Part I: Structural Bioinformatics
Chapter 2: From chain polypeptide 1D configuration to 3D

2.1 From chain polypeptide 1D configuration to folded 2D

- Amino acids
- Peptide bond
- Psi and Phi angles
- Ramachandran plot

2.2 Secondary Structure Elements

- Alpha Helix
- Beta sheets
- Turns and Loops
- Coiled coil
- TIM Barrels

2.3 Motifs and Domains

- Homeodomains
- Leucine Zipper
- Zinc Finger
- Transmembrane helices

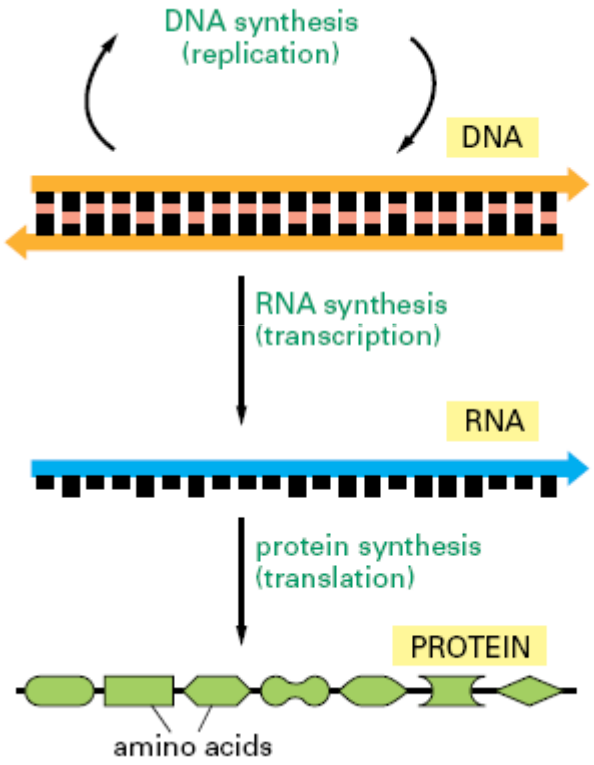
2.4 Tertiary Structure

- Viewers

2.5 First approximation

- PDB- function
- SCOP-Classes
- CAT

2.1 From chain polypeptide 1D configuration to folded 2D



Primary structure: chain translated from the genetic code 20 different amino acids linked by specific type of bond, the peptide bond

Secondary structure: non covalent hydrogen bonds are being formed between the –N-H and -C=O groups a helices or b strands

Tertiary (globular) structure: 2D bonded by loops, turns, non defined structures, etc

Quaternary structure: Association of more than one polypeptide folded chain

Covalent bonds as the strongest can NOT explain the complexity of molecular structure in biology SO it is necessary the inclusion of weaker- non covalent bonds

2.1 From chain polypeptide 1D configuration to folded 2D

Non Covalent

Van der Waals: Determine the shape of molecular surfaces and the maximal packing macromolecules can adopt

Water: Universal environment the life has selected: permanent dipole + excellent solvent due to its hydrogen bounding potential

Hydrophobic- Hydrophilic Interactions

Surround the compound by **hydratation shells** covering the acceptor group

Hydrogen bonds: Determine the conformation and folding ways of macromolecules

- Responsible for the 2D,3D and 4D structure of proteins and nucleic acids
- Fundamental importance in biological processes (water)
- In biological compounds only N and O as hydrogen bond donors
- Highly directional: donor H tends to point directly to the acceptor e^- pair
- Greater energy than most other non covalent interactions

Hydroxy compounds (-OH) ; amines (-NH2); sulfydryl compounds(-SH); esteres(-CHO) ketones (-C=O)

2.1 From chain polypeptide 1D configuration to folded 2D

➢ Amino acids

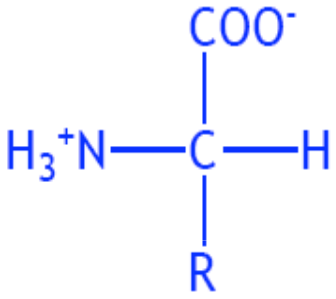
General formula NH2CaHRCOOH: differing on R group attached

pH= 7 amino and carboxylic acid groups ionize to NH3+ and COO- (dipole)

Chirality of Cα : enantiomers or optical isomers that can not be superimposable on its mirror image Proteinogenics L-amino acids

64 possible code combination

Single-base changes elsewhere in the codon produces a different amino acid but with similar physical-chemical properties



Four atoms linked to the Cα

-Hydrogen atom
-R side chain
-NH2
-COOH

2.1 From chain polypeptide 1D configuration to folded 2D

POLAR AMINO ACIDS

Negative

Aspartic acid	Asp D (-3.5)
Glutamic acid	Glu G (-3.5)

Positive

Arginine	**Arg R (-4.5)**
Lysine	**Lys K (-3.9)**
Histidine	His H (-3.2)

Uncharged

Asparagine	Asn N (-3.5)
Glutamine	Gln Q (3.5)
Serine	Ser S (-0.8)
Threonine	Thr T (-0.7)
Tyrosine	Tyr Y (-1.3)

Hydrophilic

NON POLAR AMINO ACIDS

Alanine	Ala A (1.8)
Glycine	Gly G (-0.4)
Valine	**Val V (4.2)**
Leucine	Leu L (3.8)
Isoleucine	**Ile I (4.5)**
Phenylalanine	Phe F (2.8)
Tryptophan	Trp W (-0.9)
Methionine	Met M (1.9)
Proline	Pro P
Cysteine	Cys C (2.5)

Hydrophobic

2.1 From chain polypeptide 1D configuration to folded 2D

	Gly	Ala	Val	Leu	Ile	Met	Cys	Ser	Thr	Asn	Gln	Asp	Glu	Lys	Arg	His	Phe	Tyr	Trp	Pro
Gly																				
Ala	58																			
Val	10	37																		
Leu	2	10	30																	
Ile		7	66	25																
Met	1	3	8	21	6															
Cys	1	3	3		2															
Ser	45	77	4	3	2	2	12													
Thr	5	59	19	5	13	3	1	70												
Asn	16	11	1	4	4			43	17											
Gln	3	9	3	8	1	2		5	4	5										
Asp	16	15	2		1			10	6	53	8									
Glu	11	27	4	2	4	1		9	3	9	42	83								
Lys	6	6	2	4	4	9		17	20	32	15		10							
Arg	1	3	2	2	3	2	1	14	2	2	12	9		48						
His	1	2	3	4			1	3	1	23	24	4	2	2	10					
Phe	2	2	1	17	9	2		4	1	1					1	2				
Tyr		2	2	2	1		3	2	2	4			1	1		4	26			
Trp				1				2							3		1	1		
Pro	5	35	5	4	1		1	27	7	3	9	1	4	4	7	5	1			

Substitution frequencies between amino acids in the same protein from different organisms

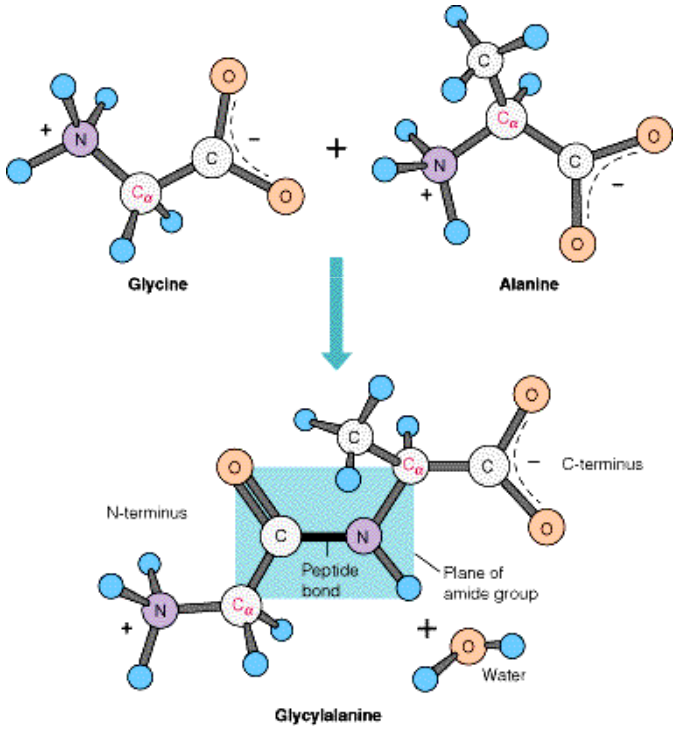
The larger the frequency the more common a substitution is

2.1 From chain polypeptide 1D configuration to folded 2D

➢ Peptide bond

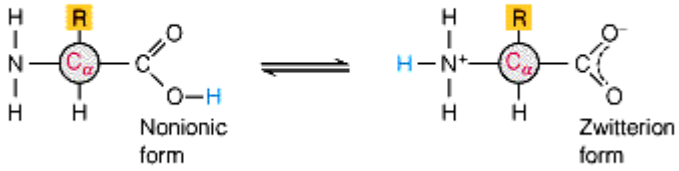
Carboxyl acid –COOH + amino -NH_2 + water

Amide bond
Covalent nature



Zwitterion: Dipolar form at pH=7

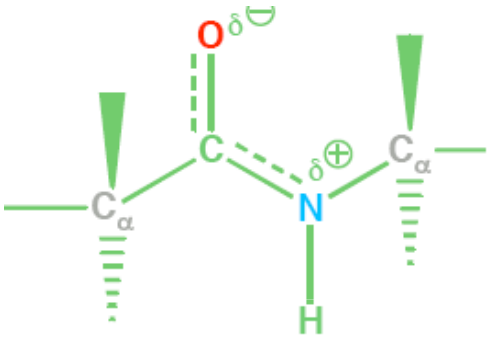
Whole charge is neutral



Protein backbone as blocks of repetitive N-Cα-N
Free amino group: N-terminus
Free carboxyl group: C-terminus

2.1 From chain polypeptide 1D configuration to folded 2D

➢ Peptide bond



Consequences

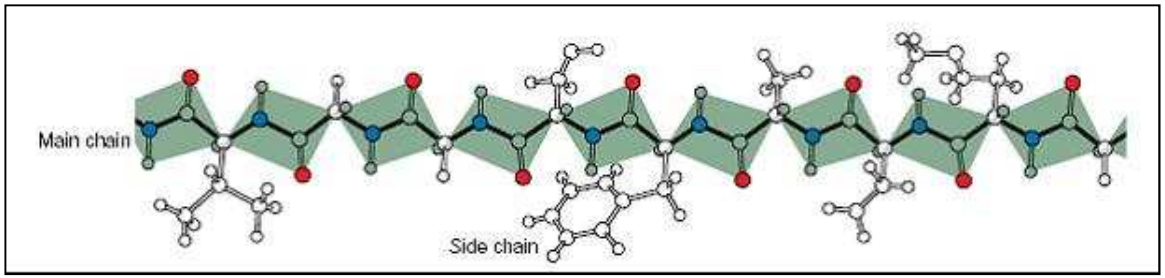
Resonance: partial double bond character (delocalized pair of e-)

Increasing polarity m = qx

Coplanarity and no free rotation for the axis O=C=N

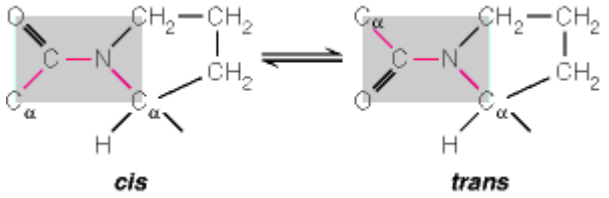
Free rotation for N-Cα and Cα-C

Stability and flexibility of polypeptide chains in water



Cis- (Π) and *trans*-($_L$) possible conformations for two adjacent Cα

Trans-configuration is the most likely except for proline



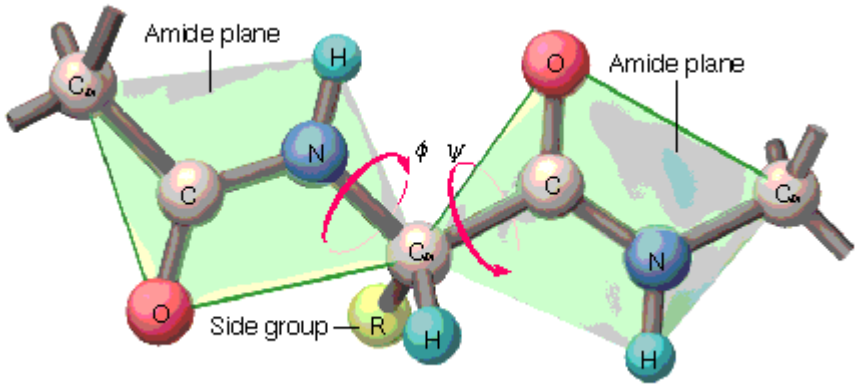
2.1 From chain polypeptide 1D configuration to folded 2D

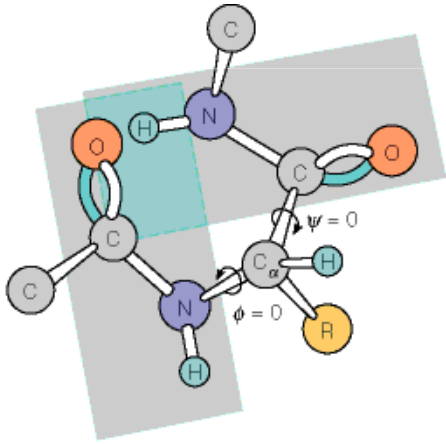
➢ Psi and Phi angles

Rotation allowed **only** for the torsion angles phi and psi
Included within the backbone *dihedral angles* of proteins

N-Cα *phi* (Φ) torsion angle : close to values of 180° (trans-conformation) or 0° (cis-conformation)

Cα-C *psi* torsion angle (Ψ)





The positive rotation is clockwise

2.1 From chain polypeptide 1D configuration to folded 2D

➢ Ramachandran Plot

How secondary structure elements are arranged

Possible conformation based on individual amino acid dihedral values in a polypeptide

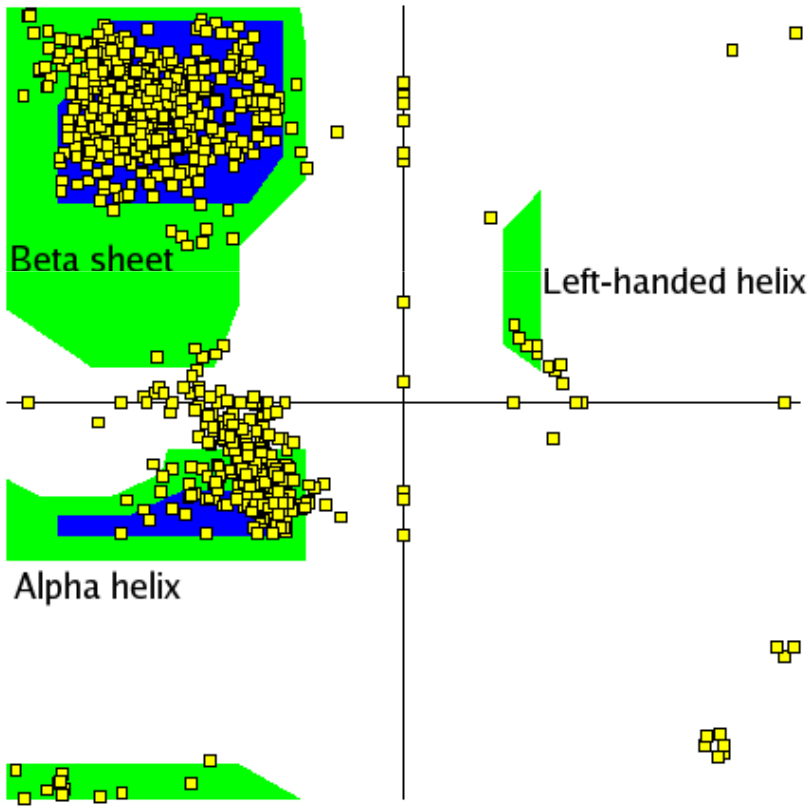
Positive rotation following clockwise (from left to right)

Negative rotation opposite direction

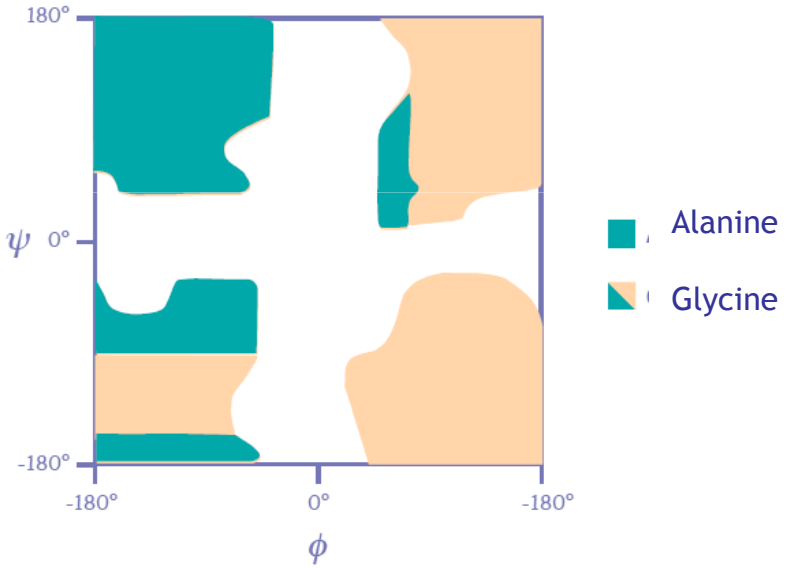
Conformation	Phi (N-Cα)(Φ°)	Psi (Cα-C)(Ψ°)
Right-handed α helix	-57	-47
Left-handed α helix	+57	+47
3_{10} helix	-49	-26
Antiparallel β sheet	-139	+135
Parallel β sheet	-119	+113
Turn II (second residue)	-60	+120
Turn II (third residue)	+90	0
Extended chain	-180	-180

2.1 From chain polypeptide 1D configuration to folded 2D

➢ Ramachandran Plot



Diagnosis method: values allowed for a experimentally solved protein structure



Gly smaller van der Waals radius (-H): less restrictive ; larger combination for phi and psi
Ala larger van der Waals radius (-CH3): more restrictions

Proline is an indicator of turns and loops due to the -N in the ring

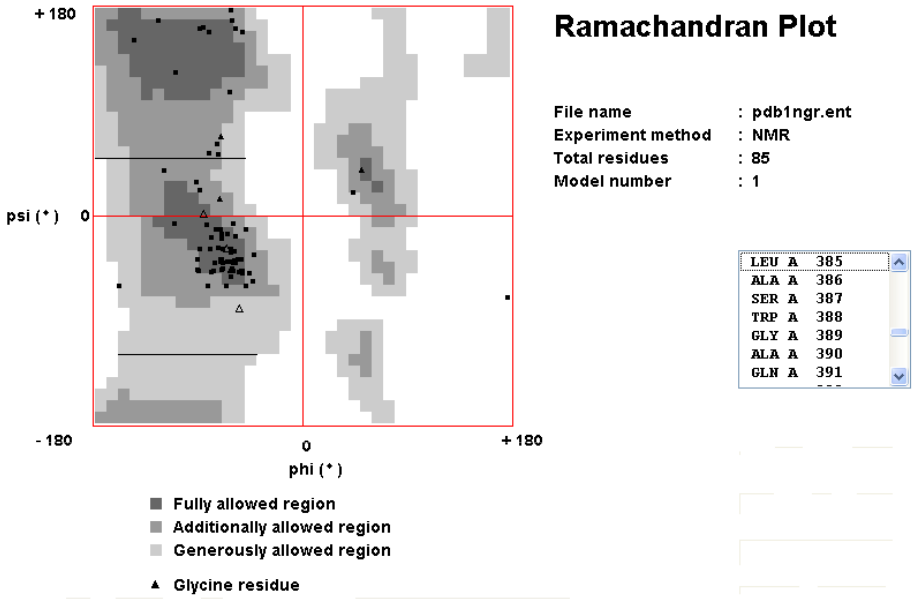
2.1 From chain polypeptide 1D configuration to folded 2D

➢ Ramachandran Plot

Web resources to generate your own plots
http://dicsoft1.physics.iisc.ernet.in/rp/

Example for A.6 Death domain of p75 PDB 1NGR

http://www.fos.su.se/~pdbdna/input_Raman.html



Conclusions

- Every backbone conformation of any particular residue in any protein could be described by specifying those **two angles**
- In similar SSs types all residues would be drawn as superimposable points because are in equivalent conformation and hence have corresponding Phi and Psi angles
- The allowed conformations of a polypeptide chain depend on the bulkiness of the side chains and consequently on the amino acids residue constitution

2.2 Secondary Structure Elements

Empirical rules to follow

- Any amino acid can be found in any type of SSE
- Whether a segment of sequence will be helical, form a turn, a coiled coil, a b sheet or adopt irregular conformation
- Normalized preferences values of individual amino acids
- Proline is the only one that has a cyclic side chain disfavored in both a helix and b sheet
- Glycine as it has a lack in one side, can adopt a much wider range of phi and psi angles values
- Pro-Gly and Gly-Pro in turns as **“beta turns predictors”**
- Proline produces a curve which arises to loops formation at the ends of a helices

2.2 Secondary Structure Elements

AMINO ACID	ALPHA HELIX	B STRAND	REVERSE TURN
ALA	1.41	0.72	0.82
LEU	1.34	1.22	0.57
MET	1.30	1.14	0.52
GLN	1.27	0.98	0.84
GLU	1.59	0.52	1.01
LYS	1.23	0.69	1.07
ARG	1.21	0.84	0.90
HIS	1.05	0.80	0.81
VAL	0.90	1.87	0.41
ILE	1.09	1.67	0.47
PHE	1.16	1.33	0.59
TYR	0.74	1.45	0.76
CYS	0.66	1.40	0.54
TRP	1.02	1.35	0.65
THR	0.76	1.17	0.90
GLY	0.43	0.58	1.77
ASN	0.76	0.48	1.34
PRO	0.34	0.31	1.32
SER	0.57	0.96	1.22
ASP	0.99	0.39	1.24

Preferences normalized values of individual amino acid to be found within specific SSEs

2.2 Secondary Structure Elements

Composition Vs interaction influence stability, function and state folding

Hydrophobic residues:

Van der Waals interactions	hydrophobic effect
Hydrogen bonds	alpha helix (Ala and Leu)

Hydrophilic residues:

Hydrogen bonds:	Water, one to another, peptide backbone
	Polar molecules
	Surface Asp, Glu, Lys (do ionize)
	Ser, Thr (Do not ionize)
	Active site His (Double donor donor-acceptor)
Disulfide bonds:	Active site Cys
	Nucleophile anion (thiolate)

Amphipatic residues (interfaces):

Van der Waals interactions	Hydrophobic side chains one to another
	Tyr (donor-acceptor)
Weak polar interactions	Trp (aromatic ring)

2.2 Secondary Structure Elements

INTERACTION	EXAMPLE	DISTANCE DEPENDENCE	TYPICAL DISTANCE(Å)	FREE ENERGY (kJ/mol) (bond dissociation enthalpies for the covalent bonds)
Covalent bond	Cα-Cα	-	1.5	356
Disulfide Bond	-Cys-S-S-cys	-	2.2	167
Hydrogen bond	-NH—O=C-	Donor(N) and acceptor(O)	3.0	2-6 in water and 12.5-21 if either donor and acceptor is charged
Van der Waals	-CH3-CH3	Short range and falls rapidly beyond 4 Å separation	3.5	4 (4-7in protein interior) depending on the size of the group

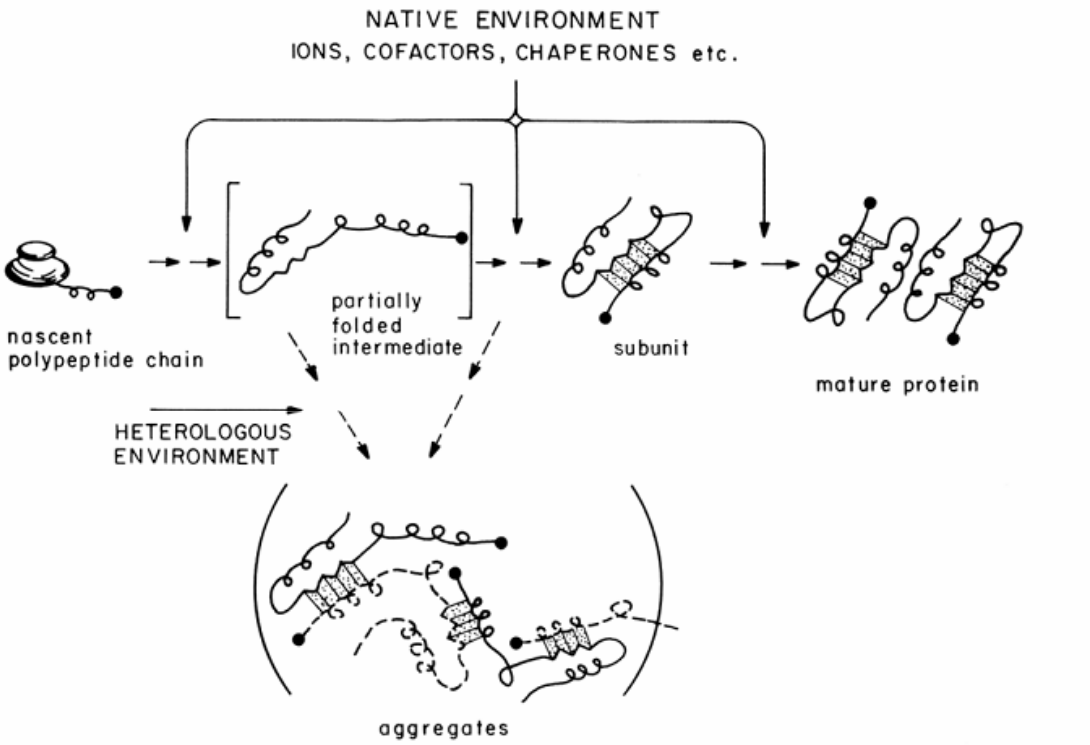
Residues and peptide bond chemical-physical properties

Folding: Space + Correctness + Time

Weak interactions addition increasing the free energy and stability

Evolution: maximal ratio Native state/time (Chaperones)

2.2 Secondary Structure Elements



Nucleation points to build up the active protein

Polar backbone hydrogen bonding with each other and hydrophilic polar side chains on the surface interacting with water

Aggregates when no optimal environment conditions

To satisfy their hydrogen-bonding potential hydrophobic residues interact with themselves leaving the secondary structure elements to form

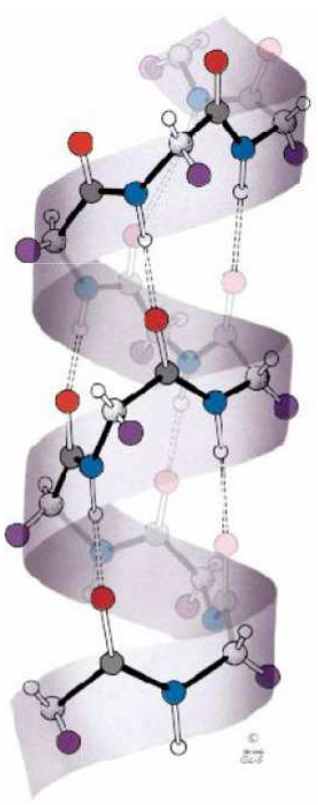
2.2 Secondary Structure Elements

Some examples for key amino acids due to their chemical and physical properties

- Movie: Active site 1 (Lactato Dehydrogenase)
  - Arg -171 and His-195
- http://www.youtube.com/watch?v=swEc_sUVz5I
- http://www.youtube.com/watch?v=BrUdCVwgJxc&feature=related
- Movie: Active-site 2
  - His-57, Ser 195 and Asp 102

2.2 Secondary Structure Elements

➢ Alpha Helix



Cylindrical structures stabilized by a network of backbone hydrogen bonds (–CO on residue n and the –NH on residue n+4)

One full turn occurs every 3.6 residue (rotation of 100°) extends the length of the helix by 0.5 nm

Distance between consecutive residues 1.5Å

Interactions do not involve side chains

Right –handed favored due to steric constrains of the L-Aas

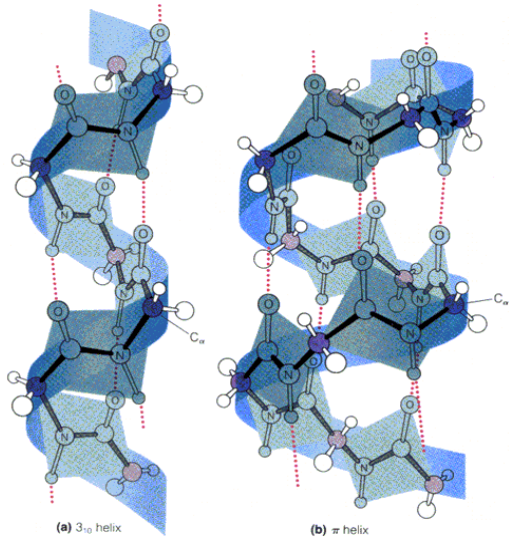
Interaction with other helices, charged chains, ions and molecules

Amphiphatic property: Protuberating formed by amino acids projected outward from the same face and regular rotation (helix-helix packing)

Macrodipolo formed by the accumulative effect of every individual peptide dipolo (NH_3^+ terminus and $-COO^-$ terminus)

➢ Alpha Helix

CONFORMAIION	PHI (°)	PSI(°)	RESIDUES PER TURN	TRANSLATION PER RESIDUE(distance from two consecutive residues) (Å)
Alpha helix	-57	-47	3.6	1.5
3-10 helix	-49	-26	3.0	2.0
Pi-helix	57	-70	4.4	1.15
Polyproline I	-83	+158	3.33	1.9
Polyproline II	-78	+149	3.0	3.12
Polyproline III	-80	+150	3.0	3.1



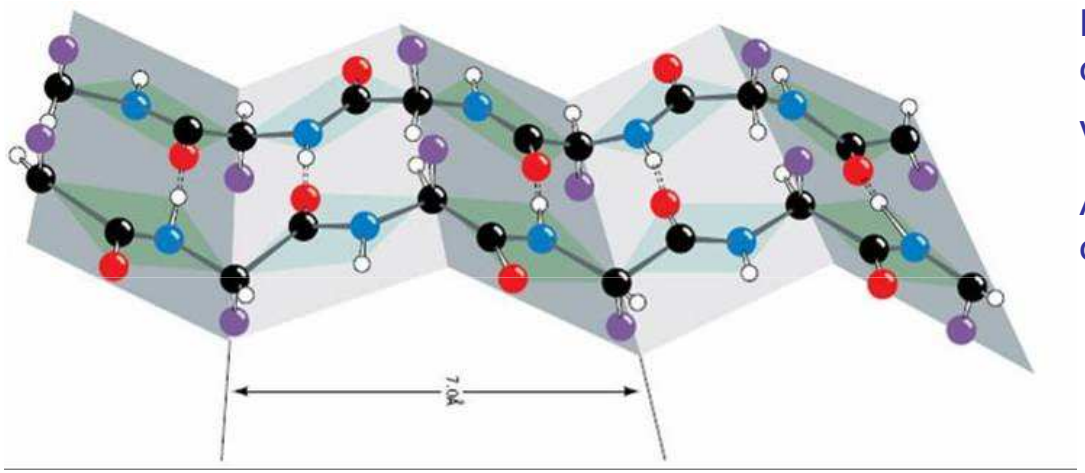
Low stability

No length limit BUT for longer length helices it would coil about the helix axis and for same pattern of hydrophobic groups, four residues apart they would form a coiled coil

Pi-helix sterically possible but not yet observed

2.2 Secondary Structure Elements

➢ Beta Sheet



Interactions do NOT involve side chains

Right –handed favored due to steric constrains

Val and Ile

Amphiphatic property due to trans-conformation of amino acids

Hydrogen bonds between backbone atoms on adjacent regions

Two or more strands separated in the protein are arranged side by side

Distance between two consecutive residues is 3.3 Å

Represented as a series of flattened arrows pointing towards the protein's Carboxy terminal end

2.2 Secondary Structure Elements

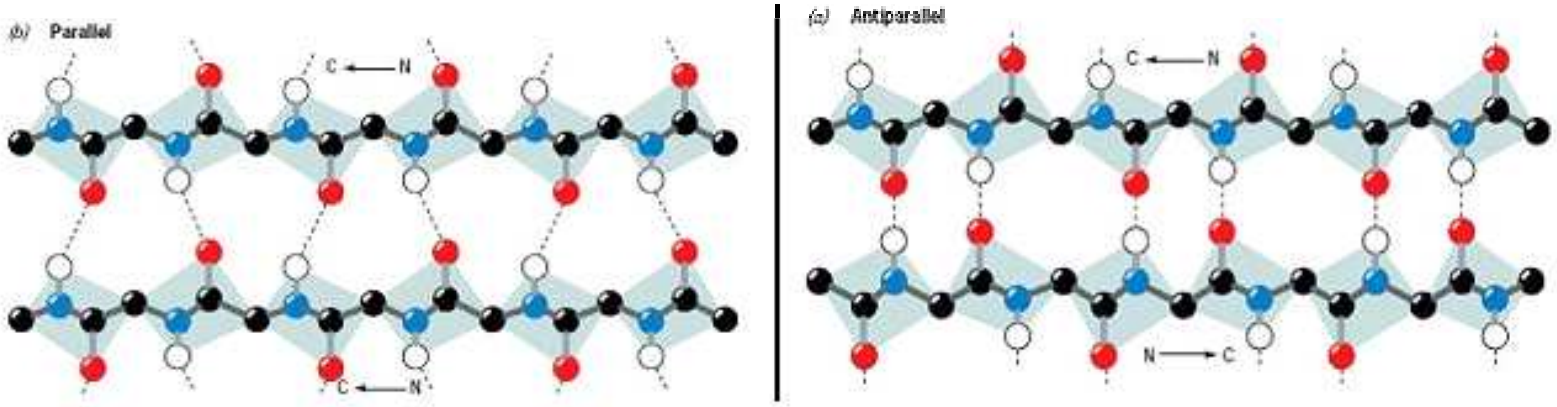
➢ Beta Sheet

Interactions between -NH and –COOH groups on the outer side with water, adjacent b strands, helices, etc

Beta barrels or cylinders formation:

Last strand of the edge interacts with the first one

Stabilization of quaternary structure



Less stable: Internally buried

Connected via complex unions (helices)

Stronger final molecule

More stable: Exposed

Connected via turns reversing direction

2.2 Secondary Structure Elements

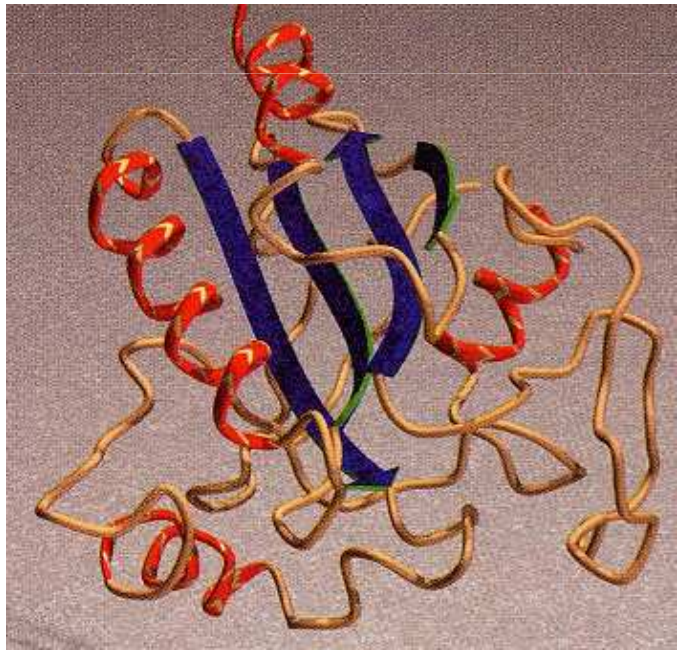
➢ Turn and Loops

Simplest SSEs

Or hairpin reverse turn or beta turn

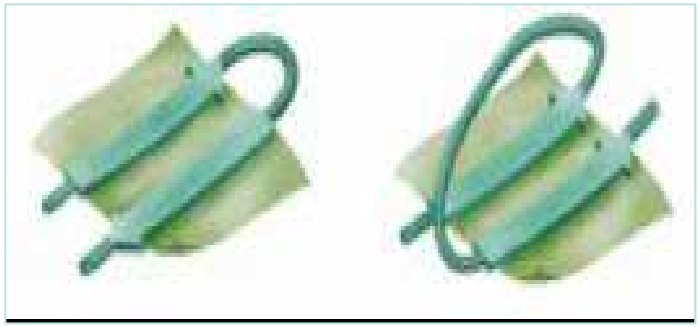
Hydrogen bond between the –CO on residue n and the –NH on residue n+3

Reversion in the direction



Limit the size of the molecule and maintain the compact state

Hydrogen bond with water molecules avoiding the four residues to interact

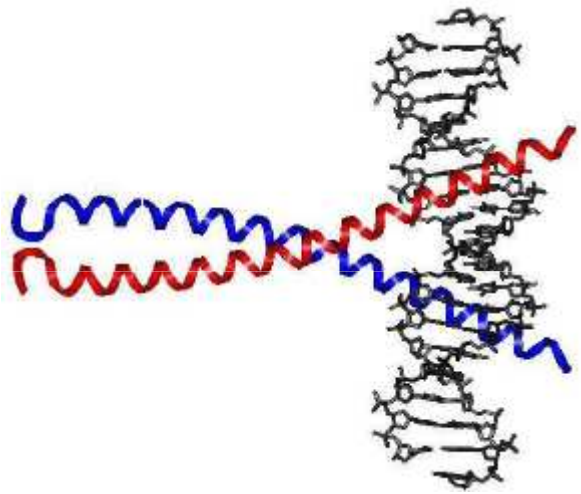


Placed in the surface of folding proteins

Gluthatione peroxidase (1GP1)

2.2 Secondary Structure Elements

➢ Coiled coil



Two to five right-handed amphiphatic α helices wrapped around each other with a left-handed super-helical twist

Associated in parallel or antiparallel orientation

May be the same (homo-oligomer) or different (hetero-oligomer)

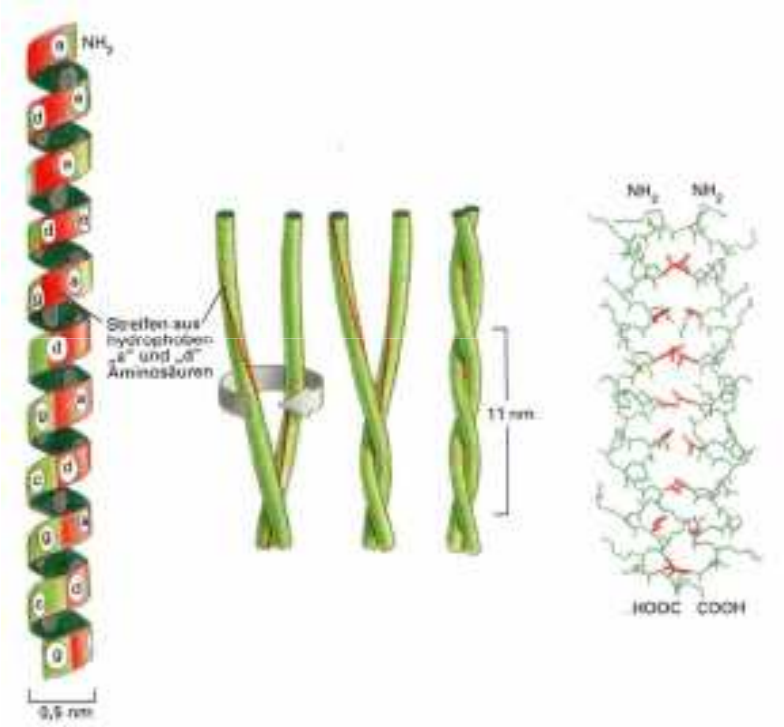
Amphiphatic property

Their hydrophobic sides snuggle tightly together in the center

Stable hydrophobic core

2.2 Secondary Structure Elements

➢ Coiled coil



"Peptide Velcro hypothesis" as the most favorable way for helices to arrange in an aqueous environment: wrap around each other so hydrophobic surface is buried

High ubiquity: 3-5% on the sequence database

Heptad repeat $(abcdefg)_n$ spread out along two turns of the helix

Positions **a** and **d** are hydrophobic , **e** and **g** are charged and **b, c, f** are hydrophilic

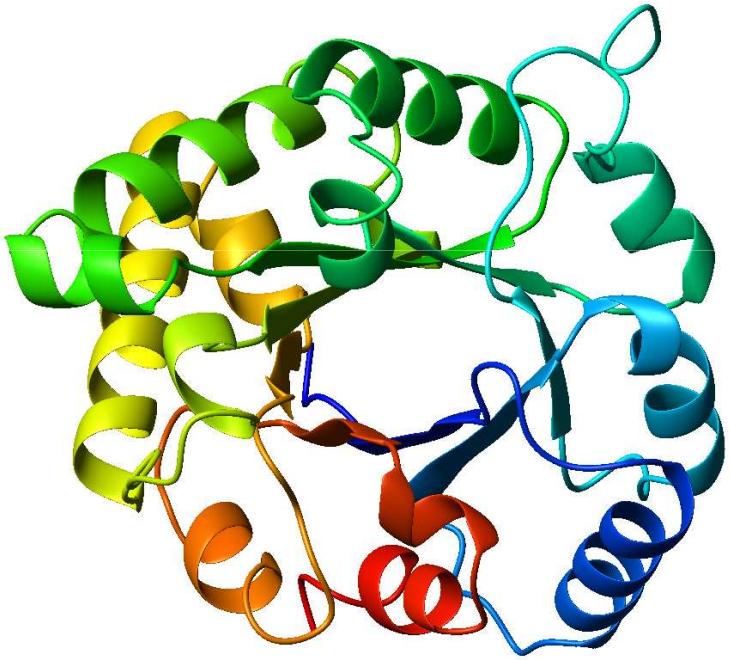
Found in elongated, fibrous proteins as fibrinogen (Blood clotting)

Transcription factor in yeast GCN4

Avian Flu Virus

2.2 Secondary Structure Elements

- **TIM barrels**



Triosephosphateisomerase

A b sheet strand followed by an a helix repeated eight times

Catalytical function of the protein

α-helices and ß-strands form a solenoid that curves around to close on itself in a ring shape (toroid)

The parallel ß-strands form the inner wall of the ring→ ß-barrel
The α-helices form the outer wall of the ring

2.3 Motifs and domains

➢ Motifs

"*A hree-dimensional structural element or fold within the chain, which appears also in a variety of other molecules* "

-Does not need to be associated with a sequence motif

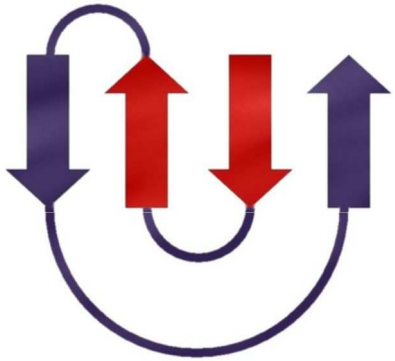
-Direct involved in protein function

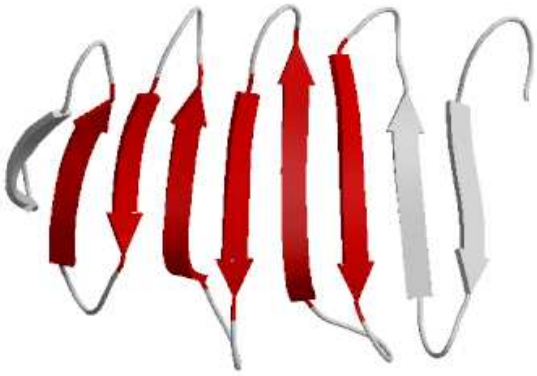
-Greek key-three antiparallel strands connected by hairpins, while the fourth is adjacent to the first and linked to the third by a longer loop

-The ß-α-ß motif (TIM barrel)-right-handed" twist linked by an helical region

-ß-meander motif-2 or more consecutive antiparallel ß-strands linked together by hairpin loops

-Psi-loop motif-two antiparallel strands with one strand in between that is connected to both by hydrogen bond





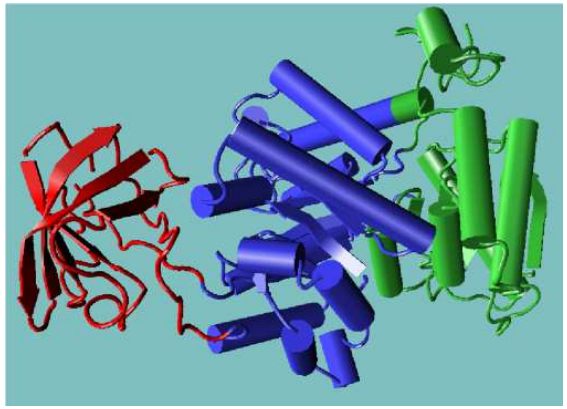
2.3 Motifs and domains

➢ Domains

" *A protein domain is a part of protein sequence and structure that can evolve, function, and exist independently of the rest of the protein chain. Each domain forms a compact three-dimensional structure and often can be independently stable and folded.*" Wikipedia

- Alpha-
- Beta-
- Alpha/beta-combination of ß-α-ß motifs that predominantly form a parallel ß-sheet surrounded by α-helices
- Alpha +beta -mixture of all-α and all-ß motifs Not used in the CATH database due to overlaps
- Cross linked domains

Are fundamental units of tertiary structure
Each domain containing an individual hydrophobic core built from SS units connected by loop regions



Pyruvate kinase

2.3 Motifs and domains

➢ Homeodomains

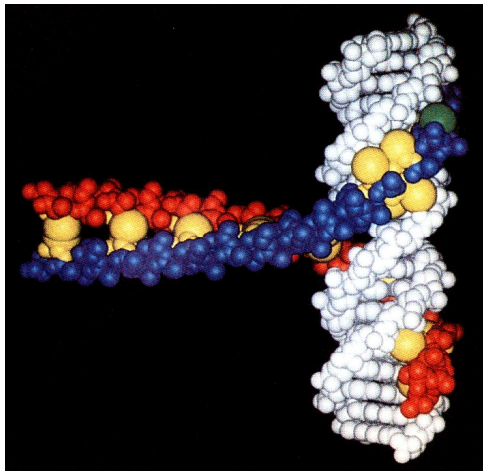
Are found in many transcription factors binding to DNA (TATA box)

Three overlapping a helices packed together by hydrophobic forces (about 60 Aas long)

Three side chains from the recognition helix form hydrogen bonds with bases in the DNA

Msx-1 Homeobox gene

Transcriptional repressor



Leucine zipper

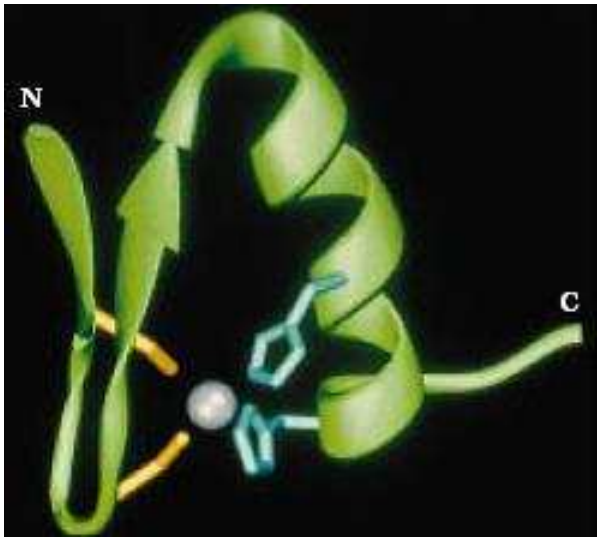
Two long intertwined a helices

Hydrophobic side chains extend out from each helix into the space shared between them

Tight packing of side chains between the leucine zipper helices especially stable

2.3 Motifs and domains

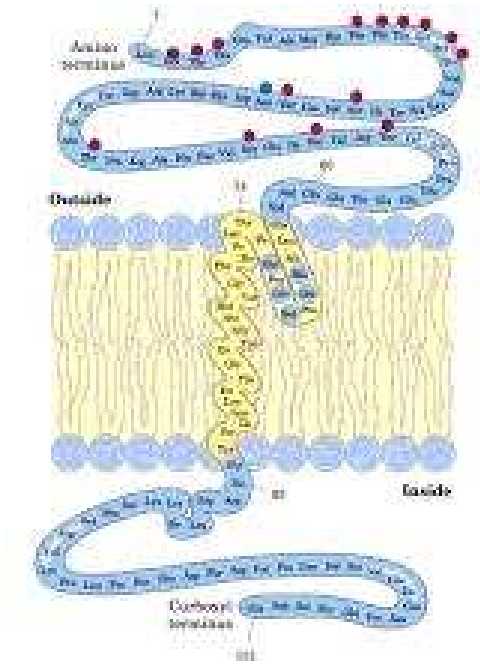
- **Zinc finger**
  - Structural motifs used by a large class of DNA-binding proteins
  - Coordinated zinc atoms as crucial structural elements
  - Single zinc finger domain is only large enough to bind a few bases of DNA (found in tandem)
  - Helical region of each zinc finger rests in the major groove of the DNA helix
  - Modulation of DNA and gene expression



HIV and potential drug target

2.3 Motifs and domains

➢ Transmembrane elements

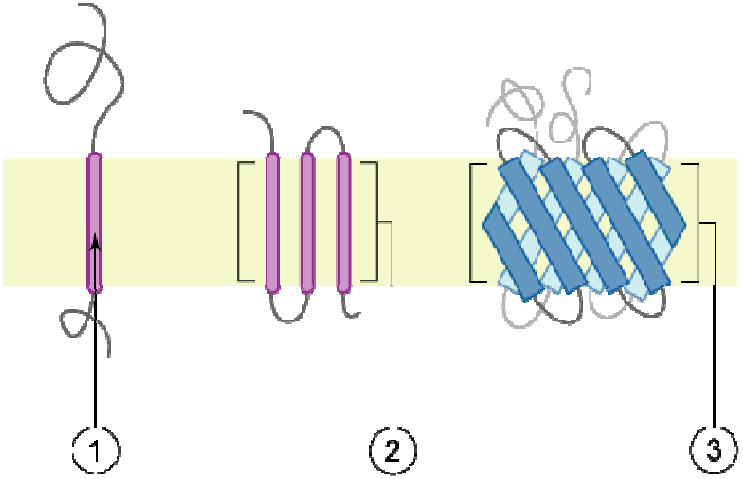


Glycophorin C protein

Single transmembrane domain

Proteins crossing the entire membrane
Chemical-physical characteristic: aggregate and precipitate in water
Elements formed very early in the folding process as nucleation point

Integral membranes proteins to be unusually stable: high levels of energy invested to break down the hydrogen bonds

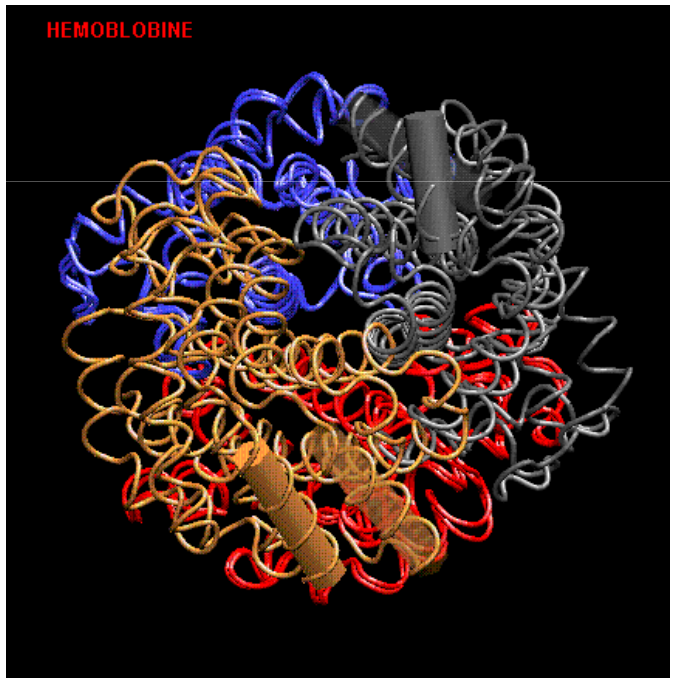


Critical for maintaining the shape and stability of erythrocyte

http://courses.cm.utexas.edu/emarcotte/ch339k/fall2005/Lecture-Ch11/Slide11.JPG

2.4 Tertiary Structure

Arrangement of SSEs into a stable and compact fold through weak interactions



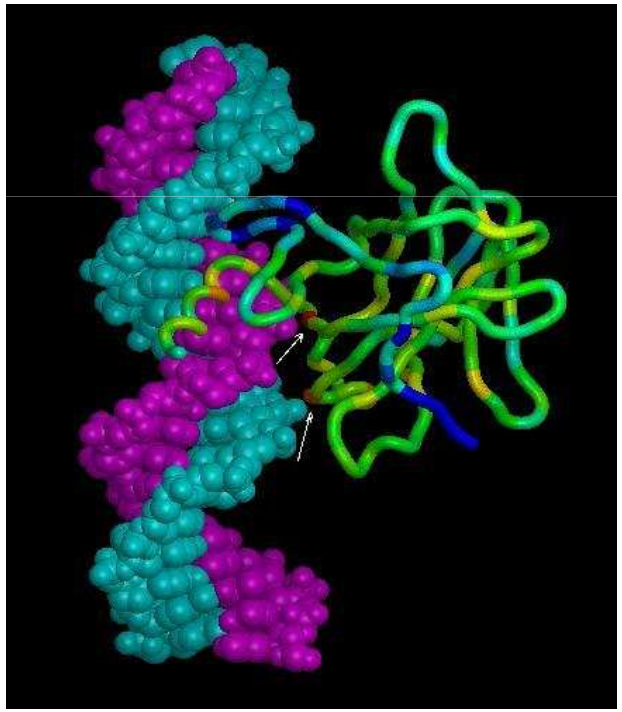
Stabilized by

- Efficient packing of atoms in the internal core
- Water binding to the polar side chains
- Potential-binding groups of the backbone
- Hydration shell surrounding the macromolecule

Source http://emerson.free.fr/images/divers/schemas/hemoglobine.png

2.4 Tertiary Structure

Topologies directly related with the
function surface or region complementarity



Interfaces holding subunits make possible
the communication through them

Three-dimensional structure
in which the protein performs its biological
function

Source http://www.bioinf.org.uk/p53/p53.jpg

2.4 Tertiary Structure

➢ Molecular viewers

Free *molecular visualization* resources

For knowing how the atoms in an a helix are connected to one another

For seeing the relative sizes of the atoms in an a helix

Ribbon b strands as arrows pointing from the N- to the C-terminus and ahelices are shown as twisted cylinders. it does not show individual atoms

Sticks bonds connecting atoms

Ball-and-stick with ball (small sphere) atoms and stick bonds

CPK Corey-Pauling-Koltun sphere full van der Waals radius. Atoms and sticks.

RasMol (Protein Explorer) displays any molecule for which a 3-dimensional structure is available

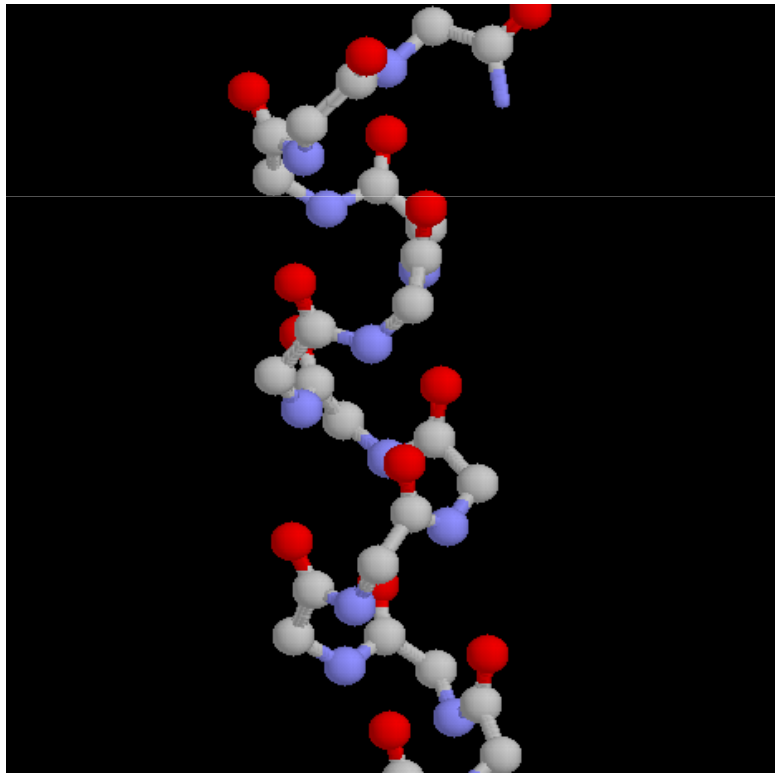
Pymol as a *molecular graphics system Python interpreter*

Chime a browser plug-in that renders 2D and 3D molecules directly within a Web page

http://av.bmbq.uma.es/av_biomo/
http://www.umass.edu/microbio/rasmol/index2.htm
http://www.mdl.com/products/framework/chime/index.jsp
http://pymol.sourceforge.net/

2.4 Tertiary Structure

➢Molecular viewers



Carbon: Grey
Oxygen: Red
Hydrogen: White
Nitrogen : Blue

αHelix **Ball** and **Stick View**

Lysozyme

http://project.bio.iastate.edu/Courses/BIOL202/Proteins/secondary_structure.htm

2.5 First approximation

- **PDB**

http://www.rcsb.org/pdb/home/home.do

Binding function
- TATA binding protein (1tgh)
- Mioglobin (1a6k):

Catalysis Function
- HVI protease (1a8k):

Switching
- Ras protein (121p “on”)
- Ras protein (1pll “off”)

Structural proteins
- Silk (1slk):

2.5 First approximation

- **SCOP**

http://scop.mrc-lmb.cam.ac.uk/scop/
http://scop.mrc-lmb.cam.ac.uk/scop/search.cgi?

Class a: Myoglobin

Class b: a-amylase inhibitor

Class a/b: Mainly parallel b strands (beta-alpha-beta patterns). Tryose phosphate isomerase

Class a+b: Mainly antiparallel b strands (separated alpha and beta section). Transglycosilase linked to membrane.

Multidomain proteins: Two or more domains each one from different classes.

Surface and membrane proteins (excluding those from immune system). aHemolysine

Proteins-Ligands

2.5 First approximation

- CATH

http://www.cathdb.info/

Class a

Cytochrome c3(2CDV)

Farnesyl diphosphate synthase (1FPS)

Class b

Ubiquitin(1UBQ)

Protein G-third Ig-binding domain(1IGD)