Part I: Structural Bioinformatics

2. Chemical and Physical Background

Introduction

2.2 Atomic Bounds: A basic introduction

2.2.1 Non covalent Interactions

2.2.1.1 Charge-Charge Interactions

2.2.1.2 Dipole Interactions

2.2.1.3 Van der Waals Forces

2.2.1.4 Hydrogen Bonds

2.2.1.5 Hydrophobic- Hydrophilic Interactions

2.3 From chain polypeptide 1D configuration to folded 2D

2.3.1 Amino acids

2.3.1.1 Peptide bond

2.3.1.2 Psi and Phi angles

2.3.1.3 Ramachandran plot

2.3.2 Amino acids Chemical-physical properties Vs interactions and folding

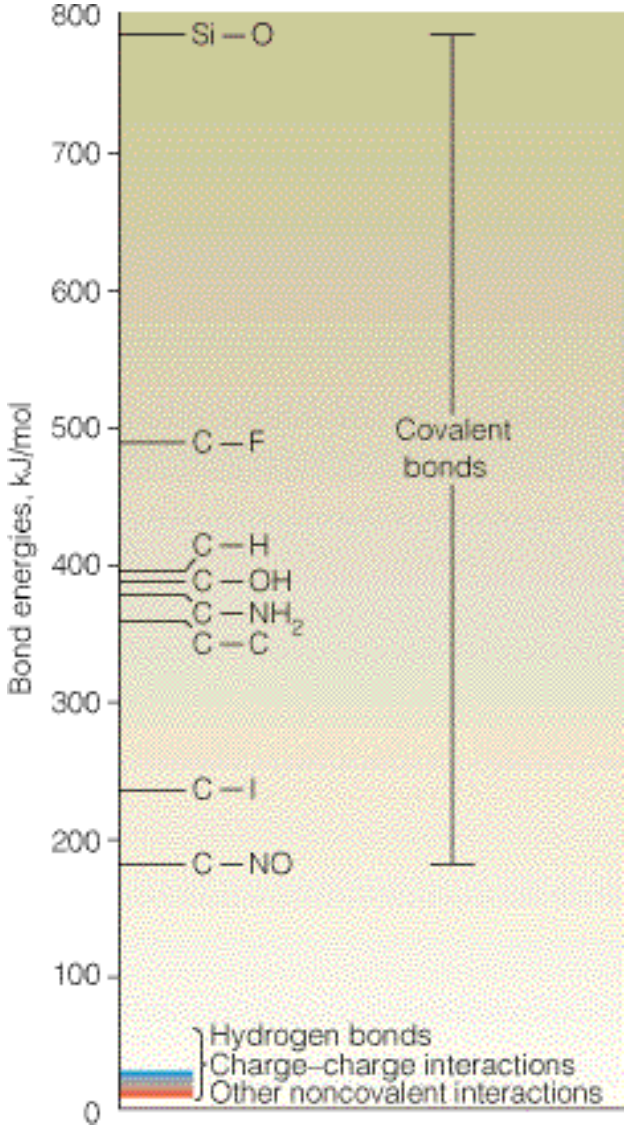
2.3.2.2 Thermodynamics

Covalent bonds:

- Sharing e^-
- Linear-primary Sequence
- C-C, C-H
- 300-400 Kj/mol

Non-covalent bonds:

- e^- Transfer
- 2D-3D structure
- 10-100 times weaker
- Conformational Flexibility
- Protein complexes
- Function stability



2. Chemical and physical Background
2.1 Introduction

Macromolecules are made up of smaller units linked one to the next by specific bonds

Individual constituent atoms: specific bonds, physical- chemical properties

Atoms: three main subatomic particles

Electron: Negatively charged.Located in shells and interacting with adjacent nuclei in the outermost shell

Proton: Positively charged. Llocated in the nucleus and determines the chemical properties

Neutron: uncharged, located in the nucleus and determines the isotope

Strenght H^+ = negative charge of an e^-

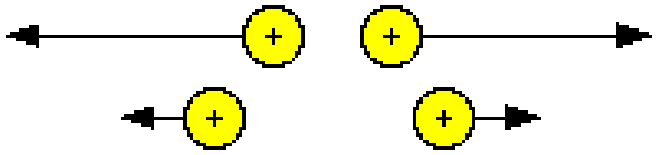
Valence shell governs the behavior: stable e^- configuration of a full v-s

- Depends on the electrostatic state of the molecules
- Weaker with growing distances (Hydrogen Bonds)
- Combined bond strengh greater than sume of the individual bonds
- Favorable Gibbs energy
- Always involve electrical changes

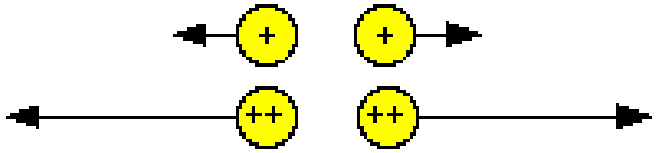
2.2.1.1 Charge-Charge Interactions

Electrostatic forces between two oppositely charged ions
Coulombs Law: force between charges q1 and q2 separated by a distance r

Direct dependency on q1-q2 and r
Greater q1-q2 stronger F
Larger r weaker F



double the distance, force drops to 1/4



double the charge,
force increases by factor of 4

Crystals of salts NaCl

Inside and outside a cell charges separated

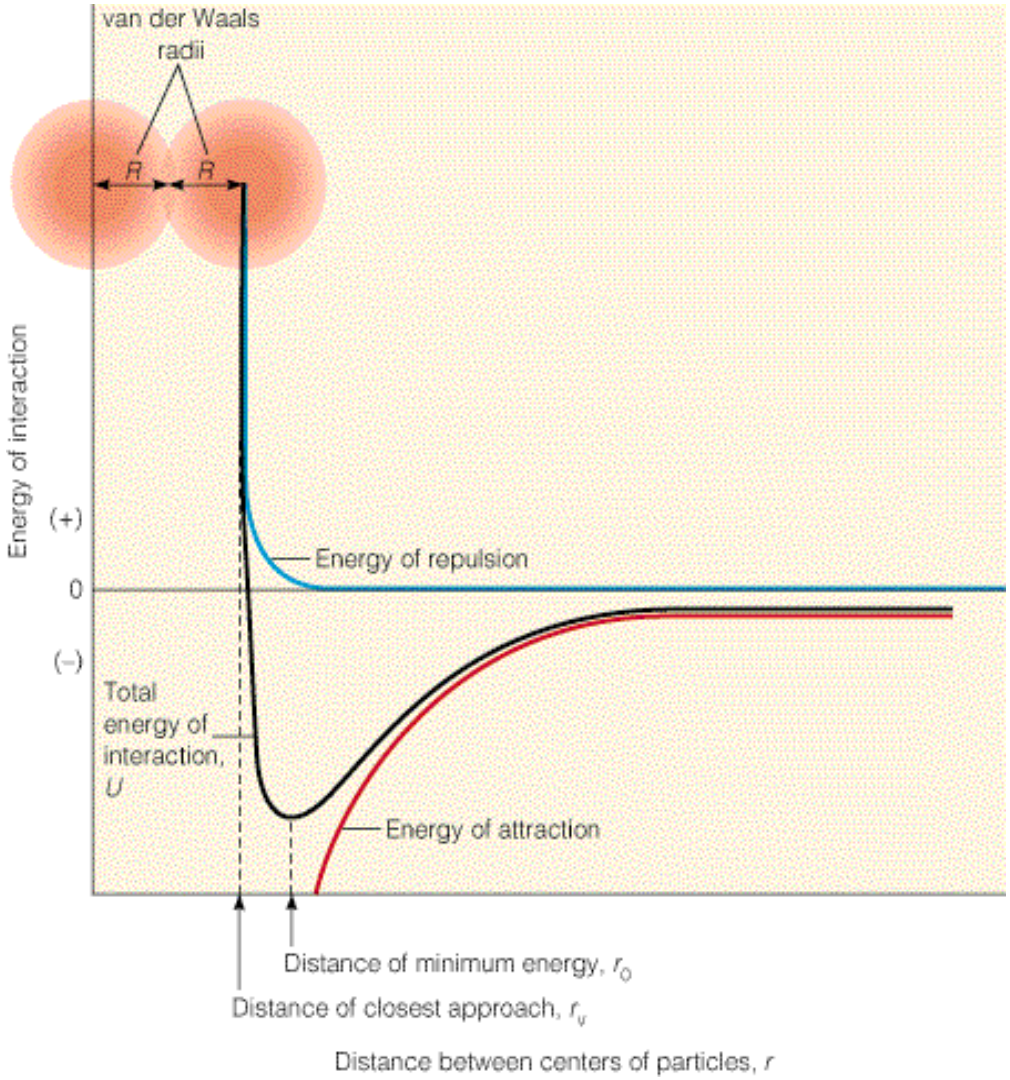
Dielectric constant : represents
the effect of the biological environment in which the actual force is

$$F = k(q1q2)/(\mathbf{e}\ r2)$$

ε= Water 80
ε = Organic substances 1-10

Interaction energy U

$$U = k(q1q2)/(e\ r)$$



- F between charges depends only on distance
- U inversely proportional to the first power of r
- A barrier approach between two charged molecules, ions or atoms does exist (r_v)
- Minimum energy close to closest aproach barrier (r_0)

2.2.1.2 Dipole Interactions

Assymmetrical e^- distribution in the outermost shell
Non charged molecules become partially charged : Dipole

Behavior depends on the surrounding medium, particles and their charges

- Induced dipole
- Permanent dipole
- Instantaneous dipole

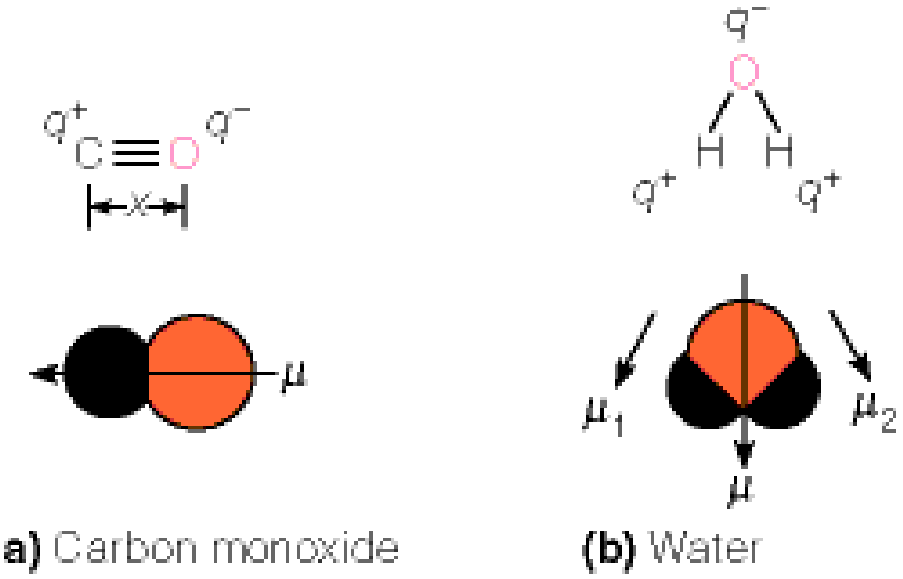
Dipole moment **m**: asymmetry of a molecule by its polarity measurement

$m = qx$

q vector pointing towards q+

x distance between the ionized-charged groups

The larger the distance the greater the dipole moment



(a) Carbon monoxide **(b)** Water

μ results of the negative partial charge of oxygen and positive charge of the carbon and hydrogen atoms

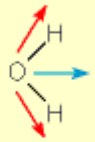
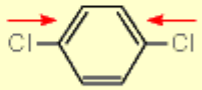
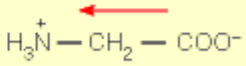
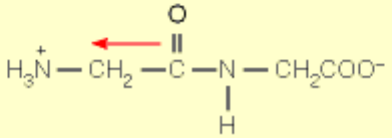
One dipolar moment μ1 CO

Two dipolar moments μ1 and μ2 OH

Global dipole moment results as sum of the single dipole moments of the sub-constitutive molecules

Molecules must be asymmetric to have a dipole moment

2. Chemical and physical Background
2.2.1 Non covalent Interactions

Molecule	Formula	Dipole Moment (D)[a]
Carbon monoxide		0.12
Carbon dioxide		0
Water		1.83
ortho-Dichlorobenzene		2.59
para-Dichlorobenzene		0
Glycine		16.7
Glycylglycine		28.6

[a]The common units of dipole moment are *debyes*; 1 debye (D) equals 3.34×10^{-30} C m.

Symmetry

Distance

– *Charge-dipole* interactions: Permanent-dipole. Strenght depends on the orientation of the molecules ($NH3^+$-H_2O)

– *Dipole-dipole* interaction: „Ionic interaction“ with only partial charges involved. Strenghts depends on mutual orientation of dipoles (H_2O- H_2O)

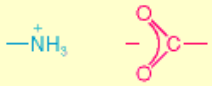
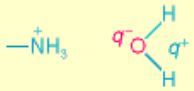
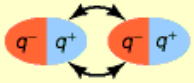
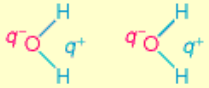
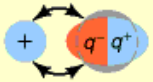
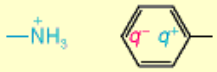
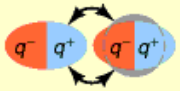
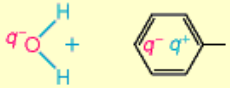
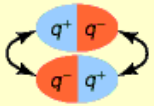
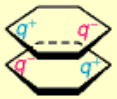
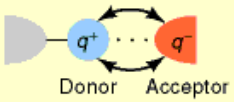
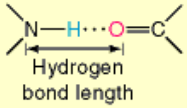
– *Induced-dipole* interactions: Molecule adopts transient partial charge within an electric field

- Induction by anion-cation charged molecule ($NH3^+$-Benzene)
- Induction by permanent dipole (H_2O-Benzene)

– *Induced-induced dipole* interactions: Instantaneous dipoles. Two uncharged molecules are close enough to synchronize the fluctuation of their e^- (attractive force) (Benzene-Benzene)

– U decreases from the *charge-dipole* interaction to *induced-induced dipole* interaction: $1/r^2$ to $1/r^6$

2. Chemical and physical Background
2.2.1 Non covalent Interactions

Type of Interaction	Model	Example	Dependence of Energy on Distance
(a) Charge–charge Longest-range force; nondirectional			$1/r$
(b) Charge–dipole Depends on orientation of dipole			$1/r^2$
(c) Dipole–dipole Depends on mutual orientation of dipoles			$1/r^3$
(d) Charge–induced dipole Depends on polarizability of molecule in which dipole is induced			$1/r^4$
(e) Dipole–induced dipole Depends on polarizability of molecule in which dipole is induced			$1/r^5$
(f) Dispersion Involves mutual synchronization of fluctuating charges			$1/r^6$
(g) van der Waals repulsion Occurs when outer electron orbitals overlap			$1/r^{12}$
(h) Hydrogen bond Charge attraction + partial covalent bond			Length of bond fixed

2.2.1.3 Van der Waals Forces

Attraction between temporarely induced, short-living dipoles in non polar molecules

Polarization induction due to

Polar molecule

Repulsion of negatively charged e^- clouds in non polar molecules

Mutual repulsion by the overlapping of outer e- orbitals

Repulsion increases proportionally to $1/r^{12}$

Each atom o molecule represented as a sphere

Chlorine dissolved in water: Electric field of the aqueous solution

+ - + - +

Permanent Dipole H-O-H-----Cl-Cl Induced Dipole

	R (nm)
Atoms	
H	0.12
O	0.14
N	0.15
C	0.17
S	0.18
P	0.19
Groups	
—OH	0.14
—NH_2	0.15
—CH_2—	0.20
—CH_3	0.20
Half-thickness of aromatic ring	0.17

Defines the minimal distance between interacting molecules

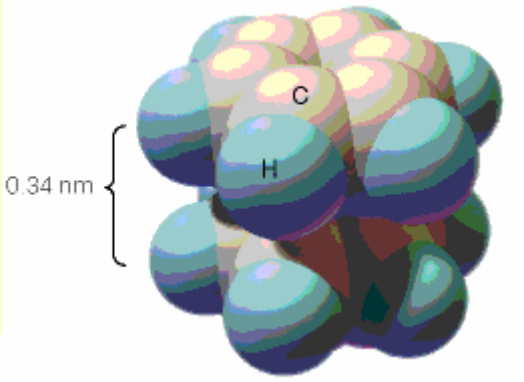
Determine the shape of molecular surfaces

Van der Waals radius R as the effective radius for closest molecular packing

Boundary distance r_v

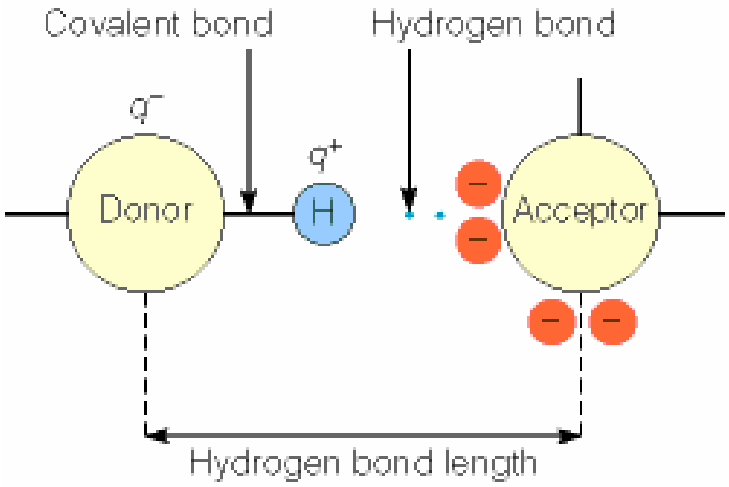
r_v = 2R when R1=E2

r_v = R1 + R2



2.2.1.4 Hydrogen Bonds

- Interaction between an hydrogen atom bound covalently to an electronegative atom and another electronegative atom



hydrogen bond donor: High electonegative + hydrogen bonds partially possitive
(Oxygen, Nitrogen, Fluorine)

hydrogen bond acceptor: Free electron pairs

2. Chemical and physical Background
2.2.1 **Non covalent Interactions**

- Intra-Intermolecular Interaction
- Stronger than Van der Waals interaction but weaker than ionic or covalent bonds
- Responsible for the 2D,3D adn 4D structure of proteins and nucleic acids
- Responsible of the hight boiling point of water(100°C)
- Bond length in biological molecules: *0.28-0.31nm* OH-OH, C=O, N-H
- Ambivalent Character
  Non covalent bonds characteristics : Charge-charge interactions

  Covalent bonds characteristic: Electron are shared i.e.
  Bond length **=N-H-O=C=**
  Non covalent Van der Waals radii : $R_1H + R_2O$ = *0.12nm* + *0.14nm*
  Covalent O-H *0.10nm*
  Actual length *0.19nm*

- Main importance in biological processes as water is the medium in which biological reactions take place

Donor···Acceptor	Bond Length[a] (nm)	Comment
—O—H ··· O(H)	0.28 ± 0.01	H bond formed in water
—O—H ··· O=C	0.28 ± 0.01	Bonding of water to other molecules often involves these
N—H ··· O(H)	0.29 ± 0.01	
N—H ··· O=C	0.29 ± 0.01	Very important in protein and nucleic acid structures
N—H ··· N	0.31 ± 0.02	
N—H ··· S	0.37	Relatively rare; weaker than above

[a]Defined as distance from center of donor atom to center of acceptor atom. For example, in the N—H···O=C< bond it is the N—O distance.

Bond length depends on

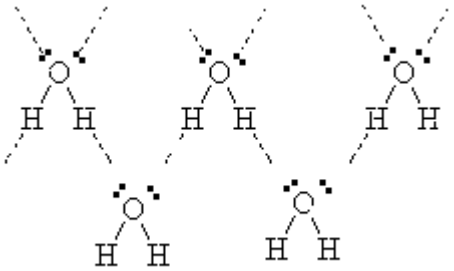
–Temperature

– Pressure

– Bond strength

– Bond angle

– Dielectric constant of environment

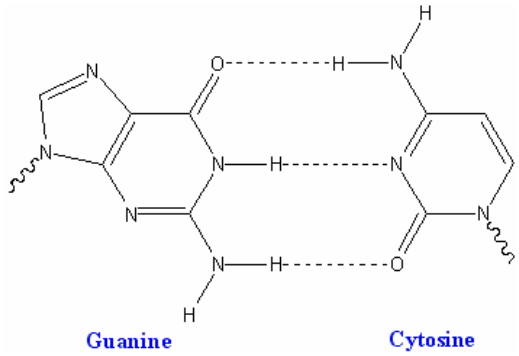


Water can hydrogen bond with up to four other molecules:

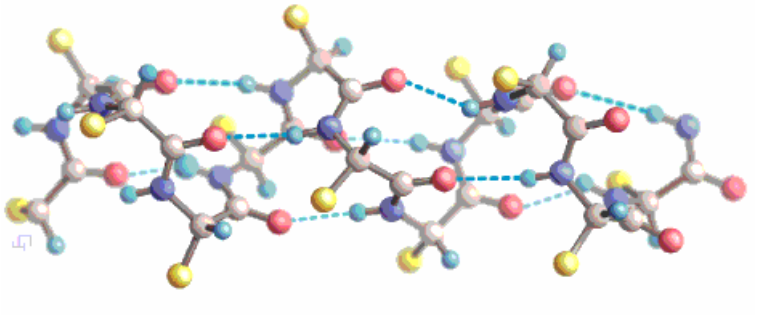
1 Oxygen two lone pairs of e- bonds to two hydrogen atoms of two other water molecules

2 Hydrogen atoms bond to two oxygen of two other water molecules

Macromolecules fold into a specific shape that determines its biological function



Base pairs linked together in DNA molecule



Oxygen and amide hydrogen in backbone of polypeptide chain

2.2.1.5 Hydrophobic- Hydrophilic Interactions

Water as excelent solvent due to ots hydrogen bounding potential

Hydrophilic interaction: ability of an atom or a molecule to engage in attractive interactions with water molecules.

Substances that are ionic or can engage in hydrogen bonding :

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

Water molecules surround the compound by **hydratation shells** covering the acceptor groups

Hydrophobic interaction: unable to engage in attractive interactions with water molecules. Substances are non-ionic and non-polar.

Hydrocarbon (C-H)

Do form clatrates: ordered cages of water molecules around non polar molecules

Pink: Oxygen

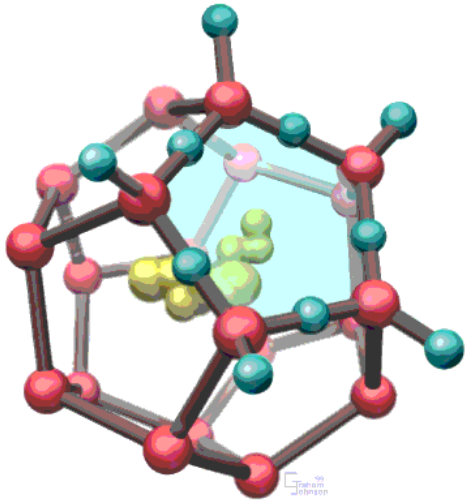
Blue: Hydrogen

Yellow: Non polar molecules

Decrease in entropy

Low solubility

Formation of aggregates



Amphiphatic interactions: both hydrophilic and hydrophobic interactions

Phospholipids tails in biological membranes

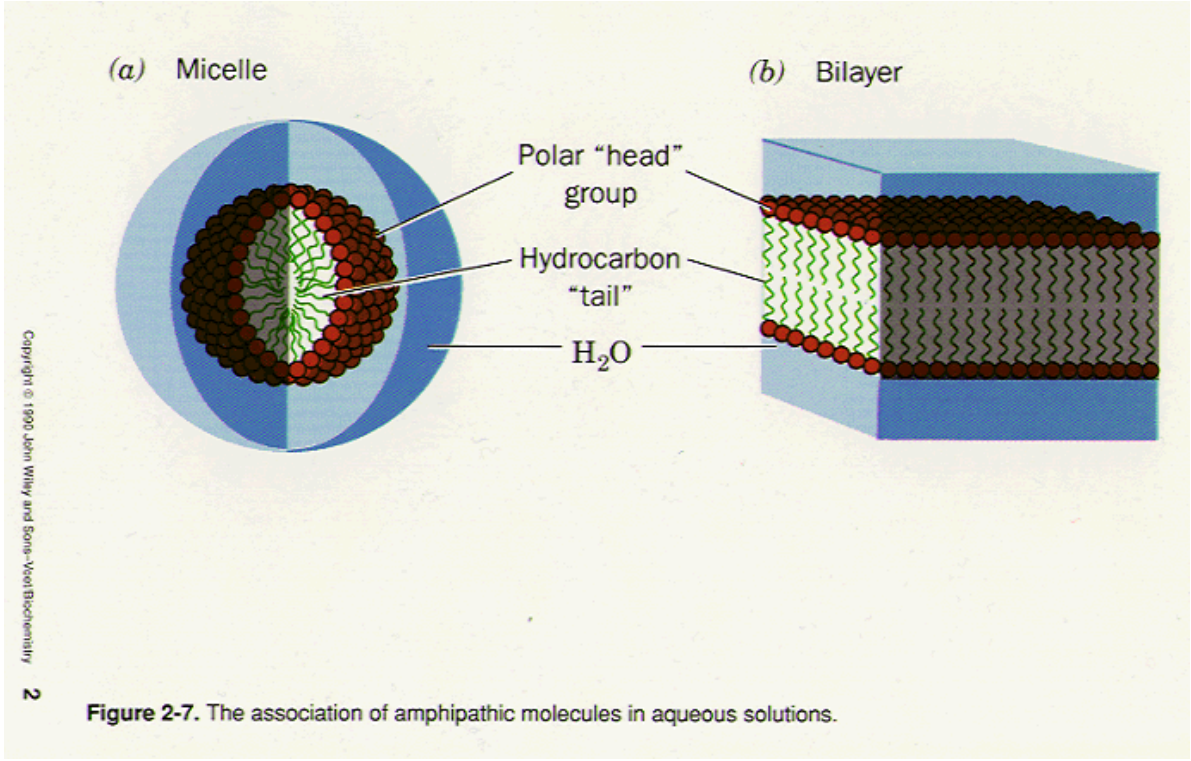


Figure 2-7. The association of amphipathic molecules in aqueous solutions.

Dissolved in water:

Monolayer

Micelles (single layer of molecules)

Bilayer vesicles (Van der Waals interaction between the hydrocarbon tails)

Compound	Molecular Weight	Melting Point (°C)	Boiling Point (°C)	Heat of Vaporization (kJ/mol)
CH_4	16.04	−182	−164	8.16
NH_3	17.03	−78	−33	23.26
H_2O	18.02	0	+100	40.71
H_2S	34.08	−86	−61	18.66

High Viscosity and suface tension

High boiling point

Decrease of density when solid state

Permanent Dipolar character hydrogen bonds:

- Oxygen as perfect acceptor
- OH as perfect donor

Conclusions

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

Hydrogen bonds:

- –Responsible for the secondary, tertiary and quaternary structure of proteins and nucleic acids
- –Determine the conformation and folding ways of macromolecules
- –In biological compounds only N and O as hydrogen bond donors
- –Fundamental importance in biological processes (water)
- –Highly directional: donor H tends to point directly to the acceptor e^- pair
- –Greater energy than most other non covalent interactions

Water is the Universal environment the life has selected:

- –permanent dipole + hydrogen bonds capacity

2. Chemical and physical Background
2.2.1 Non covalent Interactions

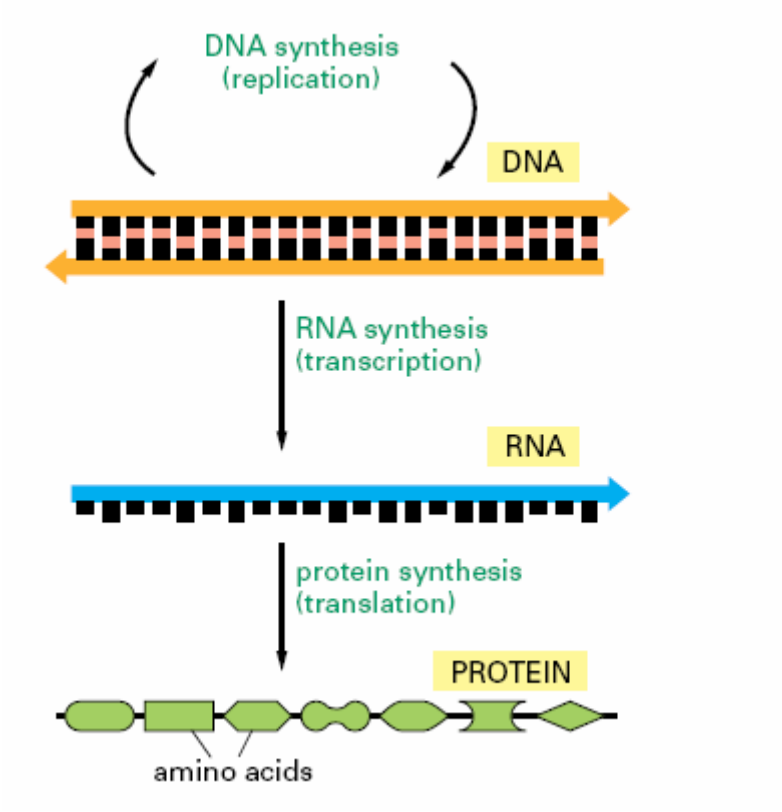
Interactions

- Electrostatic
  - Ionic, dipole-dipole and hydrophilic-hydrophobic interactions and hydrogen bonds
  - Strength charge difference: Ionic 1000 (integer charge)> hydrogen 100 (partial charge) > dipole-dipole 10 (partial charge)

- Electrodynamic : Van der Waals and London dispersion forces 1

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

2.3 **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 α helices or β strands

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

Quaternary structure: Association of more than one polypeptide folded chain

2.3.1 **Amino acids: classification and chemical-physical properties**

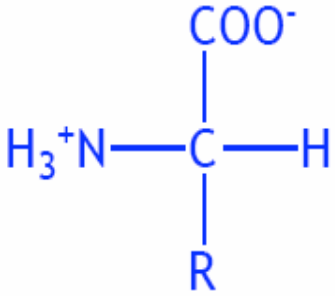
General formula NH2CαHRCOOH: differing on R group attached

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



2.3.1 **Amino acids: classification and chemical-physical properties**

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)

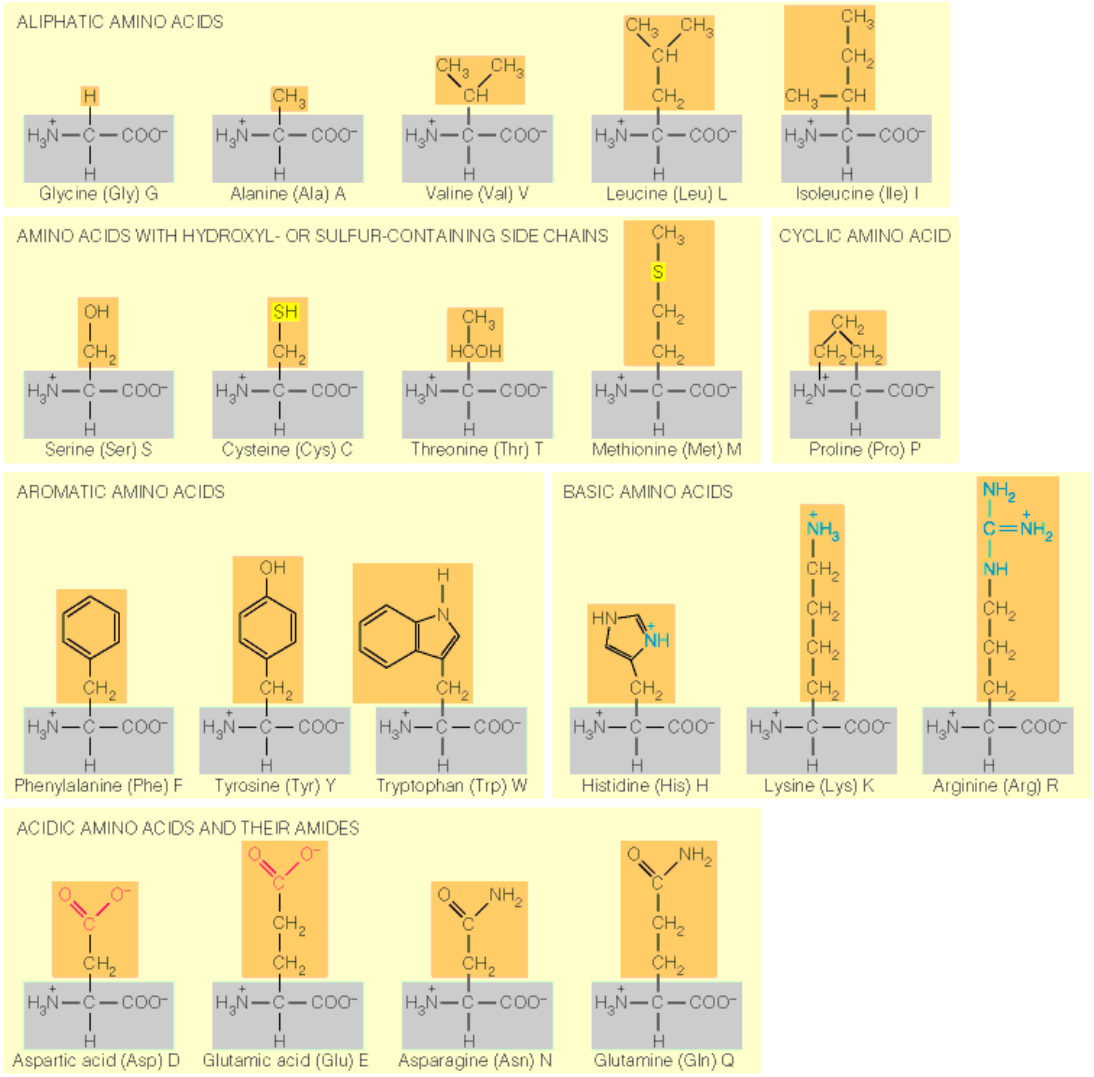
Hydrophilic

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

2.3.1 **Amino acids: classification and chemical-physical properties**



- Hydrophobic groups $(-CH)_n$
- Hydrophilic groups (-OH)
- Sulfur-containing: disulfide bonds (-SH)
- Cyclic side chain in Proline
- Heterocyclic group: Alanine + phenyl + OH + indol
- Guanidinium group
- Imidazol group
- Acid and their non charged residues
- Free carboxy groups and amide terminal groups

2.3.1 **Amino acids: classification and chemical-physical properties**

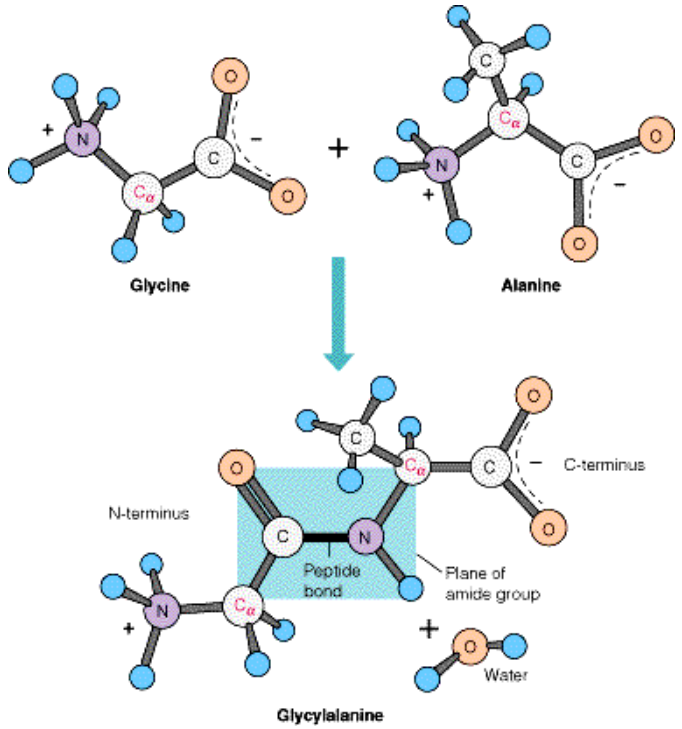
	Gly	Ala	Val	Leu	Ile	Met	Cys	Ser	Thr	Asn	Gln	Asp	Glu	Lys	Arg	His	Phe	Tyr	Trp	Pro
Gly																				
Ala	58																			
Val	10	37																		
Leu	2	10	30																	
Ile		7	66	25																
Met	1	3	8	21	6															
Cys	1	3	3		2															
Ser	45	77	4	3	2	2	12													
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

2.3.1.1 **Peptide bond**

Carboxyl acid –COOH + amino -NH_2 + water



Zwitterion: Dipolar form at pH=7

Whole charge is neutral

Amide bond

Covalent nature

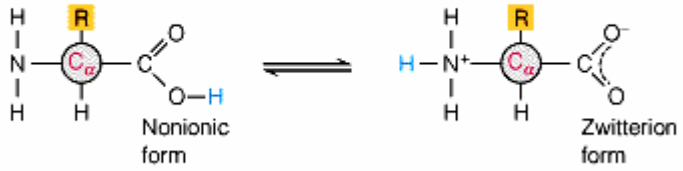
Four atoms linked to the Cα

- Hydrogen atom
- R side chain
- -NH2
- -COOH

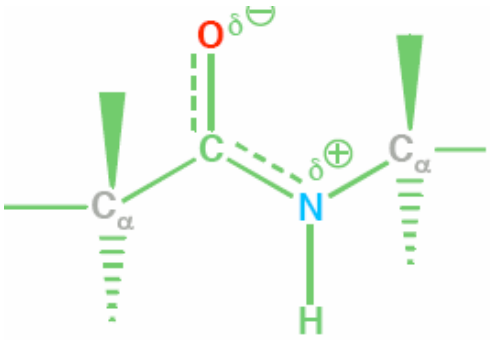
Protein backbone as blocks of repetitive N-Cα-N

Free amino group: N-terminus

Free carboxyl group: C-terminus



2.3.1.1 **Peptide bond**



Consequences

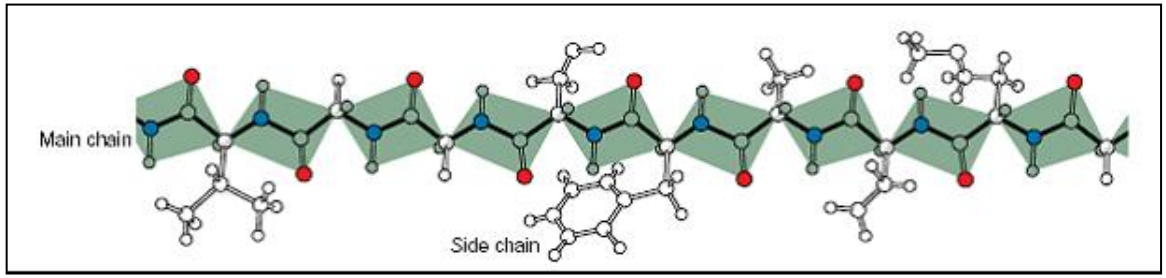
Resonance: partial double bond character (delocalized pair of e-)

Increasing polarity m = qx

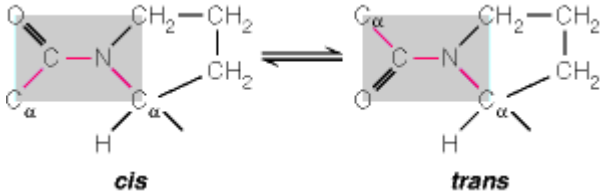
Coplanarity and no free rotation for the axis O=C=N

Free rotation for N-Cα and Cα-C

Stability and flexibility of polypeptide chains in water



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



Trans-configuration is the most likely except for proline

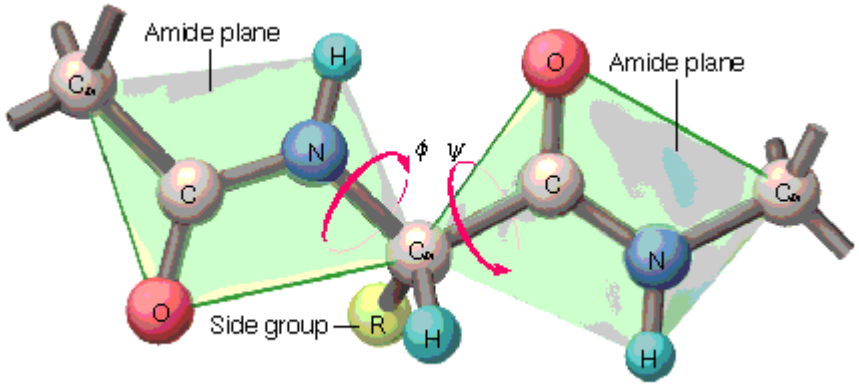
2.3.1.2 **Psi and Phi angles**

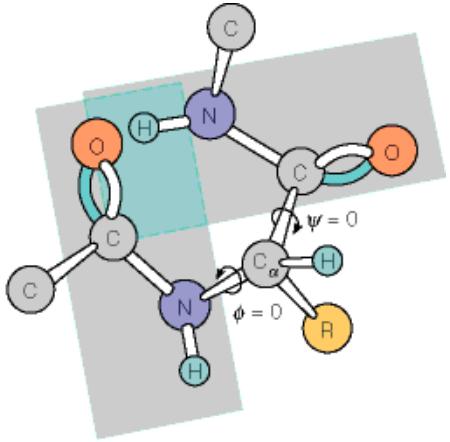
Rotation allowed **only** for the torsion angles phi and psi

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)*

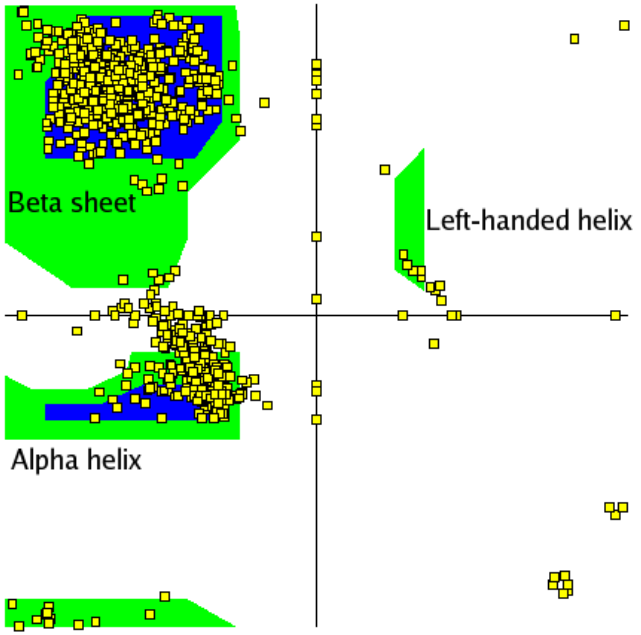




The positive rotation is clockwise

2.3.1.3 **Ramachandran plot**

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



How secondary structure elements are arranged

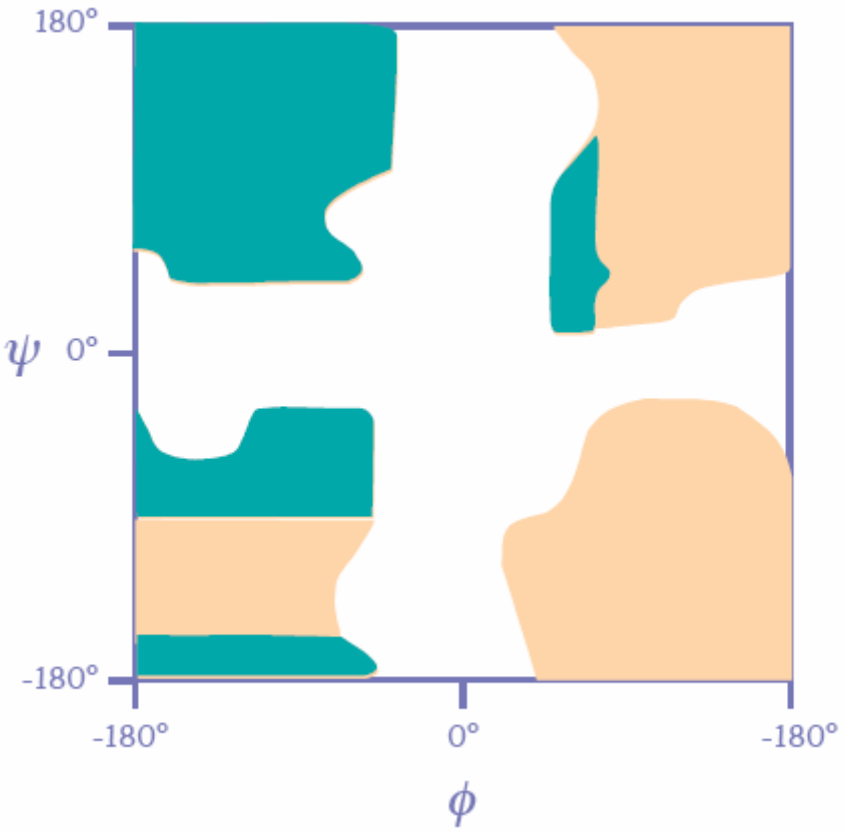
Possible conformation based on individual amino acid dihedral values in a polypeptide

Positive rotation following clockwise (from left to right)

Negative rotation opposite direction

Diagnosis method: values allowed for a experimentally solved protein structure

2.3.1.3 **Ramachandran plot**



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

2.3.1.3 **Ramachandran plot**

Conclusions

- Every backbone conformation of any particular residue in any protein could be described by specifying those **two angles**
- In similar secondary structure 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

2.3.2 **Interactions and folding**
2.3.2.1 Bonds

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)

2.3.2 **Interactions and folding**
2.3.2.1 Bonds

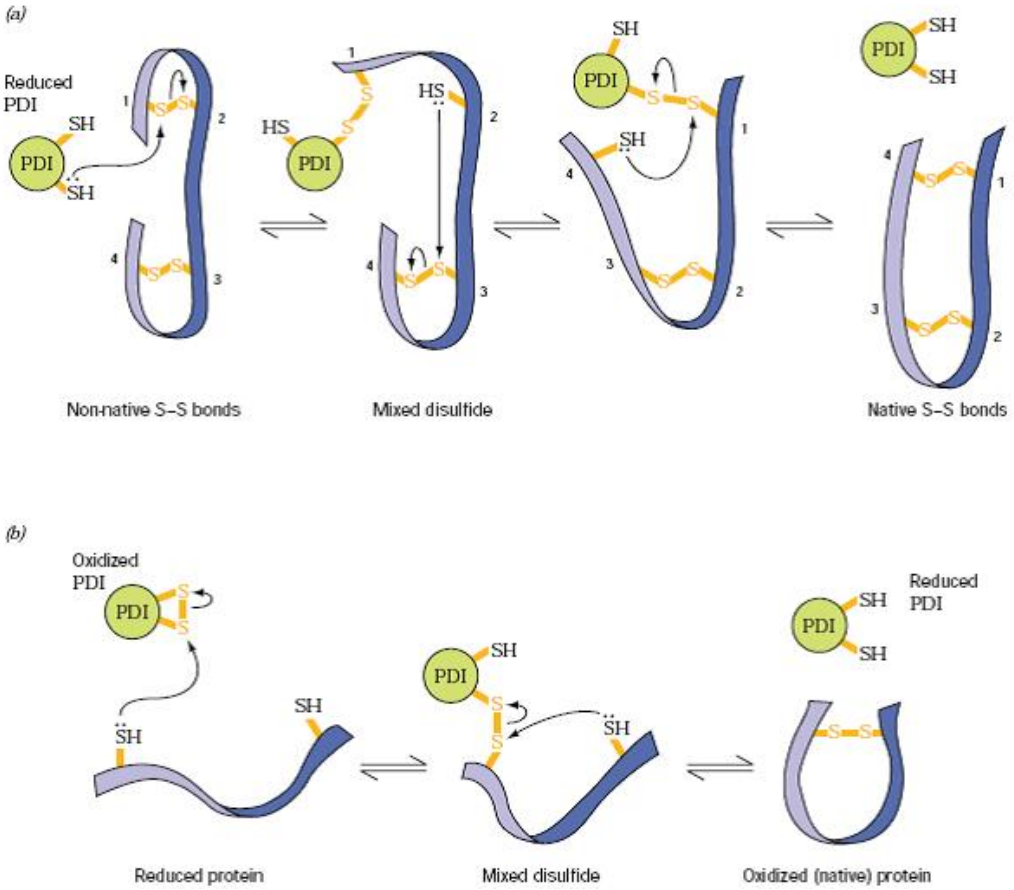
INTERACTION	EXAMPLE	DISTANCE DEPENDENCE	TYPICAL DISTANCE(Å)	FREE ENERGY (kJ/mol) (bond dissociation enthalpies for the covalent bonds)
Covalent bond	Cα-Cα	-	1.5	356
Disulfide Bond	-Cys-S-S-cys	-	2.2	167
Hydrogen bond	-NH—O=C-	Donor(N) and acceptor(O)	3.0	2-6 in water and 12.5-21 if either donor and acceptor 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)



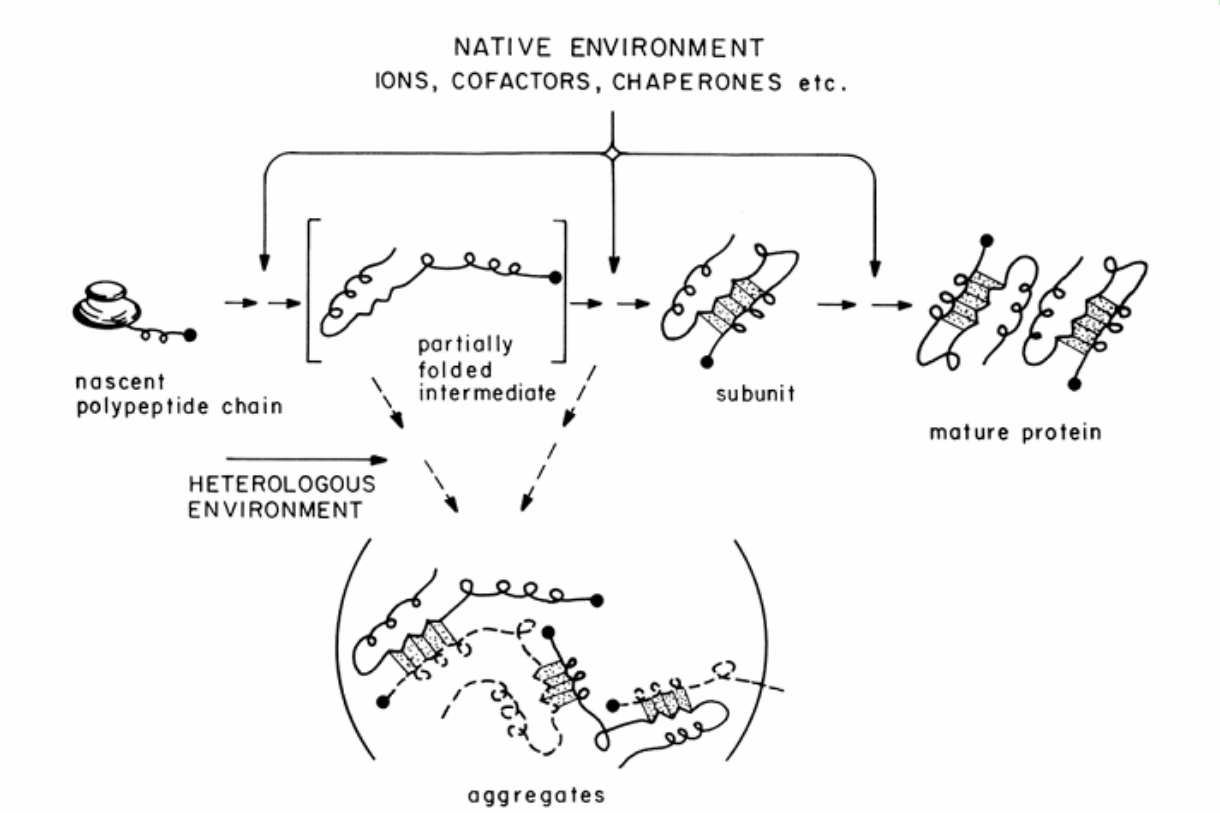
Protein Disulfide Isomerase.

Fully reduced protein is unfolded and it does not fold until the cysteine residues are oxidized to disulfide bridges

$\Delta G = \Delta H - T\Delta S$

Heat released
System disorder
Combination of both sign and value

- Net loss of free energy and $\Delta G<0$ spontaneous reaction of folding
  - ΔH decrease bonds formation
  - ΔS increase disorder
- Driven force:
  - Hydrogen bonds formation: small contribution to ΔG
  - Hydrophobic effect : gain in ΔS system (protein and environment)
- SSEs formation as consequences of the burial of hydrophobic side chains early in the folding process
- Thermodynamic compromise : ΔH equilibrated with decrease ΔS
- Broken equilibrium – Denaturalization (T^a or SDS)



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

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

- Movie: Active site 1 (Lactato Dehydrogenase)
  - Arg -171 and His-195
- Movie: Active-site 2
  - His-57, Ser 195 and Asp 102

Part I: Structural Bioinformatics

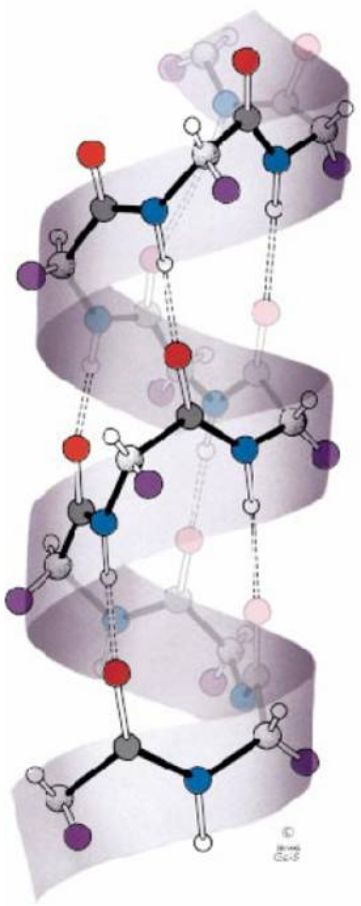
2.3.3 Secondary Structure Elements
- 2.3.3.1 Formation
- 2.3.3.2 Main Types
  - Alpha Helix
  - Beta sheets
  - Turns and Loops
  - TIM Barrels
  - Coiled coil
- 2.3.3.3 Motifs and Domains
  - Homeodomains
  - Leucine Zipper
  - Zinc Finger
  - Transmembrane helices

2.4 Tertiary Structure

2.5 Major methods for structure determination
- 2.5.1 X-ray crystallography
- 2.5.2 NMR

2.7 First approximation
- 2.7.1 PDB- function
- 2.7.2 SCOP-Classes

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

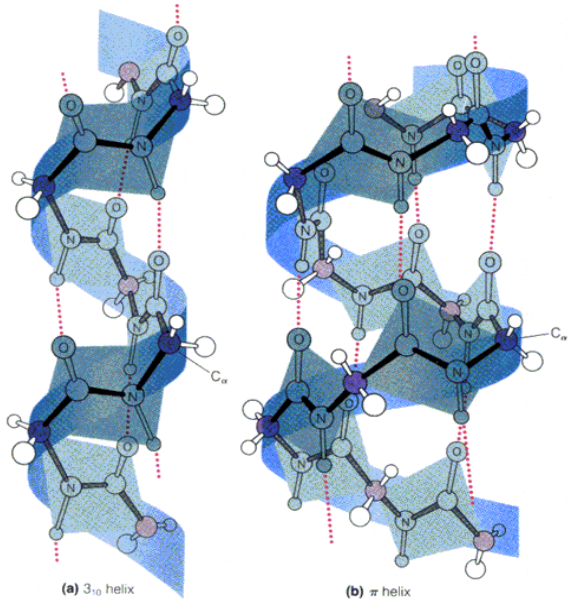
Interactions do not involve side chains

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)

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

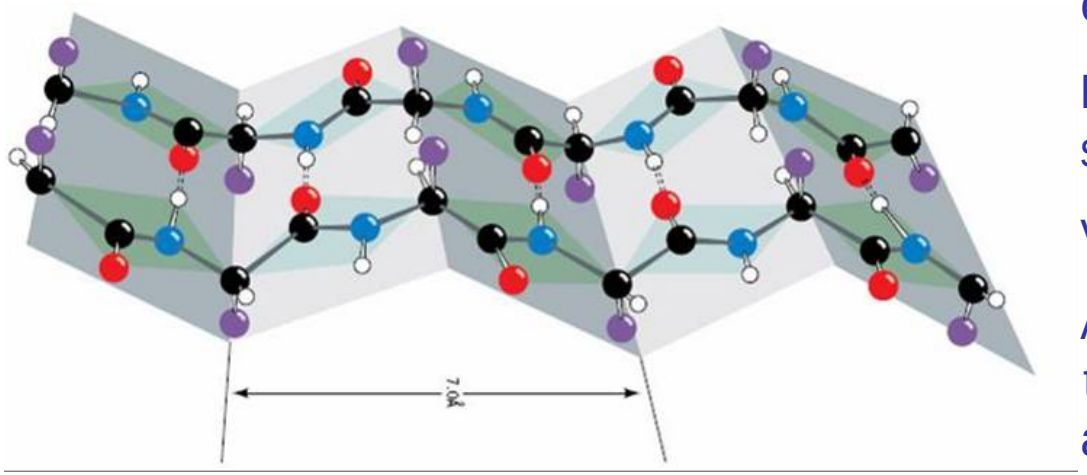


Low stability

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

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

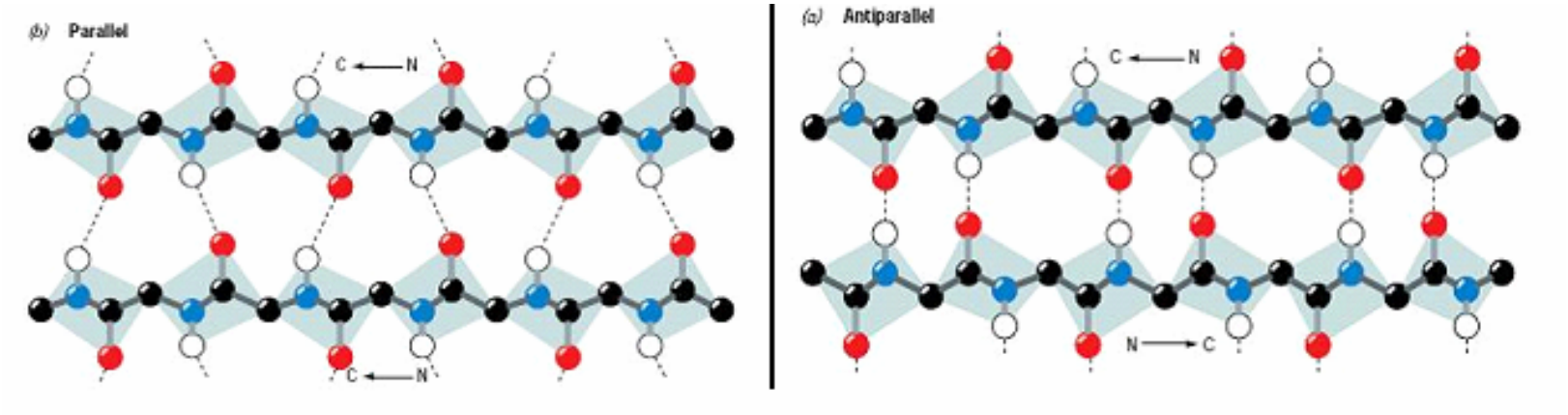
Distance between two consecutive residues is 3.3 Å

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

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

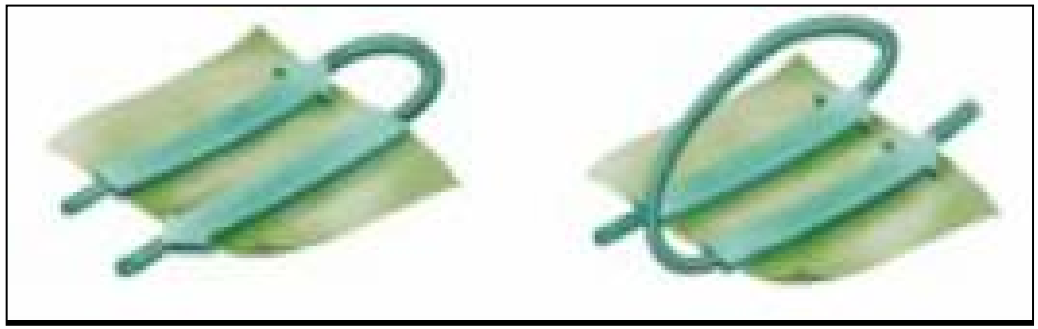
More stable: Exposed

Connected via turns reversing direction

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

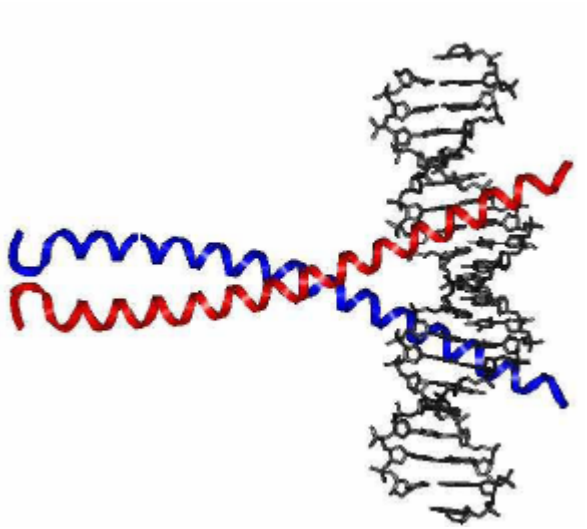


Limit the size of the molecule and maintain the compact state

Hydrogen bond with water molecules avoiding the four residues to interact

Placed in the surface of folding proteins

Coiled coil



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

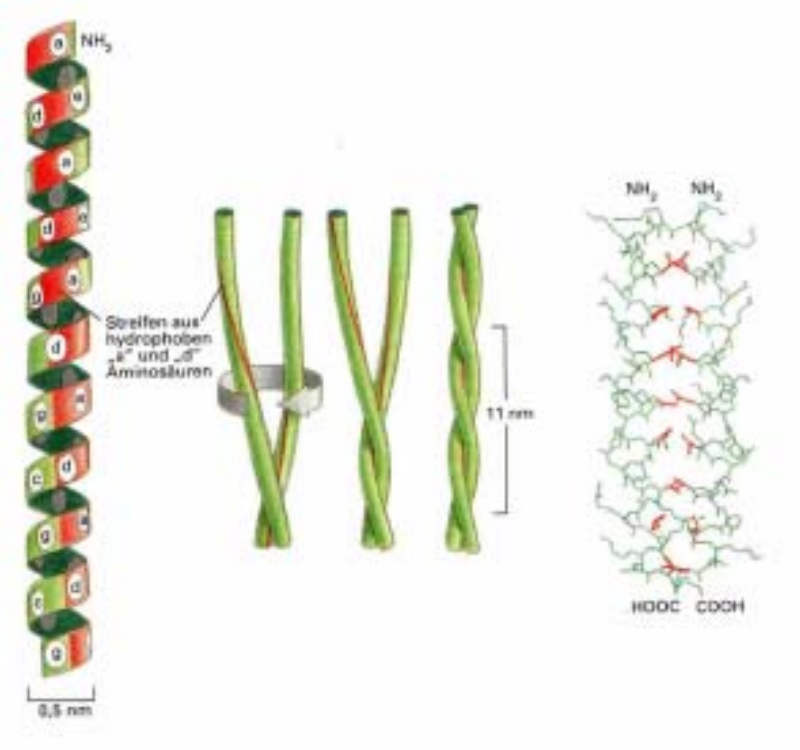
Associated in parallel or antiparallel orientation

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

Amphiphatic property

Their hydrophobic sides snuggle tightly together in the center

Stable hydrophobic core



"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

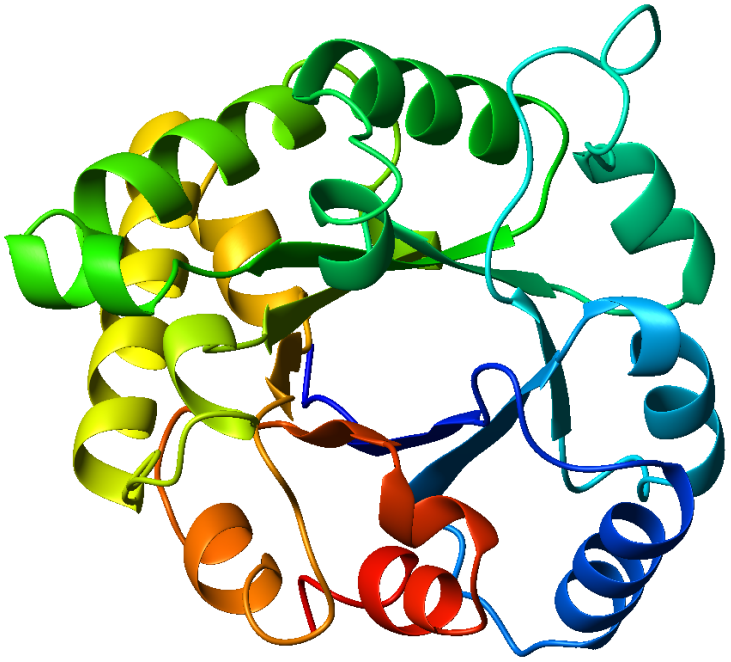
Positions **a** and **d** are hydrophobic , **e** and **g** are charged and **b, c, f** are hydrophilic

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

Transcription factor in yeast GCN4

Avian Flu Virus

TIM barrels



Triosephosphateisomerase

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

Catalytical function of the protein

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

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

- **Motifs:**
  - Particular amino acid sequence i.e. Zinc Finger (biochemical function)
  - Contiguous set of SSEs

- **Domains:**
  - Alpha
  - Beta
  - Alpha/beta
  - Alpha +beta
  - Cross linked domains

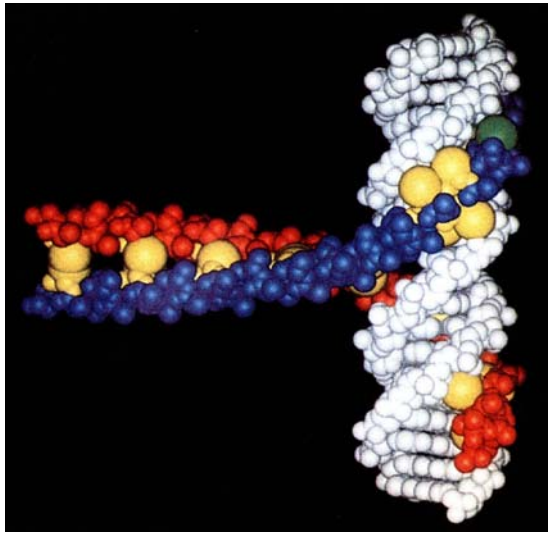
Homeodomains are found in many transcription factors binding to DNA (TATA box)

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

Zinc finger

Structural motifs used by a large class of DNA-binding proteins

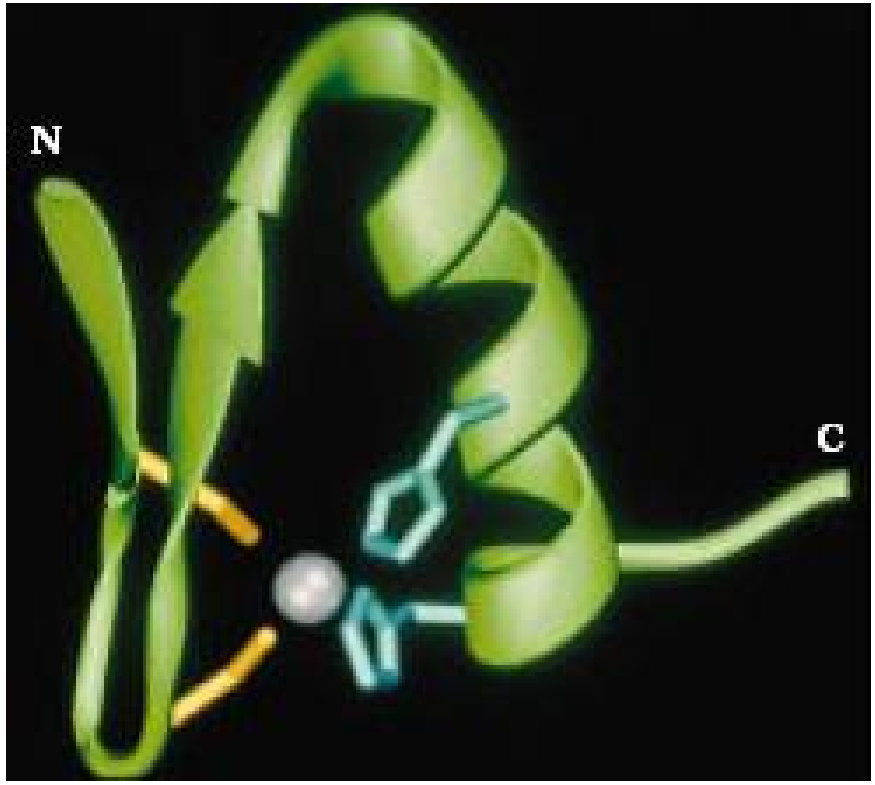
Coordinated zinc atoms as crucial structural elements

Single zinc finger domain is only large enough to bind a few bases of DNA (found in tandem)

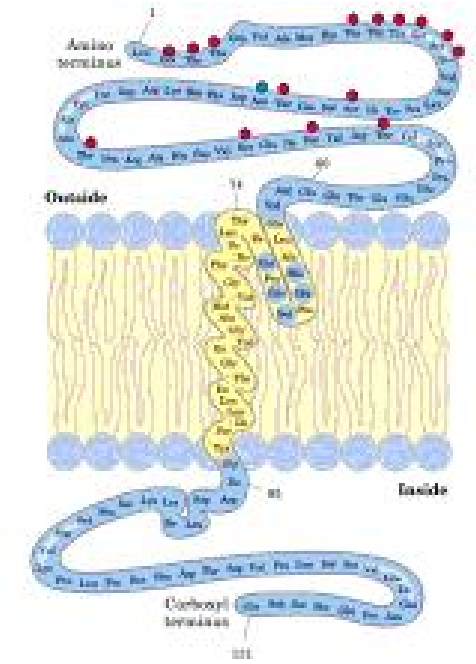
Helical region of each zinc finger rests in the major groove of the DNA helix

Modulation of DNA and gene expression

HIV and potential drug target



Transmembrane elements



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

Glycophorin C protein

Single transmembrane domain

Critical for maintaining the shape and stability of erythrocyte

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

Empirical rules to follow

Any amino acid can be found in any type of secondary structure

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 α helix and β 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 α helices

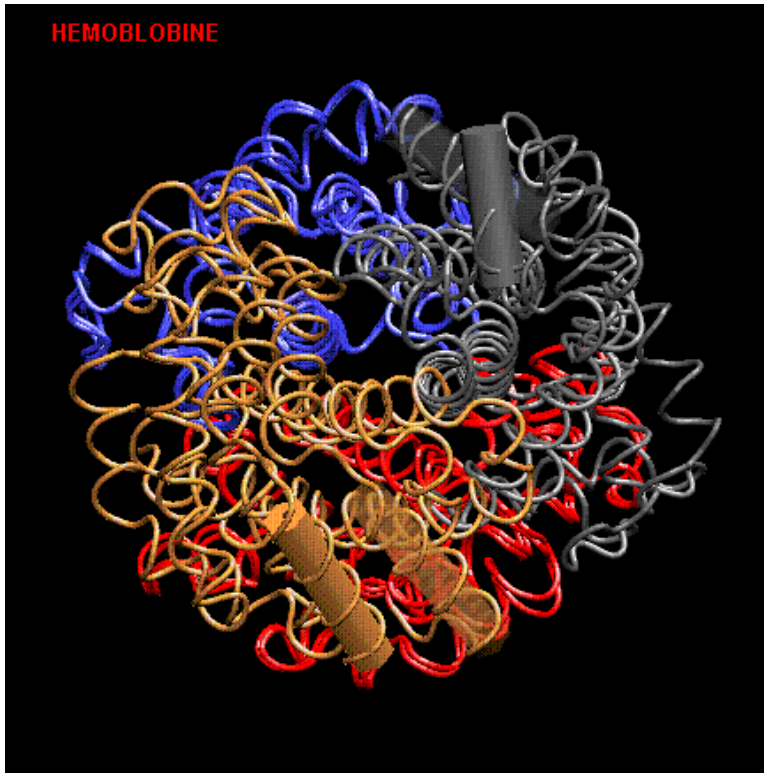
2.3.3 Secondary Structure Elements

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

2.4 Tertiary Structure

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



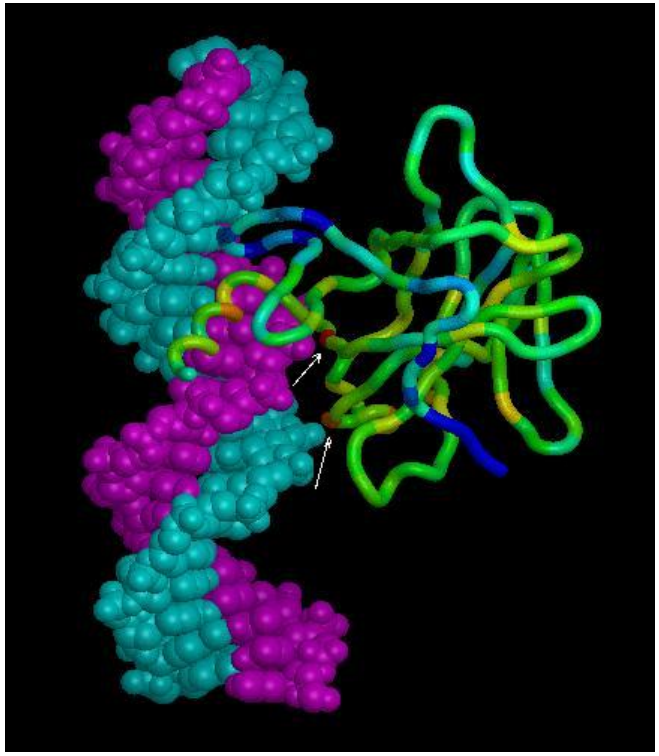
Stabilized by

- 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

2.6 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 βstrands as arrows pointing from the N- to the C-terminus and αhelices 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

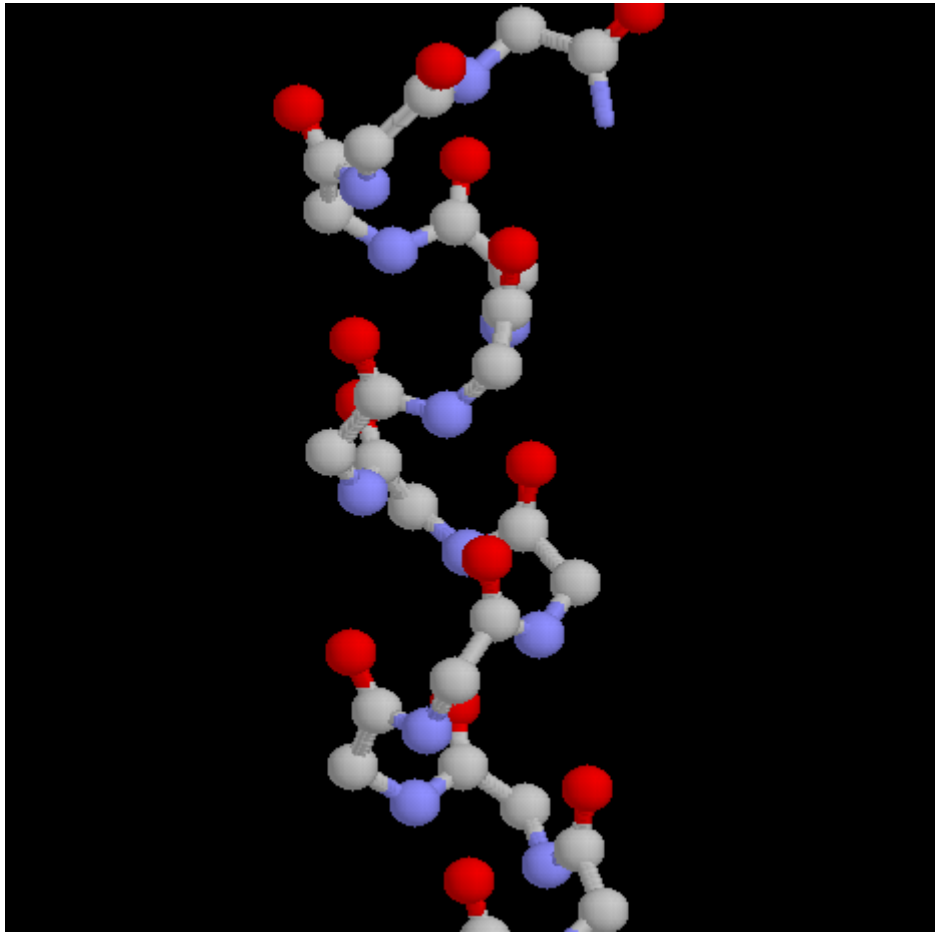
Chime a browser plug-in that renders 2D and 3D molecules directly within a Web page

Pymol as a *molecular graphics system Python interpreter*

http://av.bmbq.uma.es/av_biomo/
http://www.mdl.com/products/framework/chime/index.jsp

http://www.umass.edu/microbio/rasmol/index2.htm
http://pymol.sourceforge.net/

αHelix **Ball** and **Stick View**



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

Lysozyme

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

2.7 First approximation

2.7.1 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):

2.7 First approximation

2.7.2 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

CHAPTER 3. STRUCTURAL COMPARISON AND ALIGNMENT

3.1 Introduction

Goal:

Determination of equivalences between amino acid residues by taking into account 3D structures

Relationships between primary protein sequence, 3D structure and biological function

Four steps

- **Structure alignment**: find equivalences of amino acid residues based on known 3D structures.
- **Structure comparison**: once shared similarities are known the structures are compared.
- **Structure superposition**: find the optimal overlap of both proteins.
- **Structure classification**: assign the protein to a certain class

Structural alignments provide information that is unavailable through current sequence alignment methods

3.1 Introduction

Motivation:

From Genome sequencing to amino acids/nucleotides primary structure.
From amino acids/nucleotides primary structure to 3D Structure Prediction.

2008 In PDB data base 49192 Structures structures
Feb 24, 2009 _ 56066 Structures
2008 SWISS PROT 356 194 entries sequence
10-Feb-2009 UniProtKB/Swiss-Prot Release 56.8 of : 410 518 entries

Ratio of 1 structure to 7 sequences

3.1 Introduction

- More than 3000 structures are stored in the structural protein data bank (in 2004, 5360 experimentally determined structures were deposited)
- Protein structures are more highly conserved than sequences : Evolutionary changes like insertions and deletions take place mainly in loop regions.
  No alterations in the final fold and limiting the number of possible folds
- Similar structures may be formed by alternative folding of the amino acids' Cα backbone
  Matched regions separated by unmatched segments
- Partial local similarities do not automatically transfer to similarities in structure
  Same nucleus BUT different end
- 30% Sequence identity adopt the same folds: homologous folds
- 5% Similarity can result in the same fold: analogous folds

3.2 Main Methods

3.2.1 Basic algorithms review

- Structures can be compared, assuming they adopt the same fold
- **Structural comparison and alignment as NP-hard problems**: non-deterministic polynomial time problems solved by heuristic approaches
  - Possible solution as the best analytical answer but NO biological mean
- Find the most suitable method to solve the optimization of the alignment and to reduce the computing time consuming problem: 5 of 10 structures inferred without special algorithms
  - New proteins weekly released in PDB previous all-against-all comparison
  - Known sequence-structure relationships are used
  - PDB structures are grouped (only a subset is compared)

Steps for algorithm optimization

A. **Structure comparison and alignment**

i. Representation of the pair of proteins 1 and 2, domains or fragments to be compared and aligned.

ii. Compare 1 and 2

iii. Optimize the alignment between 1 and 2

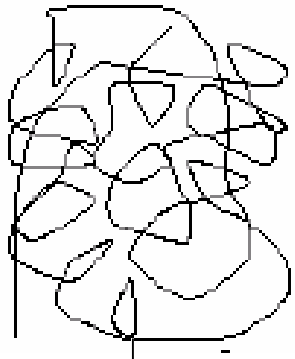
iv. Statistically significant measurement of the alignment against a random set of structures

B. **Multiple structure alignment**

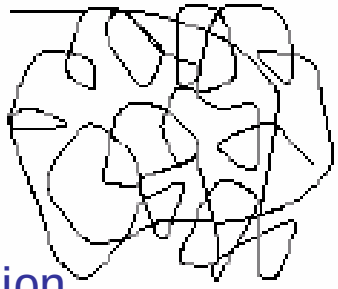
i. Starting from the initial alignment found in Aiii., the next step is running a search within a constraining sequence window to find the optimal alignment against all structures using profiles; HMMs or Monte Carlo approaches.

How to compare structures

Structure 1



Structure 2



Feature extraction

1. Structure description 1 — Structure description 2
2. Comparison algorithm
3. Scores

Statistical Significance (PDB)

Similarity-Classification

DP in Structural Bioinformatics

Solutions to NP-hard problems in a computationally cheaper way

Aligning Sequences : A row of amino acids in one sequence matches a row of identical or substituted positions in the second sequence; insertions or deletions as gaps

Aligning Structures: A scoring matrix is built to compare the positions of the atoms in both 3D structures

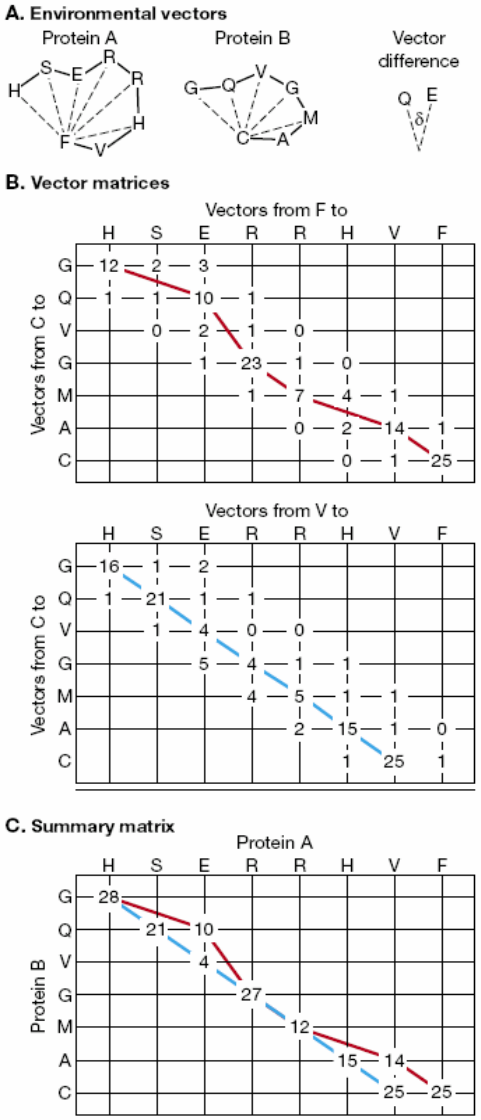
i. Scoring matrix gives scores of how well any of the 20 amino acids fits to a single position in the structure. Calculation of an optimal alignment

ii. Positions of SSEs within a domain (types, positions and numbers are similar)

iii. Distances between the Cα (NH- Cα- Cβ) and Cβ (Cα-CβO=NH) atoms within these domains, and later within the whole structure

iv. Determination of the degree of superimposition

3.2.1 Basic algorithms review
3.2.1.1 Dynamic programming



Two steps

1. Atoms or molecules as vectors:

A coded value is given describing the local environment of each amino acid

- Interatomic distances
- Bond angles
- R groups

Cartesian coordinates are assigned to each (X, Y, Z)

Direction of the bond angles is included

2. The alignment of 2D structures:

Determine of the interatomic distances between each amino acid in the polypeptide chain

"The better the arrangement, joining and 2D alignments are, the more significant and convincing is the result"

3.2.1 Basic algorithms review
3.2.1.2 Distance matrix

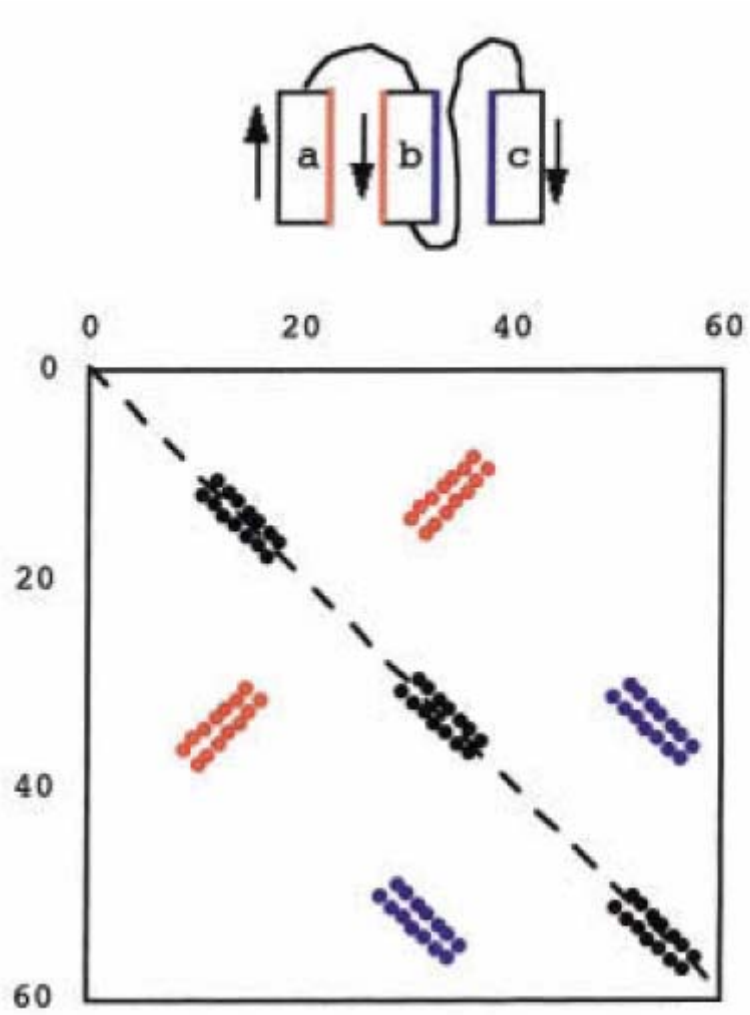
No alignments help is needed

Each position in the 2D matrix represents the distance between corresponding Cα atoms in the 3D structure

i. Distances between Cα atoms along the polypeptide chain and between Cα atoms within the protein structure are compared

ii. Similar groups of 2D structural elements are superimposed (sum distance minimization in the aligned Cα atoms resulting in a common core)

"*The smallest the distance, the most closely packed atoms within SSEs and regions of the 3D structures* "



Bases of Distance method

Degree to which all of the matched elements can be superimposed

Protein A ------- helices a and b interacting
Protein B ------- helices a' and b'

Helices superimposition-set of Cα

i^A and i^B in helix a and a'
j^A and j^B in helix b and b'

For matching pairs

dijA = distance between i^A and j^A

dijB = distance between i^B and j^B

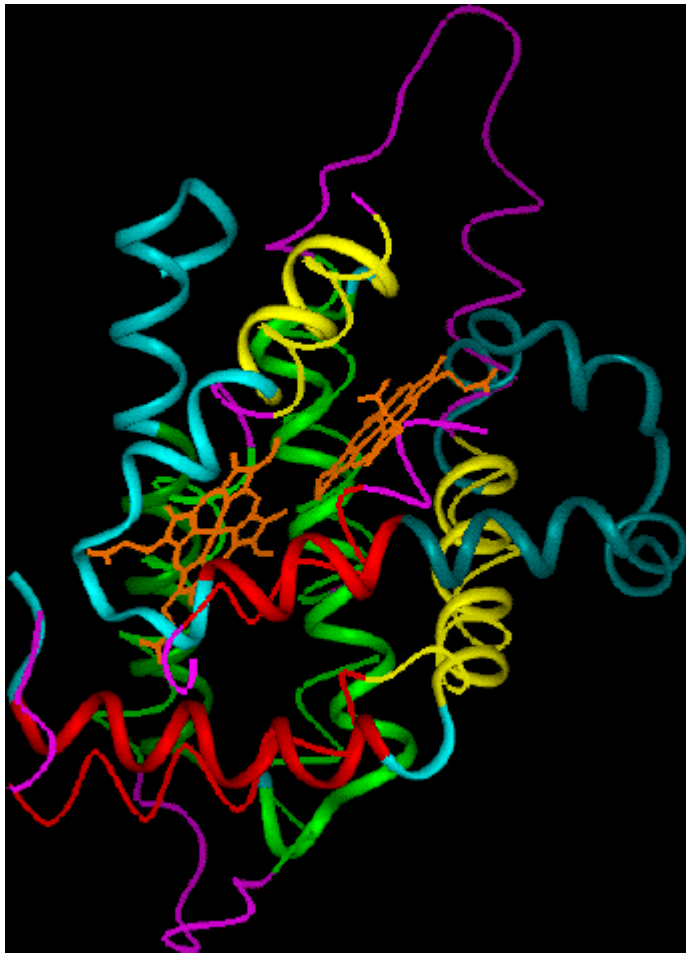
SS = |dijA – dijB| / dij*

dij* = average of dijA and dijB

3.2.2 SARF2, VAST, COMPARER
3.2.2.1 SARF2

Components of structure elements to be compared:

- Local geometry
- Side chain contacts
- Geometric hashing
- Distance matrix (Dali, 1993)
- Properties as SSs, hydrophobic clusters
- Secondary structure elements
- Distances of inter and intra aligned fragment pairs



For both SARF2 and VAST the predictions are based on vector comparisons

SSEs converted into vectors based on

- Position
- Direction
- Length

Used to compare new structures to the existing DB or to view structural similarities already in the DB

http://123d.ncifcrf.gov/sarf2.html

Spatial Arrangement of Backbone Fragments (Nickolai N Alexandrov, 1998)

Based: comparison of Cα of each residue in the SSEs of each protein

Goal: to find those SSEs which can form similar spatial arrangements but have different topological connections

How: SSEs detected through comparison with common templates for a-helices and b-strands, then larger assemblies of SSEs are constructed from the compatible pairs found

First step: pairs of SSEs are matched up

- Shortest distance between their axes
- Closest point on the axes
- Minimum and maximum distances from each SSE

3.2.2 SARF2, VAST, COMPARER
3.2.2.1 SARF2

Second Step: Largest ensembles are formed

- Graph theory and maximum clique problem approximation

Third Step: Extension of the alignment

- Additional residues included

Similarity Score: Calculated as a function of rmsd and the number of matched Cα atoms.

The significance of the comparison is considered contrasting this score with the one built up once a protein is compared with a non redundant set of structures

Blue ribbon shown as repressor 434 and recovering as red line. Yellow fragments can be superimposed with rmsd = 2.61

52 Cα matched found

No evolutionary relationship but structural stability is apparent

3.2.2 SARF2, VAST, COMPARER
3.2.2.2 VAST

Vector Alignment Search Tool (Gibrat et al,1996)

Based: SSEs-pair alignment

How: Structures as a set of vectors of secondary structural elements whose direction, type and connectivity infer the topology of the structure.

Once the alignment is achieved, it uses Gibbs sampling algorithm to examine alternative alignments

The statistical theory similar to BLAST

BLAST: probability to get the same score when aligning a test sequence against a DB sequence would be found by comparing random sequences

VAST: SS is the likelihood that the score would be the result of a random alignment of unrelated structures

Score: number of superimposed SSEs

3.2.2 SARF2, VAST, COMPARER
3.2.2.2 VAST

SS = N1 x N2

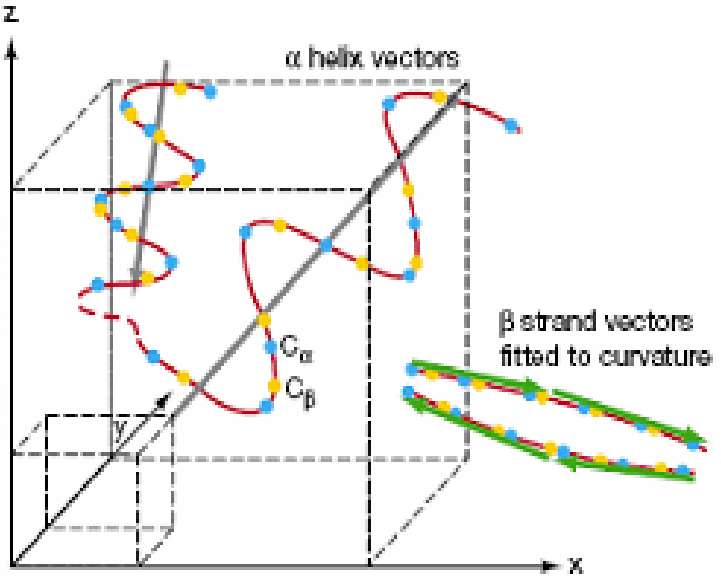
N1 = probability of picking up elements randomly and get the same score

N2 = number of alternative element pair combinations

Optimal alignment: the one with highest relation to the background distribution of Cα in the superimposed amino acid residues

Elements in two structures are similarly arranged so the expected similarity on their corresponding 3D structures

SSEs found + clustering + larger alignment group + selection



3D structures of proteins are predicted to be similar if, once the representations of their vectors were compared, the type and arrangement are alike within a rational range

3.2.2 SARF2, VAST, COMPARER
3.2.2.3 COMPARER

Based: on the sequence alignment algorithm of Needleman and Wunsch and by using equivalences between protein structures to define general topologies

How: both comparison of properties and of relationships through simulated annealing and dynamic programming

Properties: dynamic programming algorithm

Residues like identity and local conformation

Segments like secondary structure type and orientation relative to the center of gravity

Relationships: combinatorial simulated annealing technique

Relations between residues like hydrogen bonds and hydrophobic clusters.

Relations between segments like distance to one or more closer neighbors and the relative orientation of two or more segments.

For statistical analysis: E (residue equivalences) and A (gap penalties) comparing to the values obtained by using two unrelated proteins and two random sequence relationships

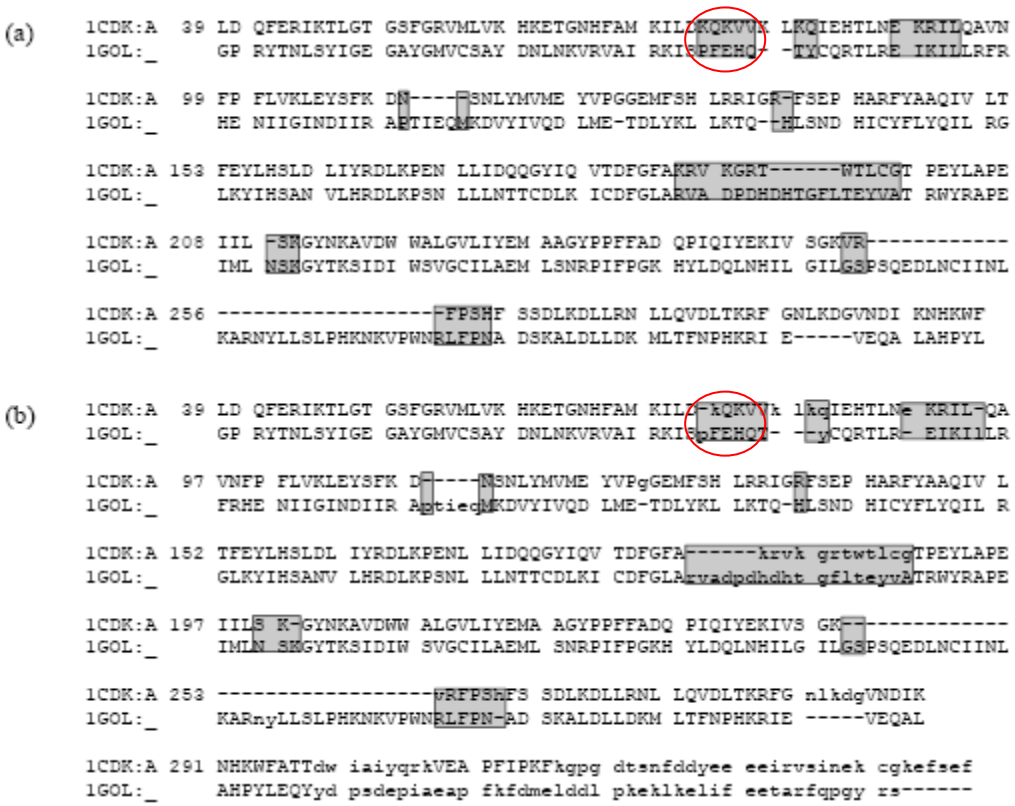
3.2.3 CE, DALI, SSAP
3.2.3.1 CE

Former heuristic methods whose results were contradictory:

Finding the best rmsd is not enough to match biologically meaningful features (the most significant structure alignment)

Different alignments produced by different methods

Close scores got from same methodology but residues in the alignments are far positioned



New algorithm that uses a combinatorial extension of the optimal path; the path is defined by the use of protein properties relevant to structural and functional features

3.2.3 CE, DALI, SSAP
3.2.3.1 CE

Basis

- Target function: heuristics assumes continuity and optimal path existence
- Compare octameric fragments – an aligned fragment pair (AFP)
- Distance matrices: distances between each Ca of each **octamer** fragment combination from both proteins is plotted and represented
- Combinations of AFP "representing" possible continuous alignment path are selected and extended
- Find the optimal path through the AFPs
- Optimize the alignment through dynamic programming
- Measure the statistical significance of the alignment

Assumed rules

- Remove highly homologous chains
- The rmsd between two chains < 2Å
- The length difference between two chains < 10%
- The number of gap positions in alignment between two chains < 20% of aligned residue positions
- At least 2/3 of the residue positions in the represented chain are aligned

3.2.3 CE, DALI, SSAP
3.2.3.1 CE

Alignment algorithm

- Input and output of alignment algorithm

Input: two proteins: $A=\{a_1,\cdots,a_m\}$ $B=\{b_1,\cdots,b_n\}$

Output: An alignment and scores

$$L(A,B)=\{(a_{i_1},b_{j_1}),\cdots,(a_{i_L},b_{j_L})\},$$

$$i_1<i_2<\cdots<i_L,\ j_1<j_2<\cdots<j_L$$

Constraints:

min rmsd:

max L

min Gaps:

$$rmsd=\min_T\sqrt{\frac{\sum_{k=1}^{L}(a_{i_k}-Tb_{j_k})^2}{L}}$$

$$Gaps=\sum_{t=1}^{L-1}\left[(i_{t+1}-i_t-1)+(j_{t+1}-j_t-1)\right]$$

Penalization gaps: Computational speed lost of non topological alignments and insertions of more than 30 residues

3.2.3 CE, DALI, SSAP

3.2.3.1 CE

Two methods for detecting structural homology

1. From ONLY structural information

   Alignment Path

   Distance Measure for Similarity Evaluation

2. From structural information AND adding composite properties

(i) Octamer A and Octamer B satisfy a similarity criterion: AFP

(ii)Three threshold

- 1st detecting AFP
- 2nd detecting the correctness of a next candidate AF relative to the current one
- 3rd threshold evaluating all alignments to find the optimal ones

(iii) Statistical significance

- Numerical table
- Two distributions corresponding to both proteins
- rmsd

Gaps values for the non-redundant

Assuming normality the final z-score is calculated by combining both z-scores

Method 1. From ONLY structural information

Alignment Path

Selection of starting point by the ones leading the longest alignment found
Longest continuous path P of AFPs in a similarity matrix S

Protein A length: n^A
Protein B length: n^B
Similarity matrix size: $(n^A - m)\ (n^B - m)$

AFPs i and i+1 extension if and only if

Condition (1): No Gaps between AFPs *i* and *i*+1

$$P_{i+1}^A = P_i^A + m \qquad P_{i+1}^B = P_i^B + m$$

Condition (2): Gaps inserted in protein A

$$P_{i+1}^A > P_i^A + m \qquad P_{i+1}^B = P_i^B + m$$

Condition (3): Gaps inserted in protein B

$$P_{i+1}^A = P_i^A + m \qquad P_{i+1}^B > P_i^B + m$$

Condition (4): Gaps on protein A ; Condition (5): Gaps on Protein B

$$P^A_{i+1} \leq P^A_i + m + G \qquad P^B_{i+1} \leq P^B_i + m + G$$

Distance Measure for Similarity Evaluation: 3 distances are measured

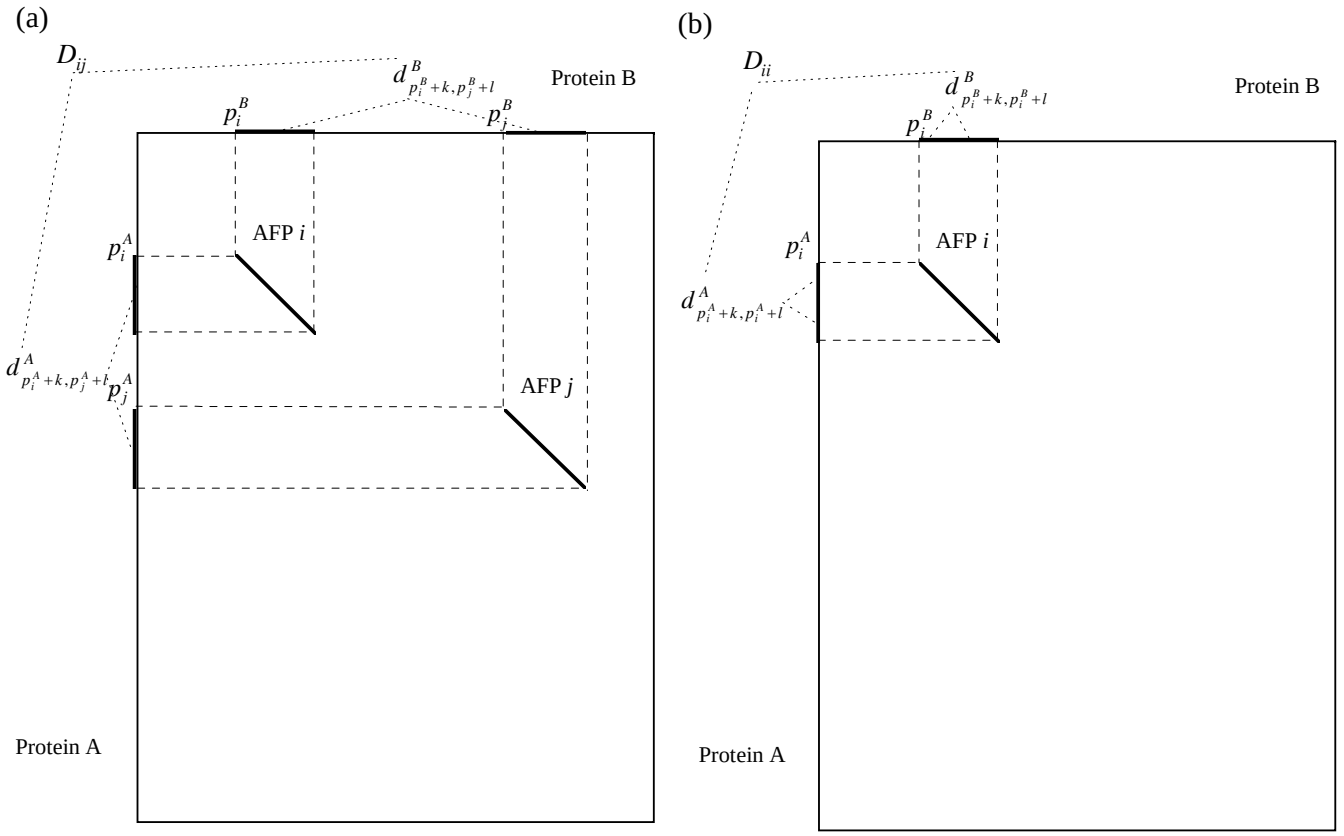
(i) Using an independent set of inter-residue distances: to evaluate combination of two AFPs

$$D_{ij} = \frac{1}{m}\left(\left| d^A_{P^A_i P^A_j} - d^A_{P^B_i P^B_j} \right| + \left| d^A_{P^A_i+m-1,\ P^A_j+m-1} - d^B_{P^B_i+m-1,\ P^B_j+m-1} \right| + \sum_{k=1}^{m-2} \left| d^A_{PA_i+k,\ P^A_j+m-1-k} - d^B_{P^B_i+k,\ P^B_j+m-1-k} \right| \right)$$

(ii) Using a full set of inter-residue distances: to evaluate a single AFP

$$D_{ij} = \frac{1}{m^2}\left(\sum_{k=0}^{m-1} \sum_{l=0}^{m-1} \left| d^A_{P^A_i+k,\ P^A_j+l} - d^B_{P^B_i+k,\ P^B_j+l} \right| \right)$$

Calculation of distance: (a) D_{ij} for alignment represented by two AFPs i and j from the path; (b) D_{ii} for single AFP i from the path.



3.2.3 CE, DALI, SSAP
3.2.3.1 CE

(ii) RMSD obtained from structures optimally superimposed: to select the best alignments and for the optimization of gaps in the final alignment

When adding the next AFP three strategies can be followed

- All possible AFPs which extend the path and satisfy the similarity criteria
- Only the best AFP which extend the path and satisfy the similarity criteria
- Intermediate criteria

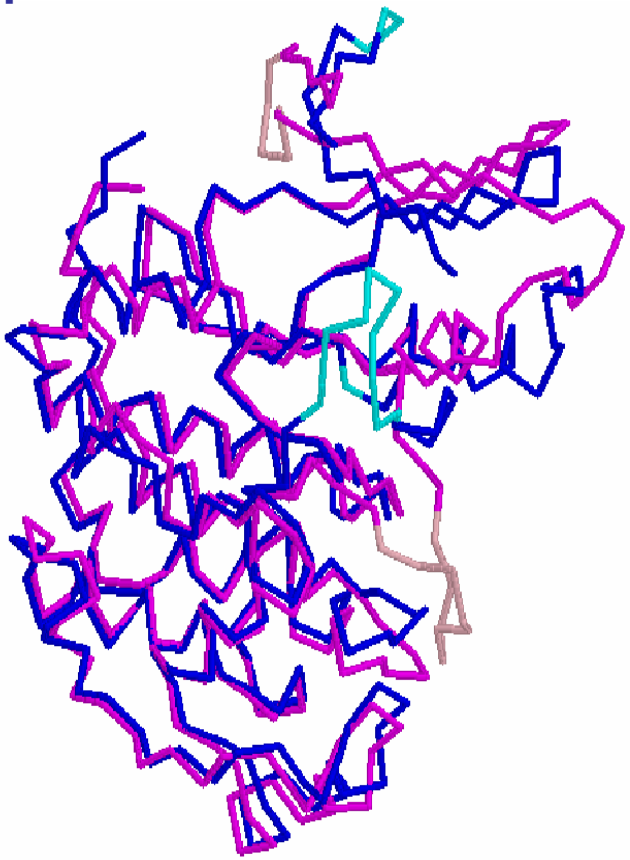
Three heuristic and three conditions to decide

Condition (6): Single AFP $D_{nn} < D_0$

Condition (7): AFP against the path $\frac{1}{n-1} \sum_{i=0}^{n-1} D_{in} < D_1$

Condition (8): Whole path $\frac{1}{i^2} \sum_{i=0}^{n} \sum_{j=0}^{n} D_{ij} < D_1$

Optimization of the Final Path



The 20 best alignments with a Z score above 3.5 are assessed based on RMSD and the best kept
This produces approximately one error in 1000 structures

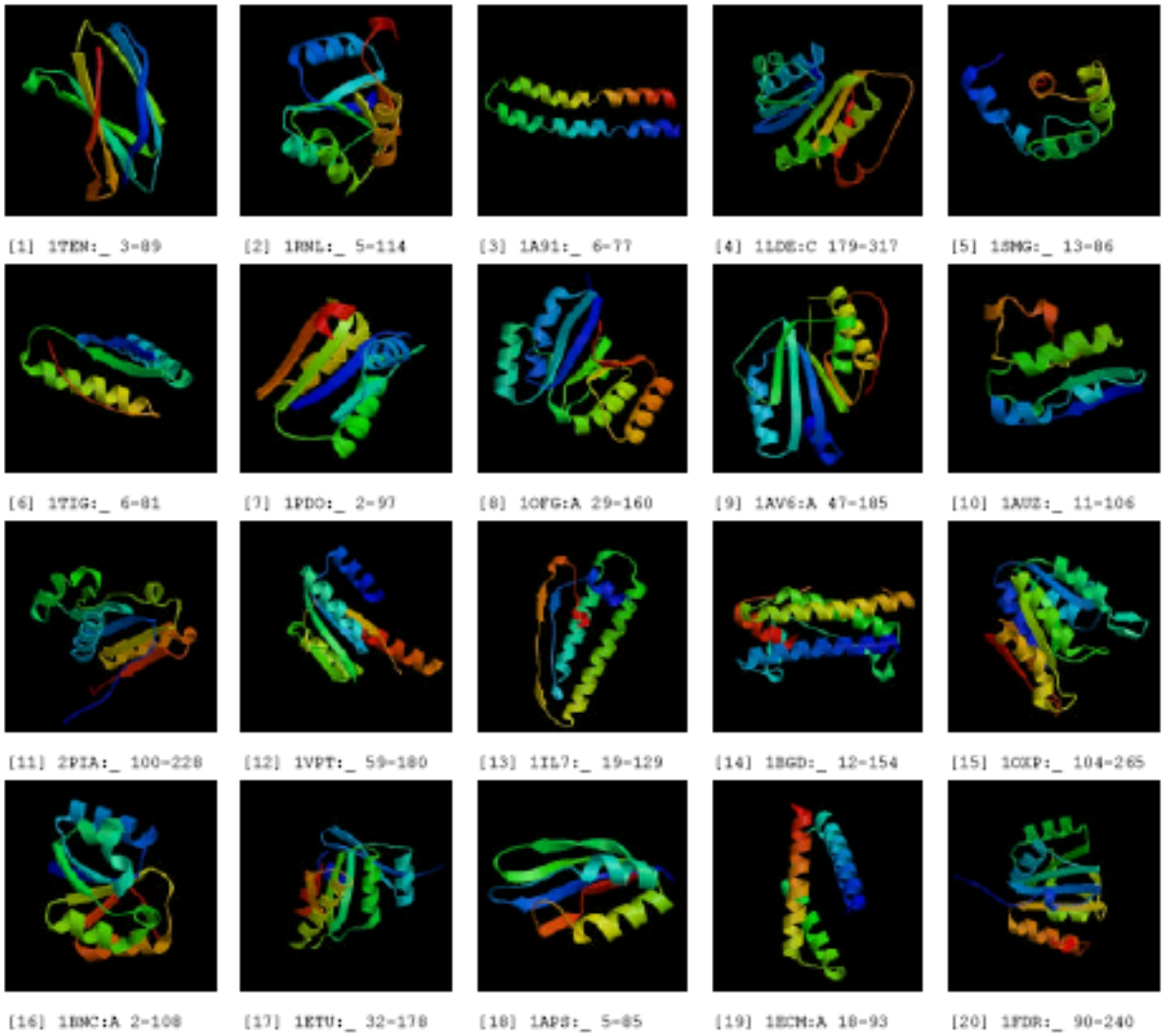
Iterative optimization using dynamic programming is performed
using residues for the superimposed structures

Open conformation in purple and the closed in blue

Red and light blue : Insertions

Will not find non-topological alignments (outside the bounds of the dotted lines)
CE works on chains and not in domains

3.2.3 CE, DALI, SSAP
3.2.3.1 CE



- Gaps included and analyzed for relocation in both directions
- RMSD improvements in superimposed structures
- New boundaries adopted
- Dynamic programming on the distance matrix using residues from the 2 superimposed structures

Method 2. From structural information AND adding composite properties

Similarity is calculated by adding the following properties represented as scores

P_{ij} measures the match between residues i and j from two proteins

Structure: Property 1, defined by coordinates of Cα

$$P_{ij} = \begin{cases} c_1 - d_{ij}, & \text{if } c_1 - d_{ij} > c_2 \\ c_2, & \text{otherwise} \end{cases}$$

Sequence: Property 2, value of PET91 matrix for amino acids at positions i and j

Secondary structure: Property 3

$$P_{ij} = \begin{cases} 1, & \text{if } s_i = s_j \\ 0, & \text{otherwise} \end{cases}$$

Solvent Exposure: Property 4

$$P_{ij} = E_0 - |E_i - E_j|$$

Conservation Index: Property 5

$$P_{ij} = 20 - |I_i - I_j|$$

3.2.3 CE, DALI, SSAP
3.2.3.1 CE

Method	PKA (1CDK:A) vs MAPK (1GOL:_) length of alignment = 248	PKA (1CDK:A) vs CDK2 (1FIN:A) length of alignment = 251
Dali	34 (13.7%)	30 (12.0%).
STR	8 (3.2%)	8 (3.2%)
STR+SEQ+CONS	3 (1.2%)	5 (2.0%)
SEQ	98 (39.5%)	76 (30.3%)
SS	76 (30.6%)	77 (30.3%).
CONS	84 (33.9%)	107 (42.6%)
EXP	45 (18.1%)	62 (24.7%)
STR+SEQ	4 (1.6%)	6 (2.4%)

STR: structure based on the rmsd calculated for the superposition of Cα atoms after optimal alignment found using the CE algorithm

SEQ: sequence based on PET91 amino-avid similarity measure by Jones and Thornton (1992)

SS: secondary structure based on the SSEs by Kabsch and Sander (1983)

EXP: solvent exposure based on the definition of Lee and Richards (1971)

CONS: conservation index based on sequences compiled for proteins with known structure

()Absolute difference between alignments

3.2.3 CE, DALI, SSAP
3.2.3.1 CE

The calculus is done one residue by residue

Dynamic programming to find the optimal alignment for the whole polypeptide chain.

The composite property that measure structural similarity at residue level is defined

$$\tilde{P}_{ij} = \sum_k w_k * P_{ij}^k$$

Gap initialization penalty of 10 and gap extension penalty of 1

$$a^D = \sum_i a_i^D$$

$$a^D = \begin{cases} 1, & \text{if } a_i^1 \neq -1 \text{ and } a_i^1 \neq a_i^2 \\ 0, & \text{otherwise} \end{cases}$$