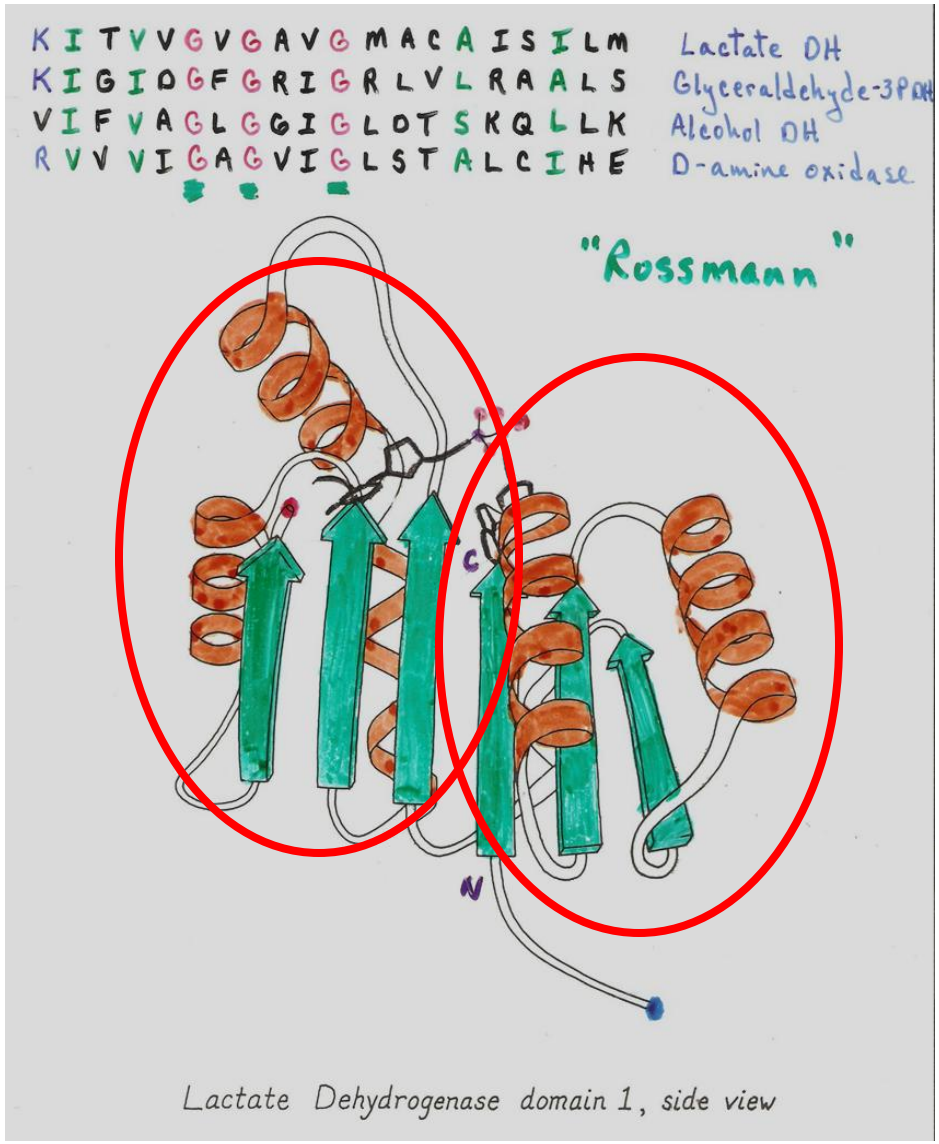


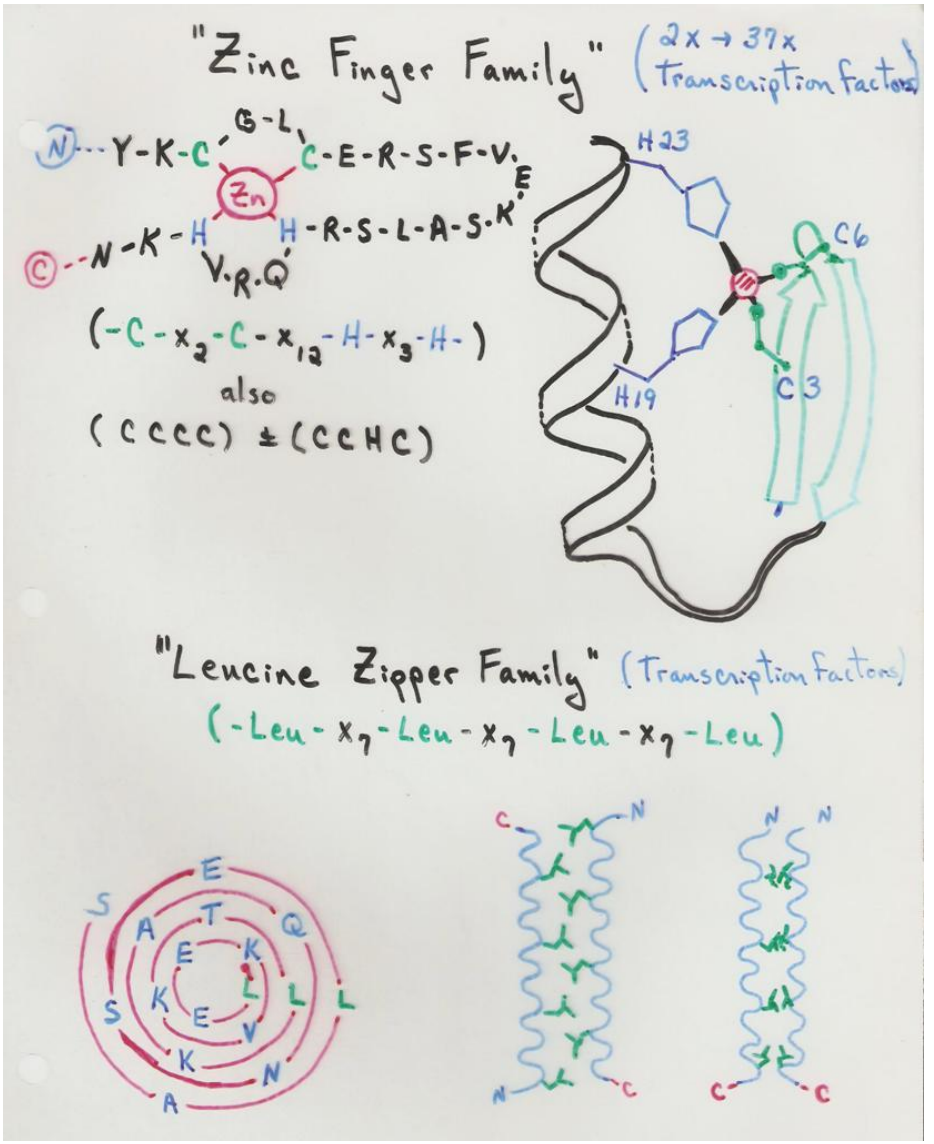
Domain:

1) A spatially separated unit of the protein structure

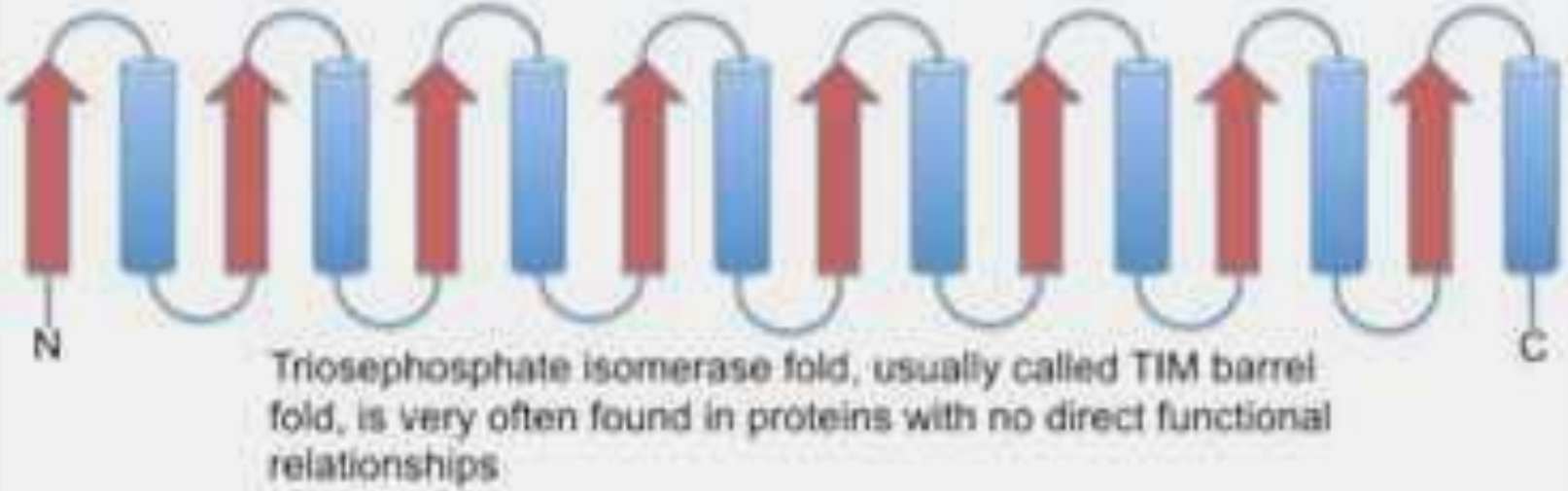
2) May have sequence and/or structural resemblance to another protein structure or domain.

3) May have a specific function associated with it.





TIM Barrel



Triosephosphate isomerase fold, usually called TIM barrel fold, is very often found in proteins with no direct functional relationships

http://www.proteinstructures.com/Structure/Structure/protein-domains.html

SCOP Structural Classification of Proteins

Structural Classification of Proteins

Root: scop

Classes:

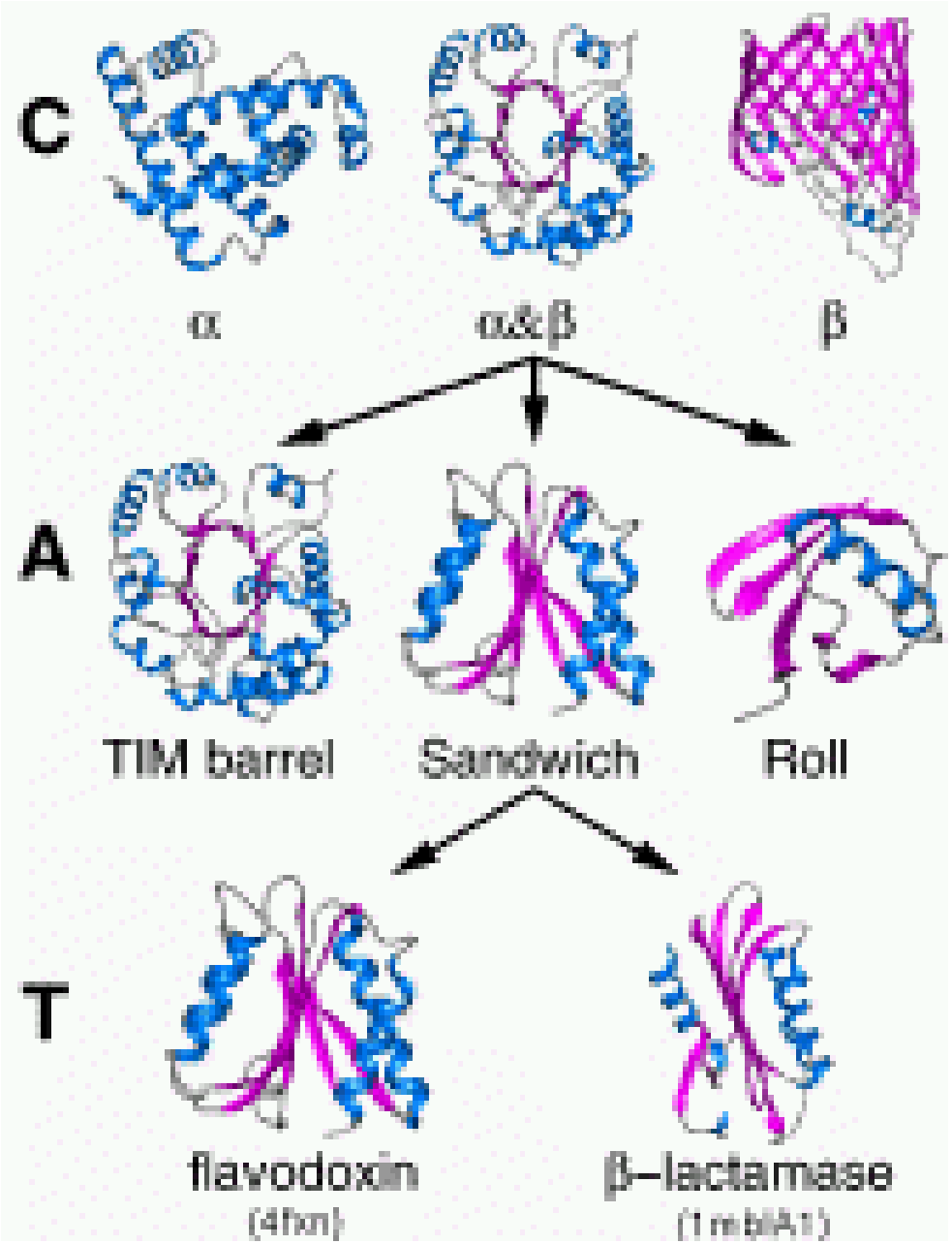
1. All alpha proteins (151)
2. All beta proteins (111)
3. Alpha and beta proteins (a/b) (117)
   Mainly parallel beta sheets (beta-alpha-beta units)
4. Alpha and beta proteins (a+b) (212)
   Mainly antiparallel beta sheets (segregated alpha and beta regions)
5. Multi-domain proteins (alpha and beta) (39)
   Folds consisting of two or more domains belonging to different classes
6. Membrane and cell surface proteins and peptides (12)
   Does not include proteins in the immune system
7. Small proteins (59)
   Usually dominated by metal ligand, heme, and/or disulfide bridges
8. Coiled coil proteins (5)
   Not a true class
9. Low resolution protein structures (17)
   Not a true class
10. Peptides (95)
    Peptides and fragments. Not a true class
11. Designed proteins (36)
    Experimental structures of proteins with essentially non-natural sequences. Not a true class

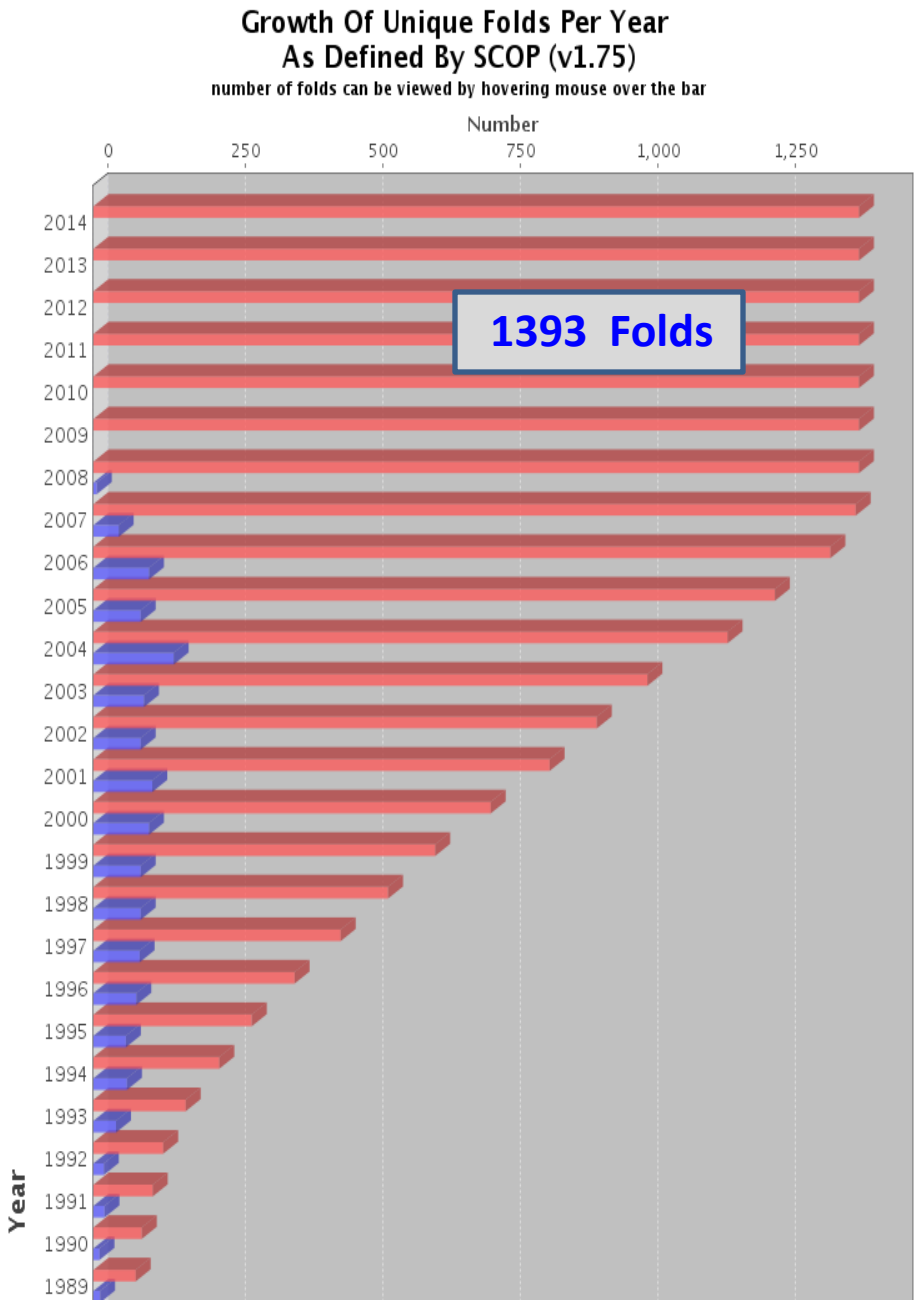
CATH - Protein Structure Classification

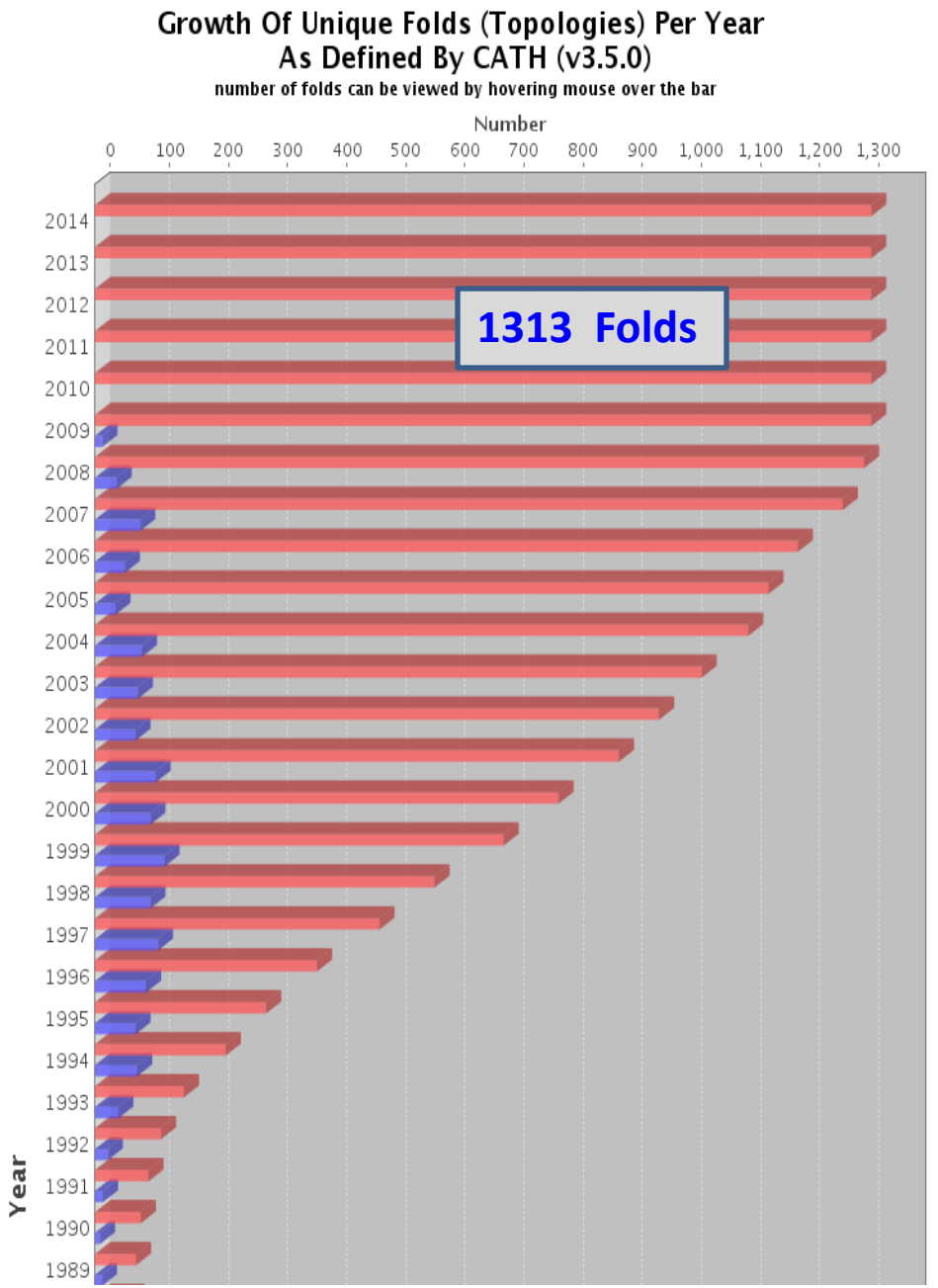
CATH is a novel hierarchical classification of protein domain structures, which clusters proteins at four major levels: Class (C), Architecture (A), Topology (T), and Homologous (H) Superfamily

Class, derived from secondary structure content, is assigned for more than 90% of protein structures automatically. **Architecture**, which describes the gross orientation of secondary structures, independent of connectivities, is currently assigned manually. The **topology** level clusters structures according to their topological connections and numbers of secondary structures. The **homologous superfamilies** cluster proteins with highly similar structures and functions. The assignments of structures to toplogy families and homologous superfamilies are made by sequence and structure comparisons.

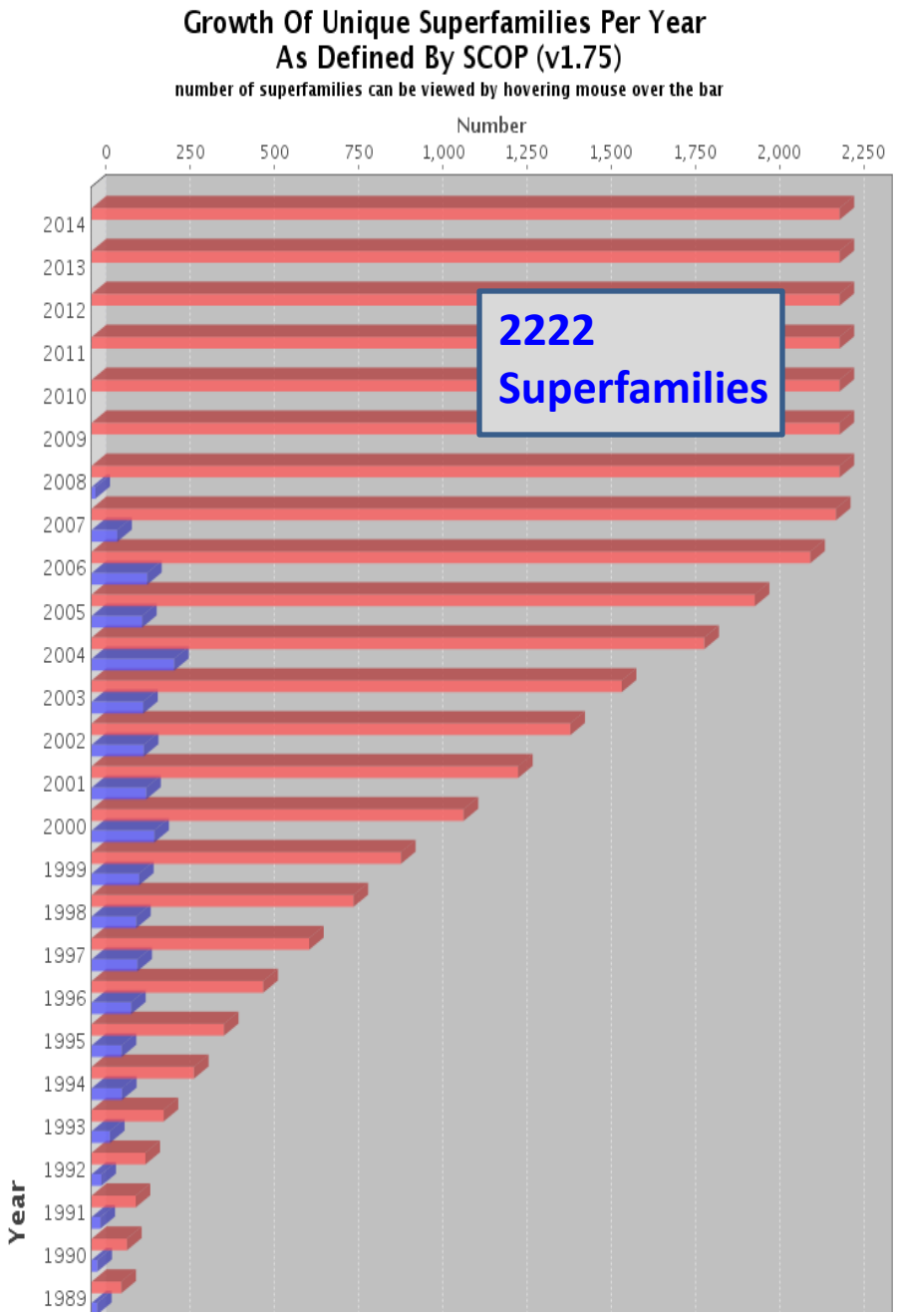
CATH

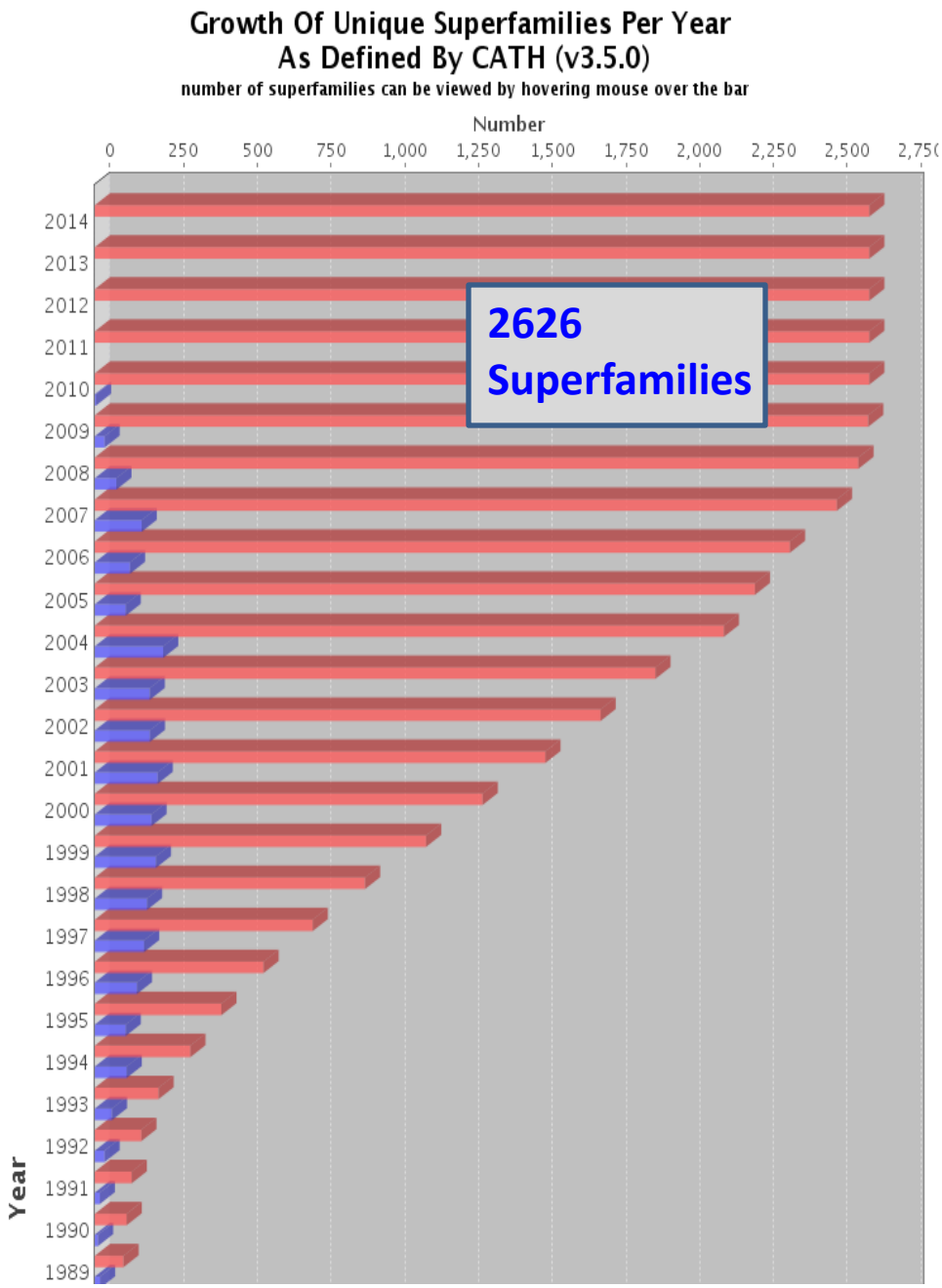

C
α
α&β
β
A
TIM barrel
Sandwich
Roll
T
flavodoxin
(4fxn)
β–lactamase
(1mbiA1)





http://pdb.org/pdb/static.do?p=general_information/pdb_statistics/index.html



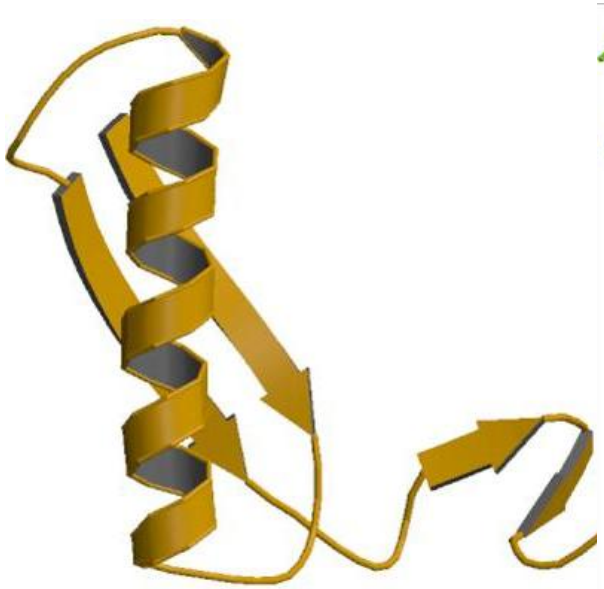


http://pdb.org/pdb/static.do?p=general_information/pdb_statistics/index.html

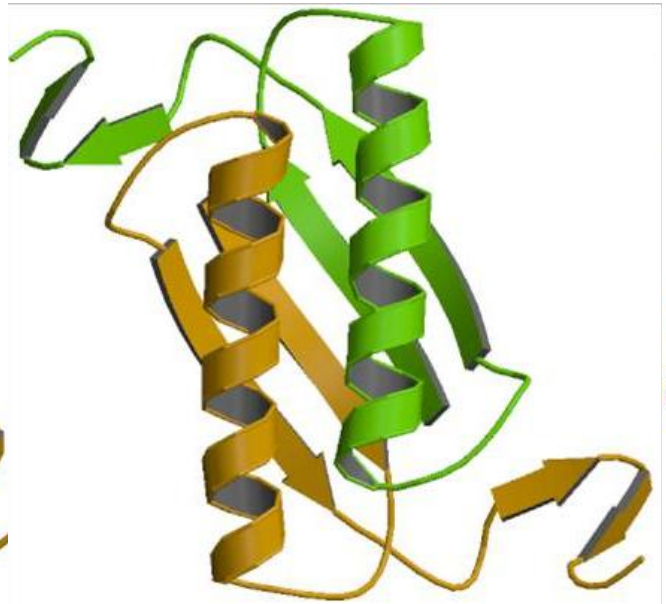
Quaternary Structure:

Arrangements of subunits in oligomers

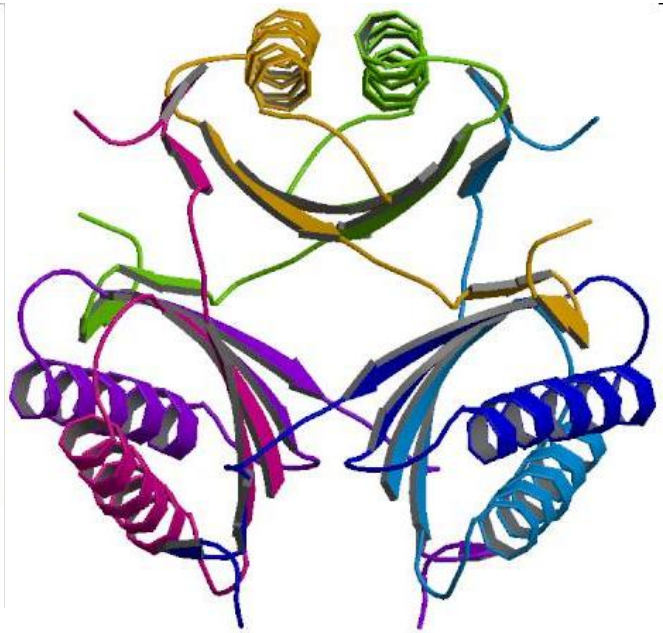
α_4 ; α_{12} ; $(\alpha\beta)_2$; $(\alpha\beta)_6$



monomer



dimer



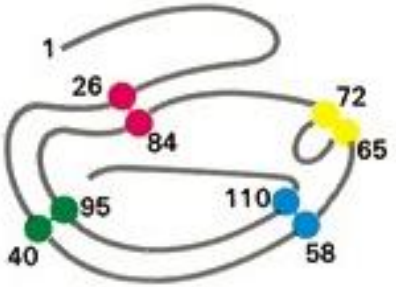
Hexamer

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

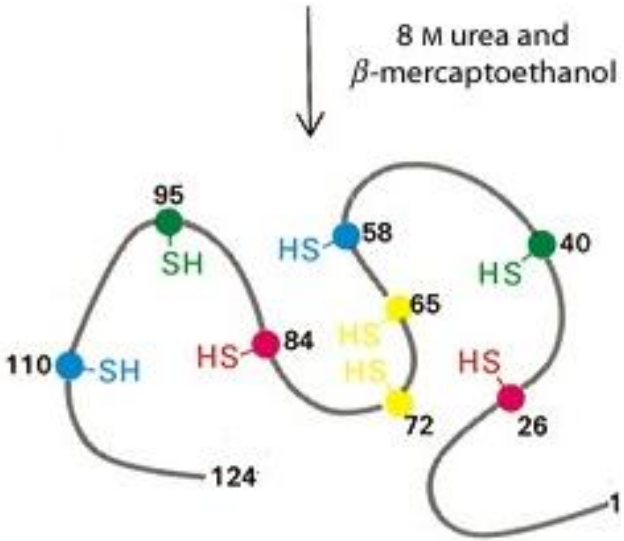
Renaturation :

Chris Anfinsen - Folding of Ribonuclease

124 a.a. + 4 disulfides

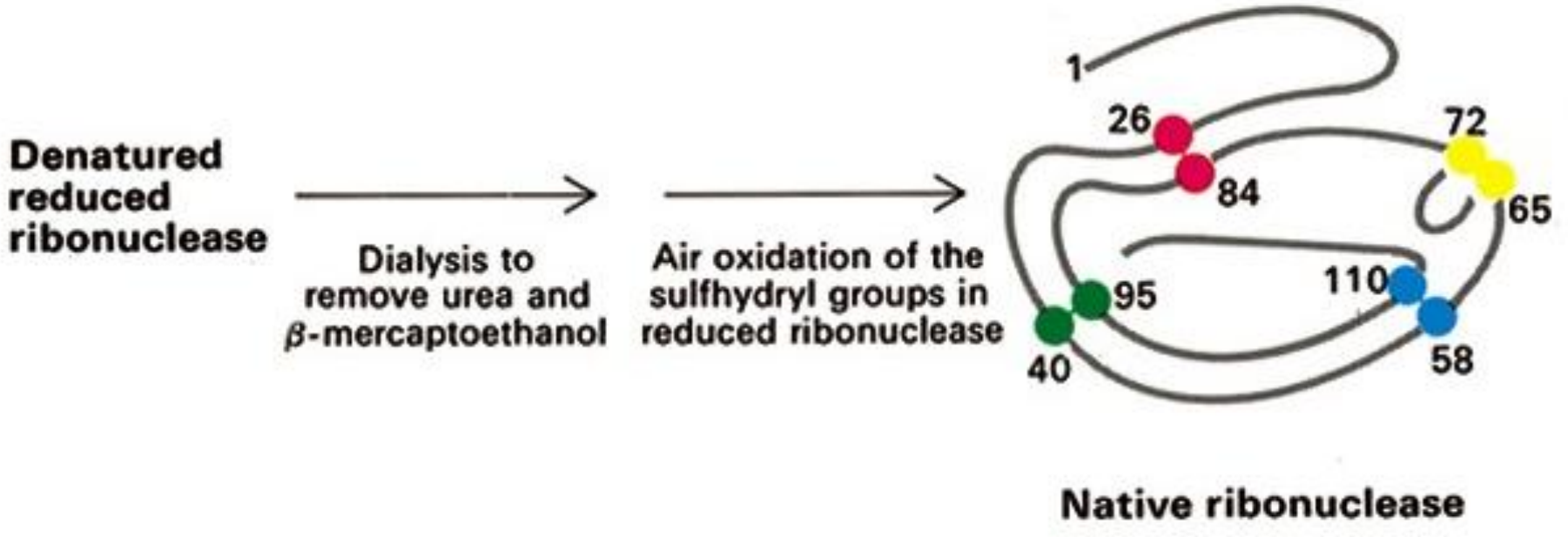
(26 → 84; 40 → 95; 58 → 110; 65 → 72)

(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.

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

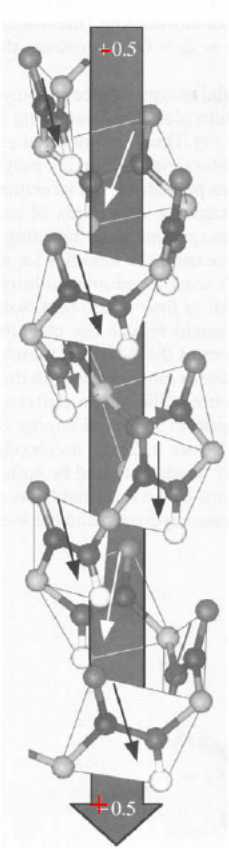
Dipole-related interaction

van de Waals interaction

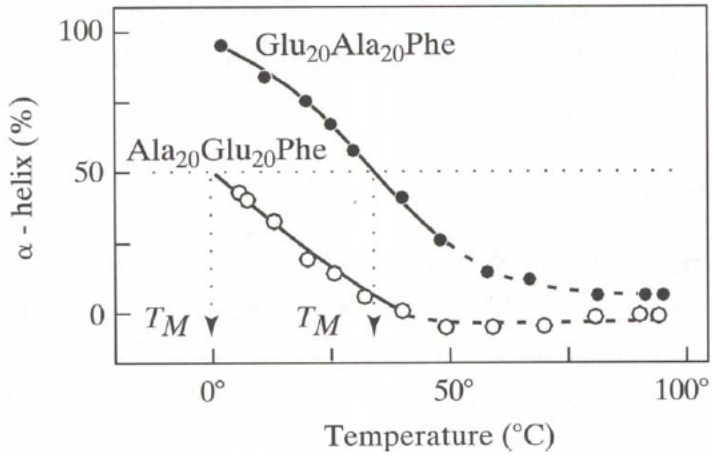
Hydrophobic interaction

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: Stability / Denaturation and Renaturation

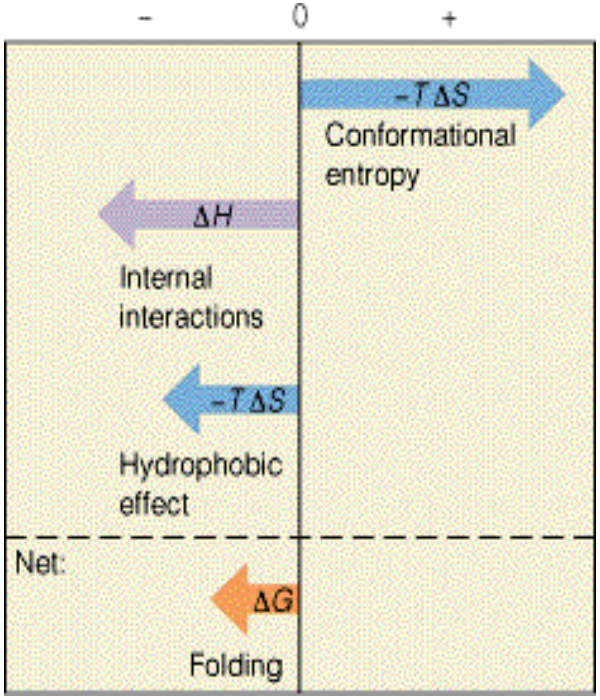
Protein folding (tertiary structure) is determined by weak interactions

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

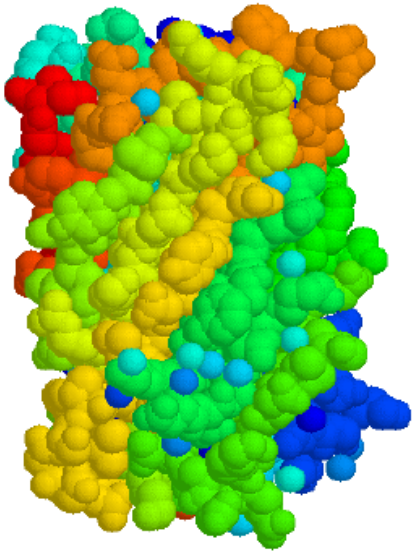


Protein	ΔG (kJ/mol)	ΔH (kJ/mol)	ΔS (J/K·mol)
Ribonuclease	−46	−280	−790
Chymotrypsin	−55	−270	−720
Lysozyme	−62	−220	−530
Cytochrome *c*	−44	−52	−27
Myoglobin	−50	0	+170

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.

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)

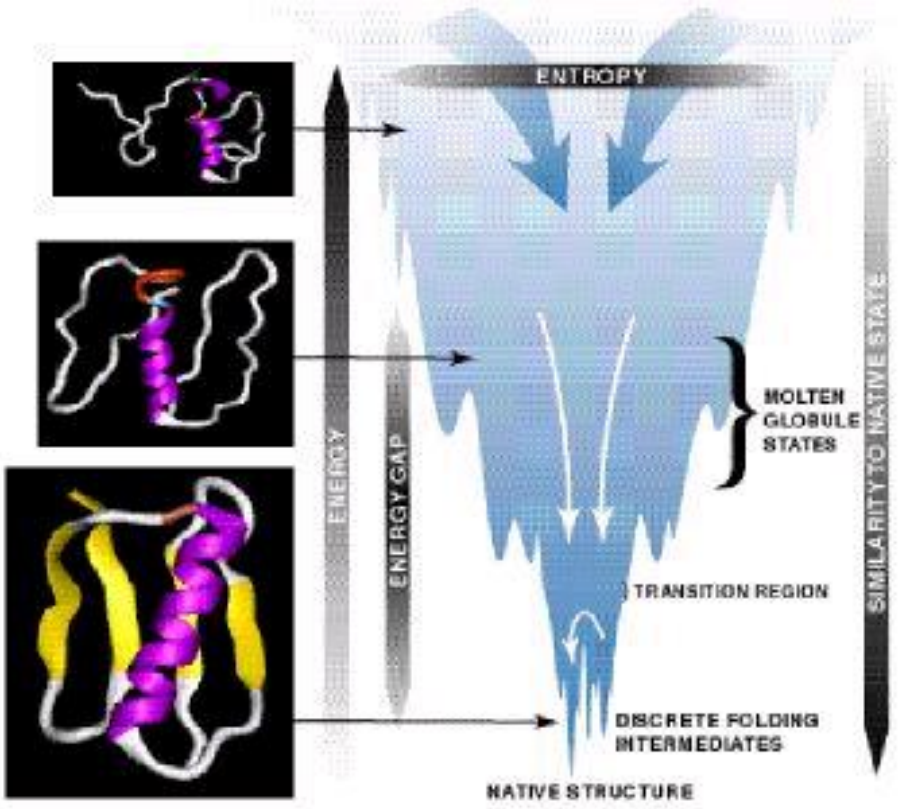
- There are too many possible conformations for a protein to fold by a random search.
- Consider just for the peptide backbone, there are at least 3 conformations per amino acid in the unfolded state, For a 100 a.a. protein we have 3^{100} conformations.
- If the chain can sample 10^{12} conformations/sec, 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)

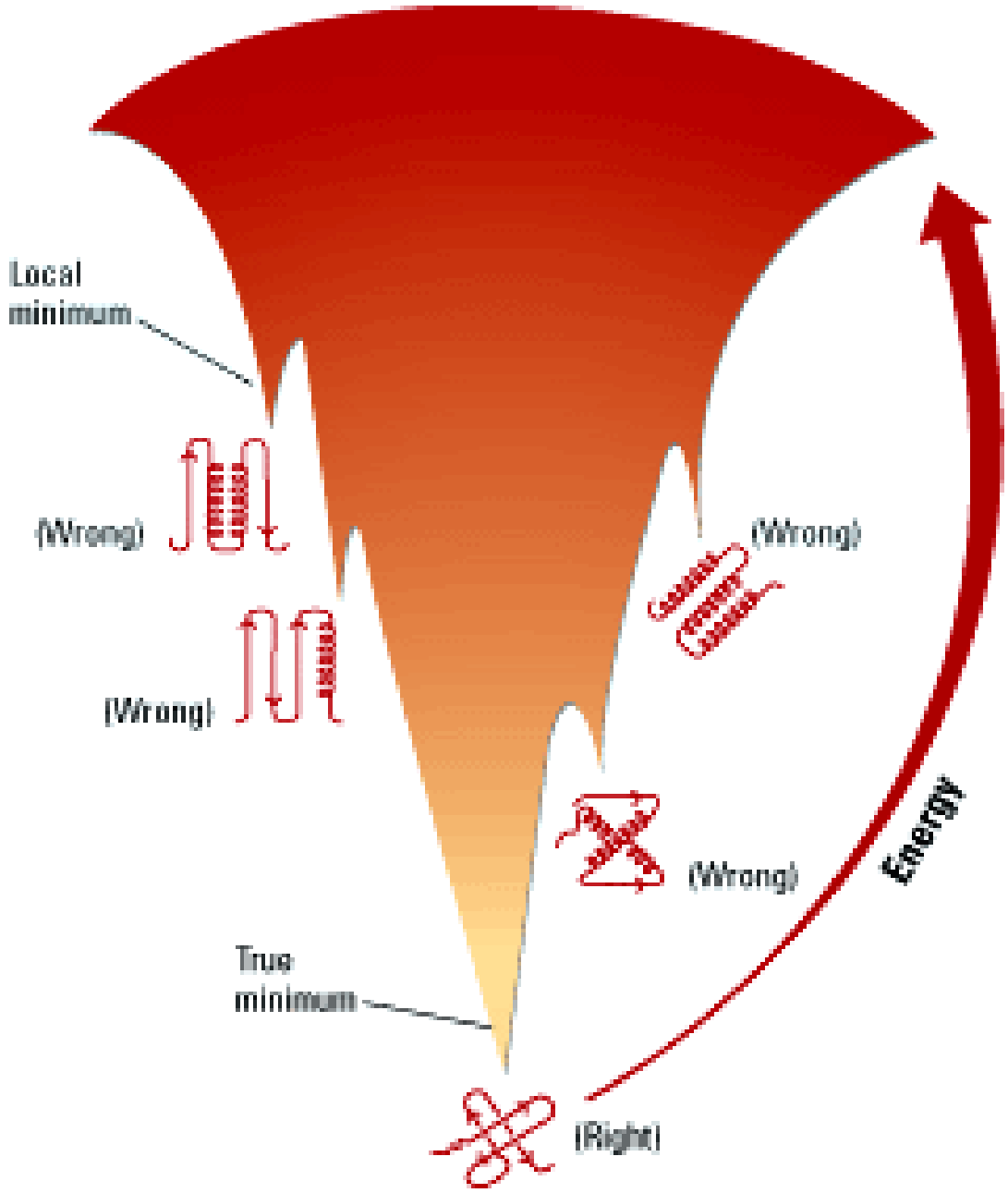
Proteins fold on timescales ranging from a microsecond to a few minutes, so they obviously drive or are driven quickly toward the native state.

- 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.

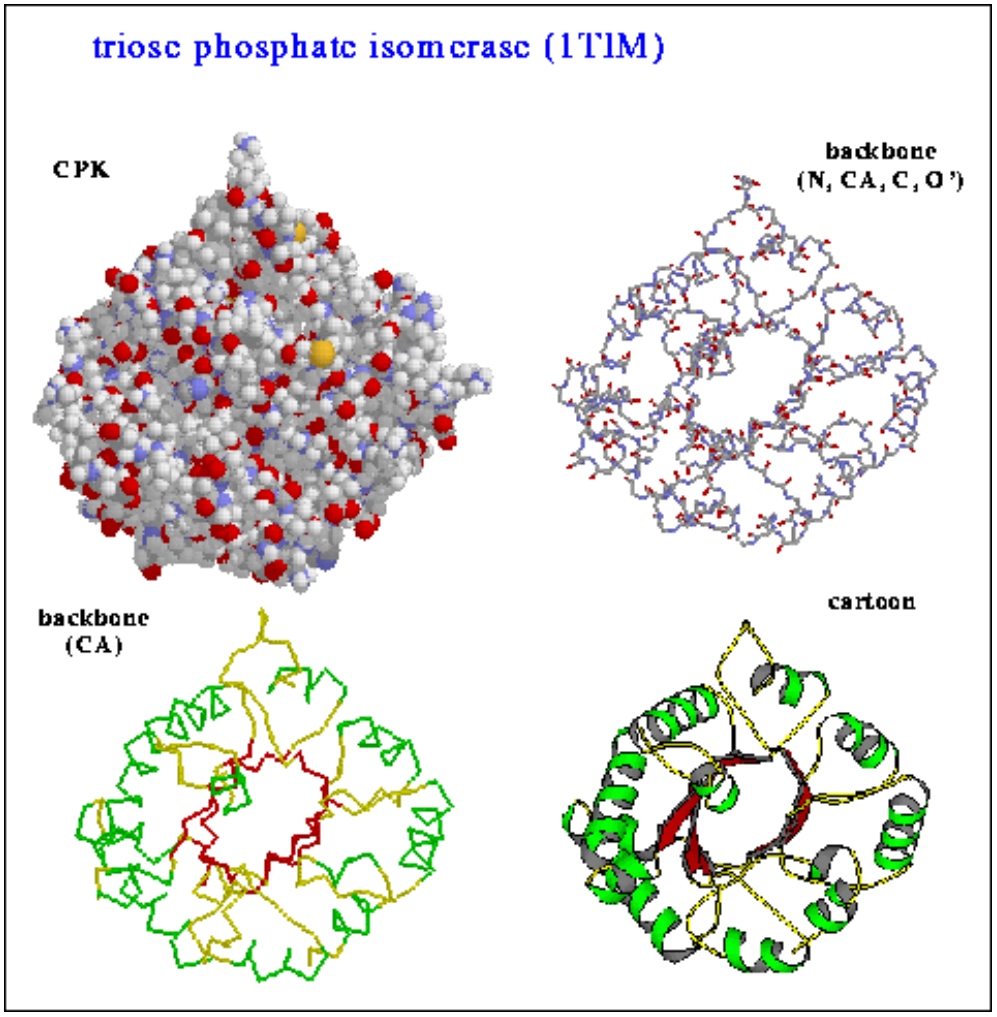
The transition state is composed of a broad ensemble of structures rather than one particular structure.



Protein Structure

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

=



Chou-Fasman

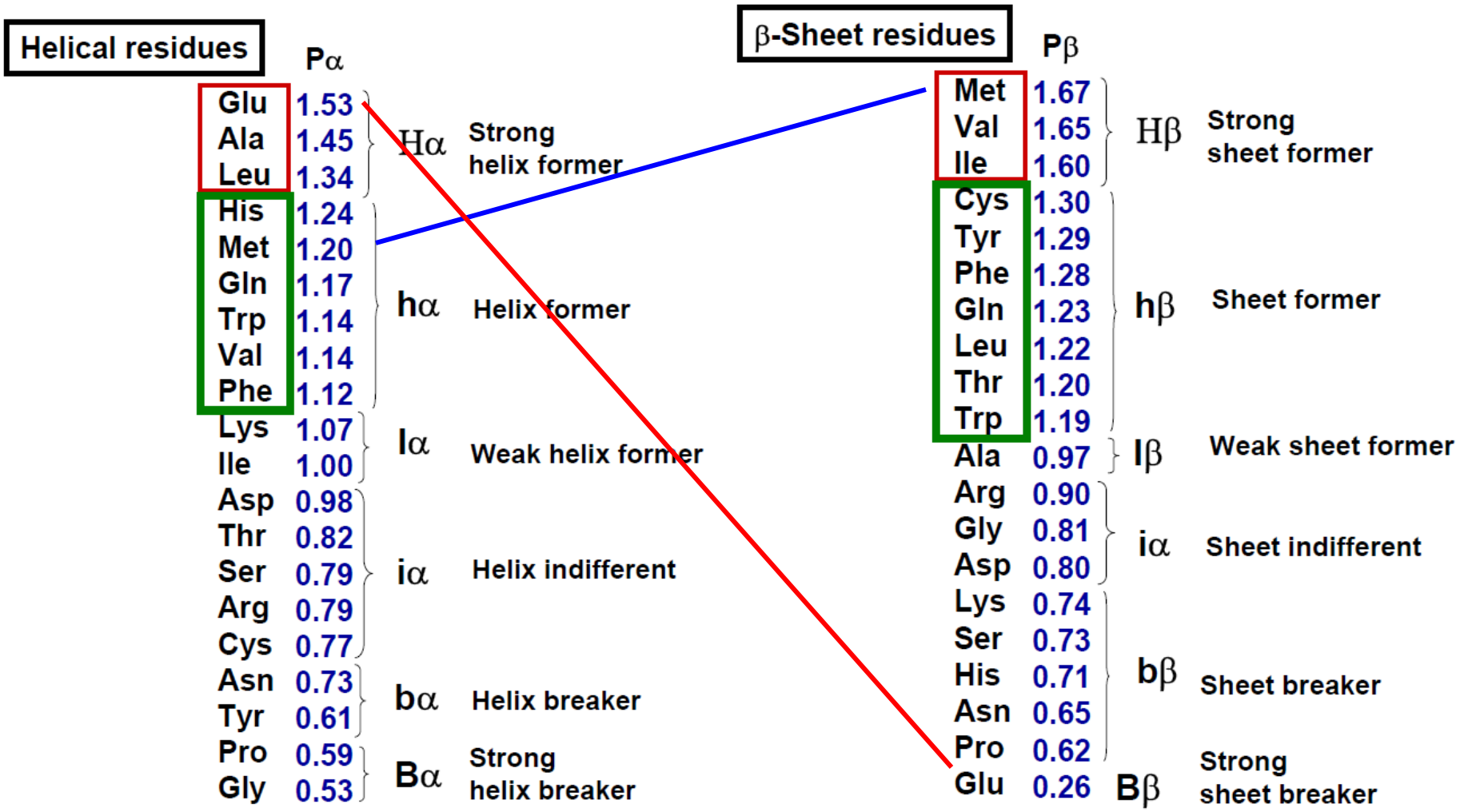
Biochemistry, *13*: 222-245, 1974

•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

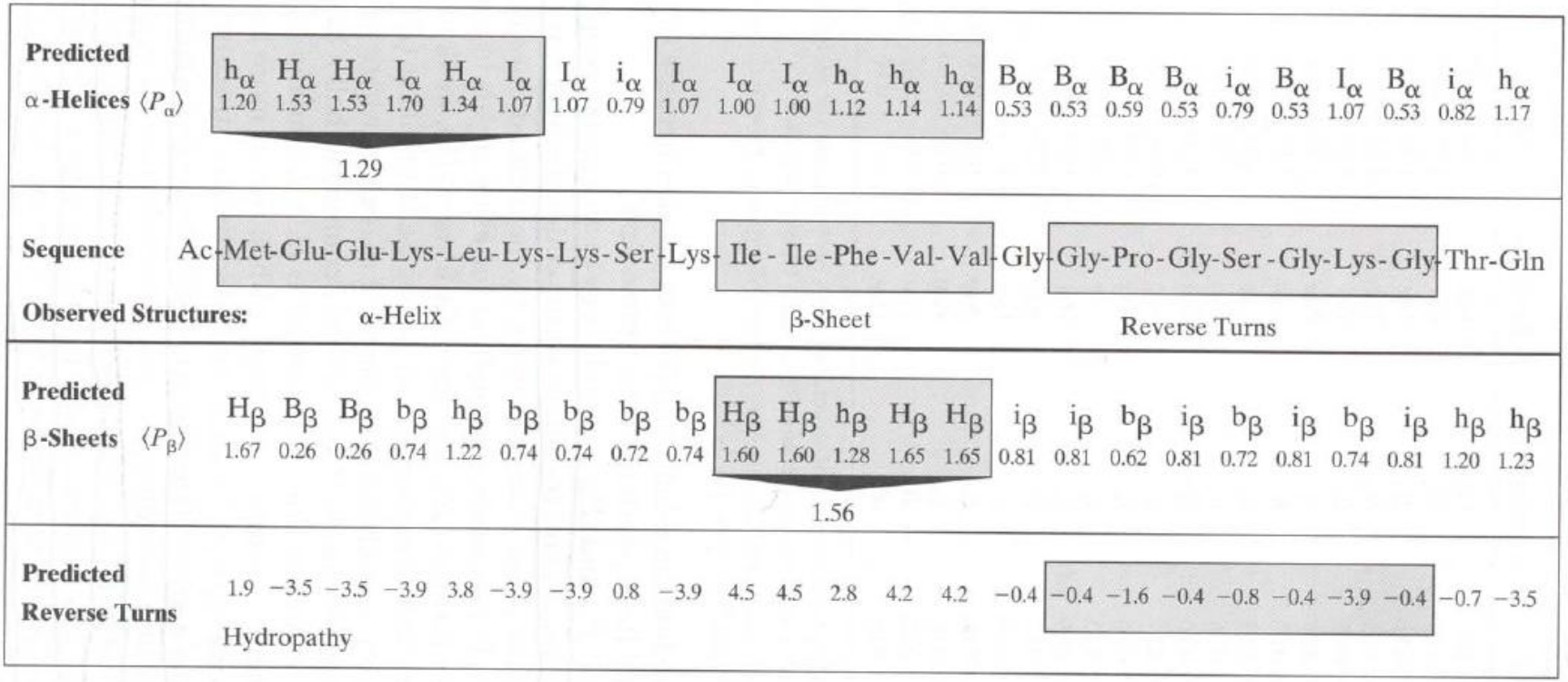
Chou-Fasman

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)

Protein Structure Prediction Center

11th Community Wide Experiment on the Critical Assessment of Techniques for Protein Structure Prediction

Menu

- Home
- FORCASP Forum
- PC Login
- PC Registration
- CASP Experiments
  - CASP ROLL
  - CASP11 (2014)
    - Home
    - My CASP11 profile
    - Targets
    - CASP11 in numbers
  - CASP10 (2012)
  - CASP9 (2010)
  - CASP8 (2008)
  - CASP7 (2006)
  - CASP6 (2004)
  - CASP5 (2002)
  - CASP4 (2000)
  - CASP3 (1998)
  - CASP2 (1996)
  - CASP1 (1994)

CASP11

CASP11 model collection is completed as of August 19, 2014. We are currently working on the evaluations in preparation for the December meeting in Mexico.

CASP11 provides an independent mechanism for the assessment of current methods in protein structure modeling. From April through July 2014, structures about to be solved by crystallography or NMR are identified, and their sequences are made available to predictors. Through the Summer and Fall, as the experimental coordinates become available, the tens of thousands of models submitted by approximately 200 prediction groups worldwide are processed and evaluated. Independent assessors bring objectivity, balance, and independent insight to this process. Tools for viewing, comparison, and analysis of submitted models are made available at this site. The results of the CASP11 Experiment will first be made public and discussed at the CASP11 Meeting to be held in December 2014.

Targets	Predictions	Meeting
Target List	Model Viewer	**Register for the meeting**
	Server Tarballs	*Early bird registration deadline - September 4, 2014*
		Abstract submission
		Abstract submission deadline - September 19, 2014

CASP11 in numbers

Detailed description of the experiment

Goals | Scope | Related | Timetable | Participation | Targets | Format | Assessment | Results | Meeting | Organizers

Rosetta@home

Protein Folding, Design, and Docking

What is Rosetta@home?

Rosetta@home Like 6,110

Rosetta@home needs your help to determine the 3-dimensional shapes of proteins in research that may ultimately lead to finding cures for some majo human diseases. By running the Rosetta program on your computer while you don't need it you will help us speed up and extend our research in way we couldn't possibly attempt without your help. You will also be helping our efforts at designing new proteins to fight diseases such as HIV, Malaria Cancer, and Alzheimer's (See our Disease Related Research for more information). Please join us in our efforts! ***Rosetta@home is not for profit.***

Site search

Follow us on Twitter: @rosettaathome Tweet

Join Rosetta@home

1. Rules and policies
2. System requirements
3. Download, install, and run BOINC
   (enter the project URL: http://boinc.bakerlab.org/rosetta)
4. A welcome from David Baker
5. Donate

About

- 10 reasons why users crunch Rosetta@home
- Quick Guide to Rosetta@home and Its Graphics
- Play the interactive rosetta game, FoldIt!
- Rosetta@home FAQ
- Rosetta@home Science FAQ
- Disease Related Research
- Research Overview
- Publications
- News & Articles about Rosetta
- David Baker's Rosetta@home Journal
- Rosetta@home promo video
- Technical news

User of the day

msnelling
just thought it was a worth while cause

Server Status as of 3 Sep 2014 1:52:41 UTC

[Scheduler running]

Total queued jobs: **4,454,043**

In progress: 791,375

Successes last 24h: 239,966

Users (last day) : 637,602 (+3786)

Hosts (last day) : 1,493,925 (+3871)

Credits last 24h : 21,842,107

Total credits : 29,794,329,221

TeraFLOPS estimate: 218.421

Sep 02, 2014

Predictor of the day: Congratulations to NC for predicting the lowest energy structure for workunit *relax_1prq.4_bbintra_chi_fit_r2_199865_0* !

...more

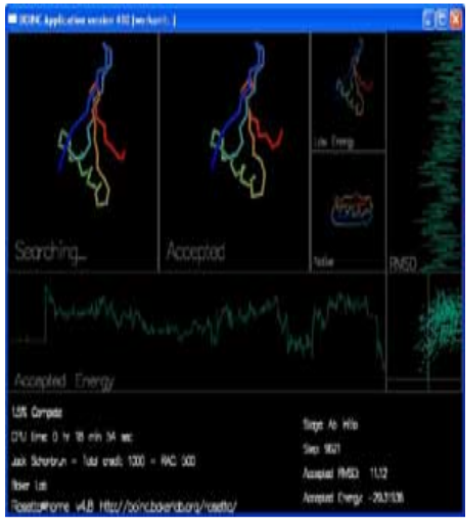
XML Available as an RSS feed..

Tweets

Follow

Quick guide to Rosetta and its graphics

About Rosetta



One of the major goals of Rosetta is to predict the shapes that proteins fold up into in nature. Proteins are linear polymer molecules made up of amino acid monomers and are often refered to as "chains." Amino acids can be considered as the "links" in a protein "chain". Here is a simple analogy. When considering a metal chain, it can have many different shapes depending on the forces exerted upon it. For example, if you pull its ends, the chain will extend to a straight line and if you drop it on the floor, it will take on a unique shape. Unlike metal chains that are made of identical links, proteins are made of 20 different amino acids that each have their own unique properties (different shapes, and attractive and repulsive forces, for example), and in combination, the amino acids exert forces on the chain to make it take on a specific shape, which we call a "fold." The order in which the amino acids are linked determines the protein's fold. There are many kinds of proteins that vary in the number and order of their amino acids.

To predict the shape that a particular protein adopts in nature, what we are really trying to do is find the fold with the lowest energy. The energy is determined by a number of factors. For example, some amino acids are attracted to each other so when they are close in space, their interaction provides a favorable contribution to the energy. Rosetta's strategy for finding low energy shapes looks like this:

1. Start with a fully unfolded chain (like a metal chain with its ends pulled).
2. Move a part of the chain to create a new shape.
3. Calculate the energy of the new shape.
4. Accept or reject the move depending on the change in energy.
5. Repeat 2 through 4 until every part of the chain has been moved a lot of times.

We call this a trajectory. The end result of a trajectory is a predicted structure. Rosetta keeps track of the lowest energy shape found in each trajectory. Each trajectory is unique, because the attempted moves are determined by a random number. They do not always find the same low energy shape because there are so many possibilities.

A trajectory may consist of two stages. The first stage uses a simplified representation of amino acids which allows us to try many different possible shapes rapidly. This stage is regarded as a low resolution search and **on the screen saver you will see the protein chain jumping around a lot**. In the second stage, Rosetta uses a full representation of amino acids. This stage is refered to