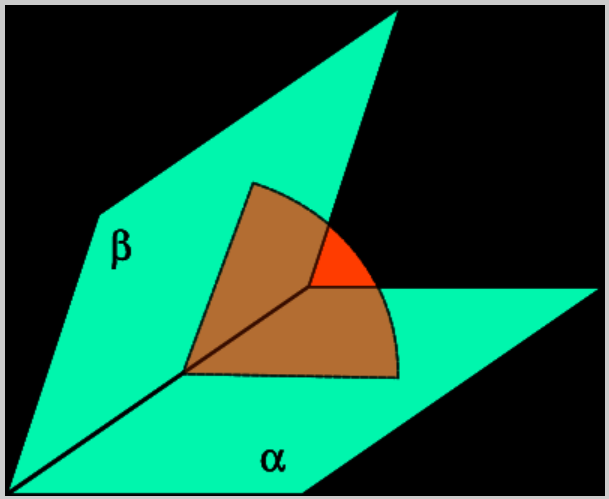
Dihedral Angles

Homayoun Valafar

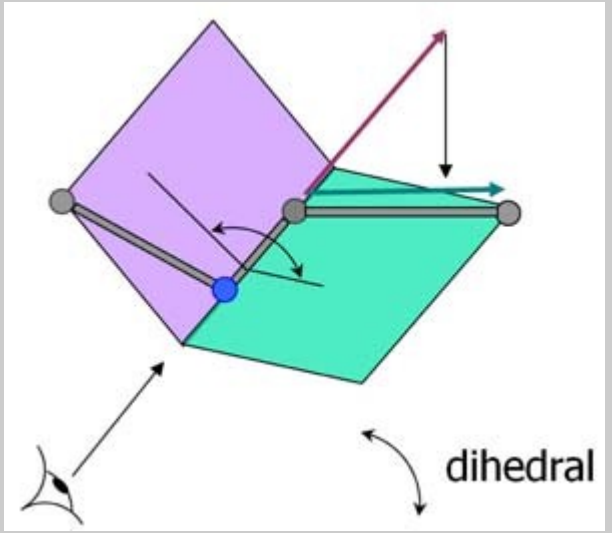
Department of Computer Science and Engineering, USC

Dihedral Angles

- The precise definition of a dihedral or torsion angle can be found in spatial geometry
  - Angle between to planes

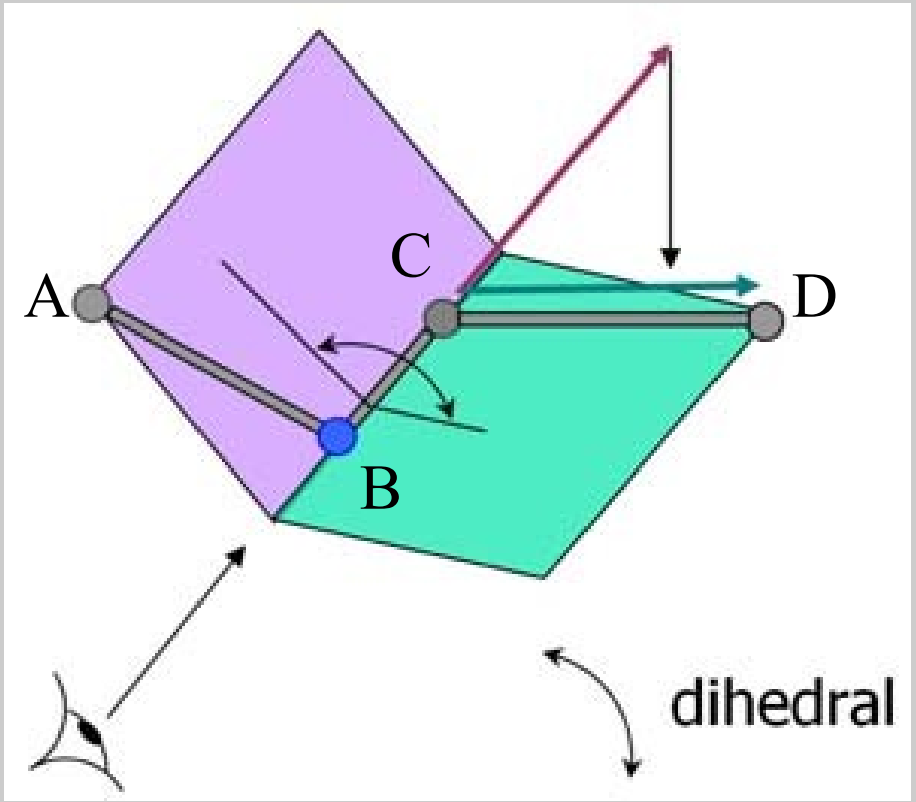


- Structural biology defines a torsion angle by four consecutive atoms



Torsion Angles in Proteins

- Defines rotation about a bond
- Four atoms are needed A – B – C – D to define the torsion angle about the B – C bond



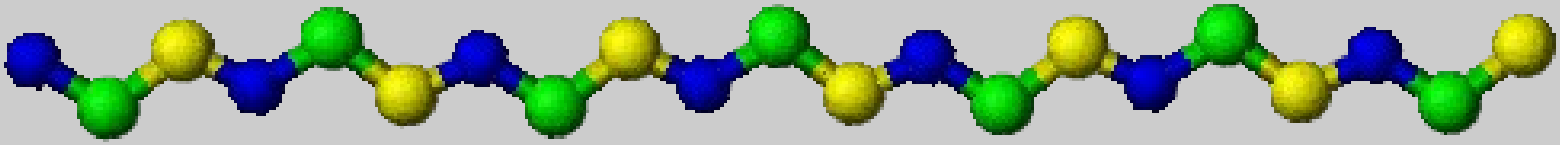
Torsion Angles in Proteins

- Backbone: repeating motif of -N_{blue} - Ca_{green} - C'_{blue} -
- Backbone can also be defined in terms of three torsion angles.
- ω : rotation about the C'-N bond (peptide bond)
  - Defined by Ca – C' – N – Ca
  - Ideally at ±180° with very small deviation allowed
- ϕ : rotation about the N – Ca bond
  - Defined by C' – N – Ca – C'
  - Restrained only by the Ramachandran space
- φ : rotation about the Ca – C' bond
  - Defined by N – Ca – C' – N
  - Restrained only by the Ramachandran space

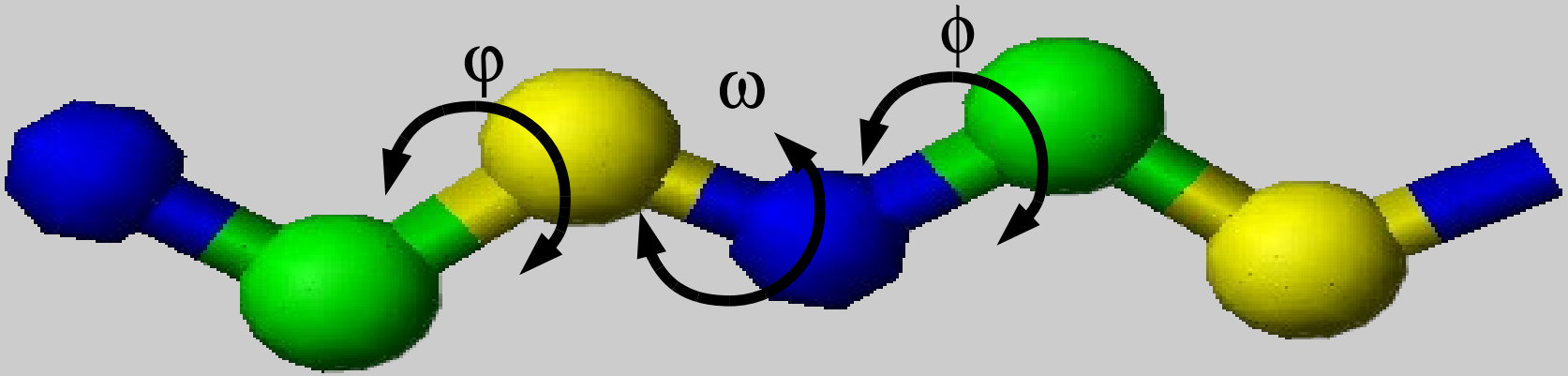
Torsion Angles in Proteins

Backbone atomic motif

N – Ca – C – N – Ca – C – N – Ca – C



Backbone torsion angle motif



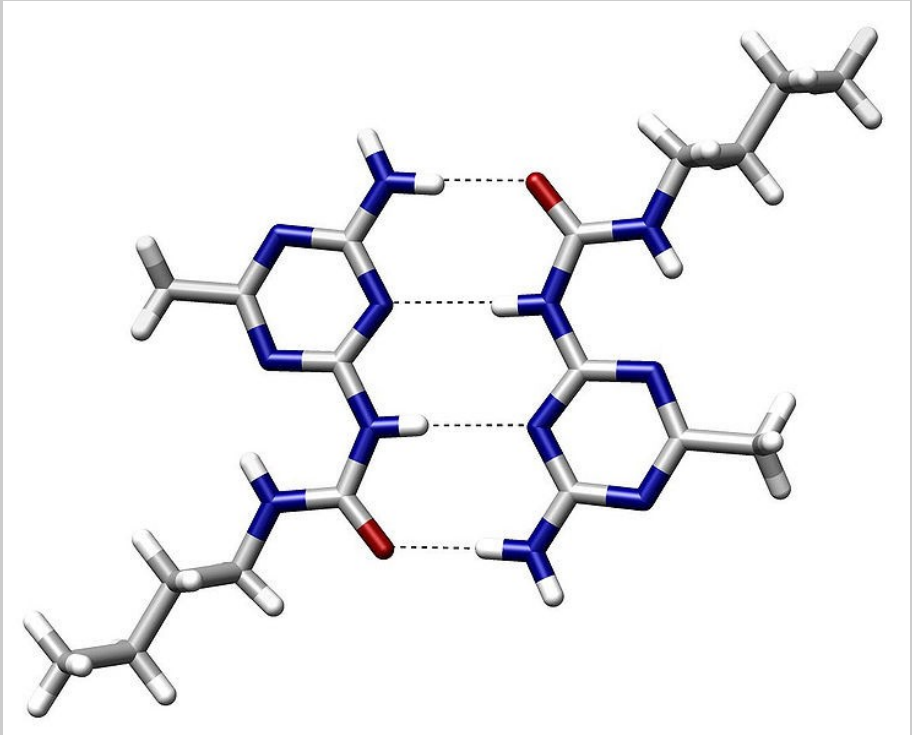
Energetics of Secondary Structural Elements

Homayoun Valafar

Department of Computer Science and Engineering, USC

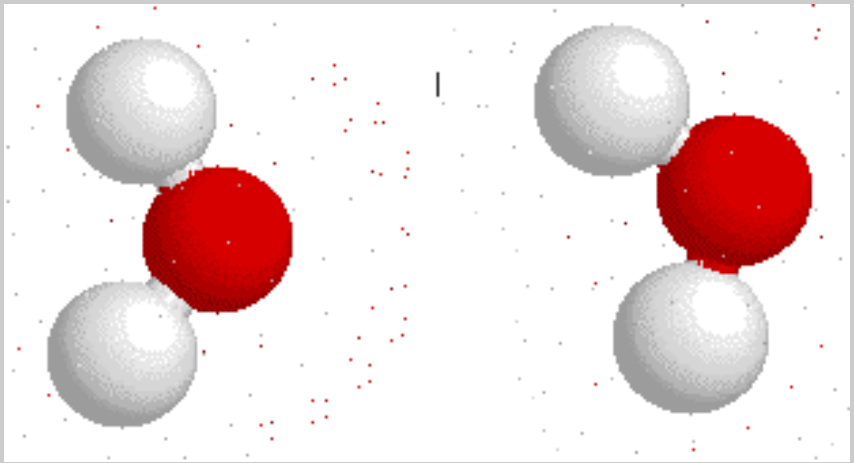
Hydrogen Bond

- A type of attractive intermolecular force that exists between two partial electric charges of opposite polarity.
- Stabilizes α-helical and β-sheet secondary structural elements (SSE).
- Hydrogen bond formation requires spatial vicinity (~ 2.0Å) and proper orientation of the electronic orbitals (<35°)



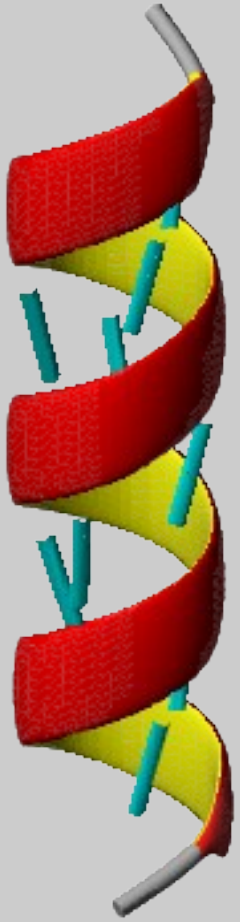
Hydrogen Bond

- Two participating atoms: donor and acceptor
- Normally between O-H or N-H
- Strongest non-bonded force



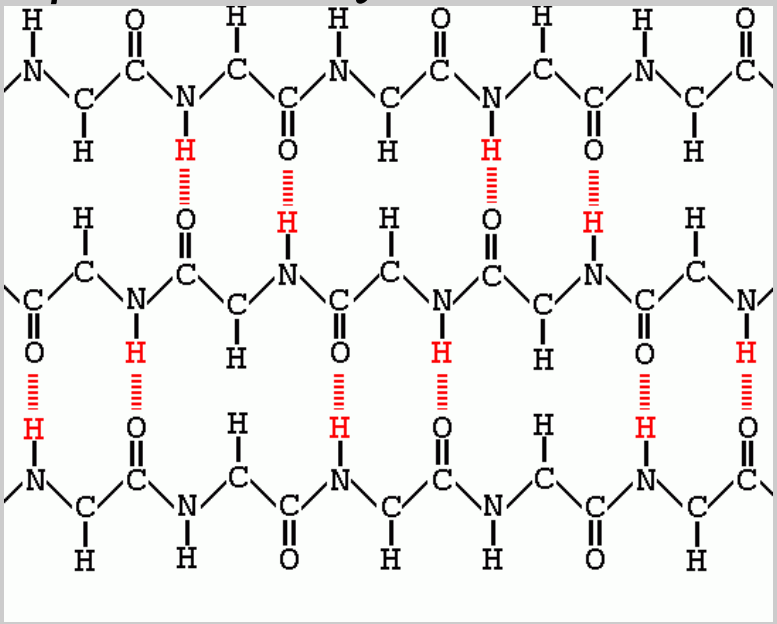
Hydrogen Bonds in Helical SSE

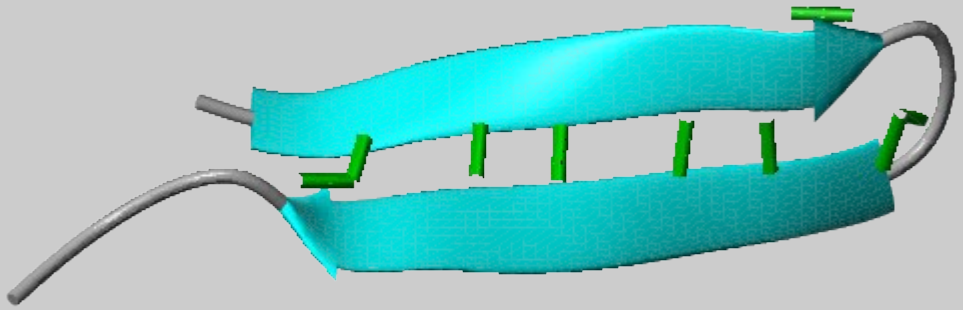
- Helices are internally stabilized by H-bonds
- A hydrogen bond between positions *i* and *i* + 4 forms an alpha helix
  - 3.6 residues per turn
  - 1.5Å rise per residue
  - Pitch of 5.4Å (rise per turn)
- A hydrogen bond between positions *i* and *i* + 3 forms a 3_{10} helix.
- A hydrogen bond between positions *i* and *i* + 5 forms a π–helix.
- Left handed helix



Hydrogen Bonds in Beta Sheet SSE

- β-Sheet are formed from individual β-strands
- Hydrogen bonds involving alternating residues on each participating strand forms a beta sheet.
- β-sheets can be parallel or anti-parallel
- β-sheets may involve discontinues and remote regions





Van der Waals Radius Ramachandran Space

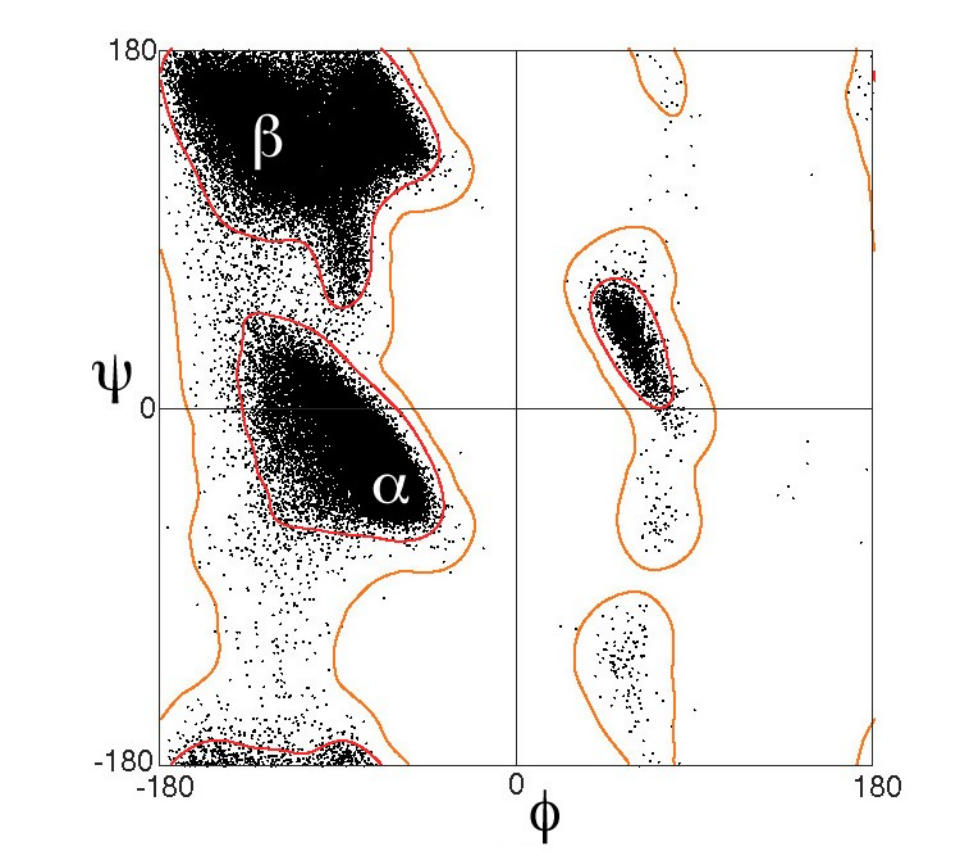
Homayoun Valafar

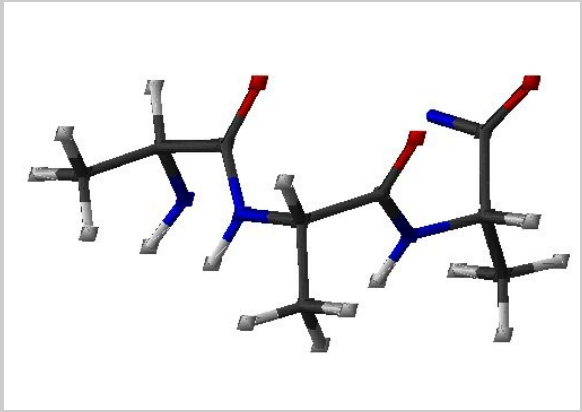
Department of Computer Science and Engineering, USC

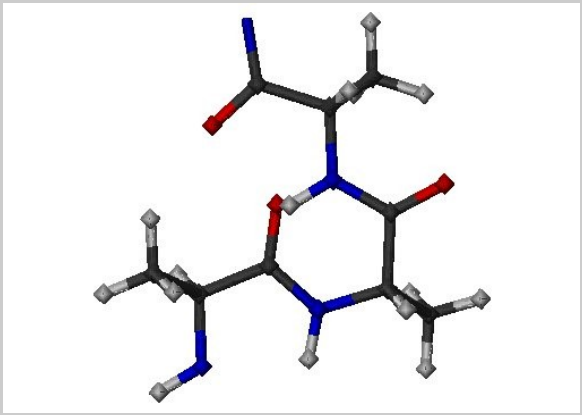
Steric Collision

- Some torsion angles may be energetically more favorable.
- Some torsion angles may be energetically very unfavorable
- Energies associated with different torsion angles can be interpreted as the probability of two peptide planes assume that local geometry

Ramachandran Space





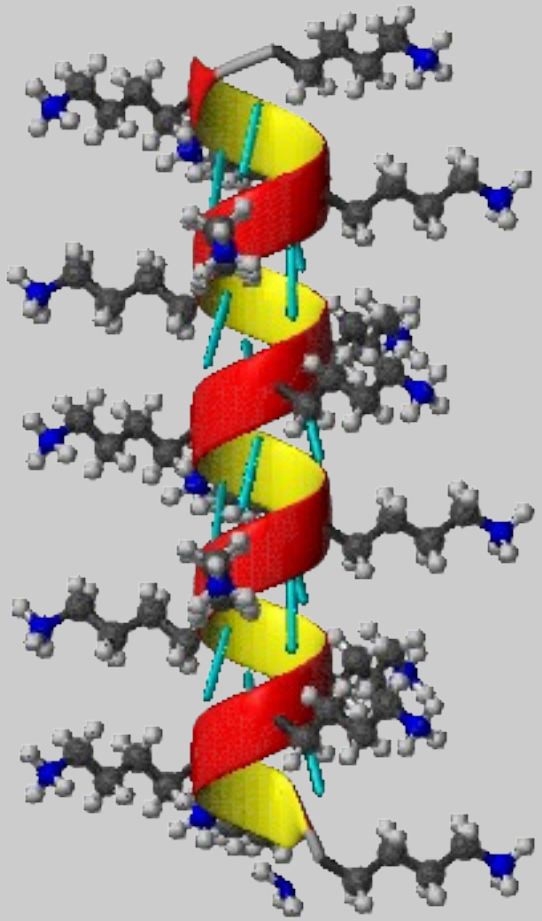


Ramachandran Space

- How would you determine Ramachandran space?
  - Theoretically
    1. Model forces
    2. Calculate forces for all torsion angles.
    3. Determine likelihood of a certain torsion angle.
  - Experimentally
    1. Collect all good structures determined experimentally.
    2. Find all torsion angles.
    3. Create a two dimensional histogram of torsion angles.

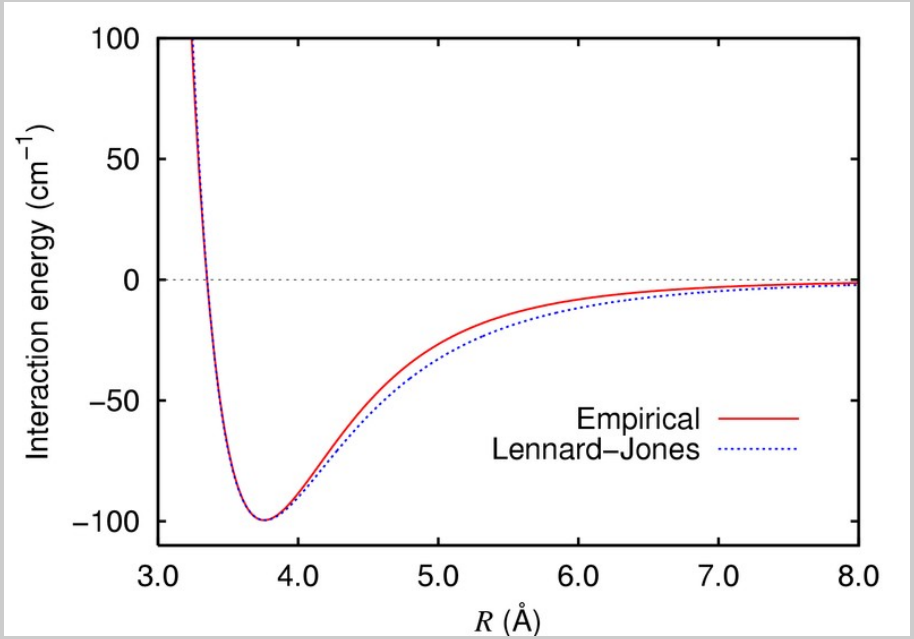
Ramachandran Space

- Ramachandran space:
  - Maximizes H-bond formation
  - Minimizes spatial occupation of atoms/groups of atoms



Lennard-Jones Potential

- Van der Waals forces may be:
  - Attractive in long range.
  - Repulsive in short range.
- Modeled by L-J poetntial:
  - ε is the well depth
  - σ is the van der Waals radia
  - Experimentally determined!
- (6-12) L-J potential is defined as:



$$V(r) = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$$

Ab Initio Protein Folding

Homayoun Valafar

Department of Computer Science and Engineering, USC

From Sequence to Structure

- Does primary sequence lead to functional structure?
  1. Take functional protein.
  2. Denature using urea or other agents.
  3. Confirm loss of function.
  4. Purify protein and reintroduce to physiological conditions.
  5. Confirm gain of function.
- In general protein sequence leads to functional structure.
- Simulation should allow computational folding of proteins.
  - Levitt, M. and A. Warshel, *Computer simulation of protein folding.* *Nature, 1975. 253: p. 694-698.*

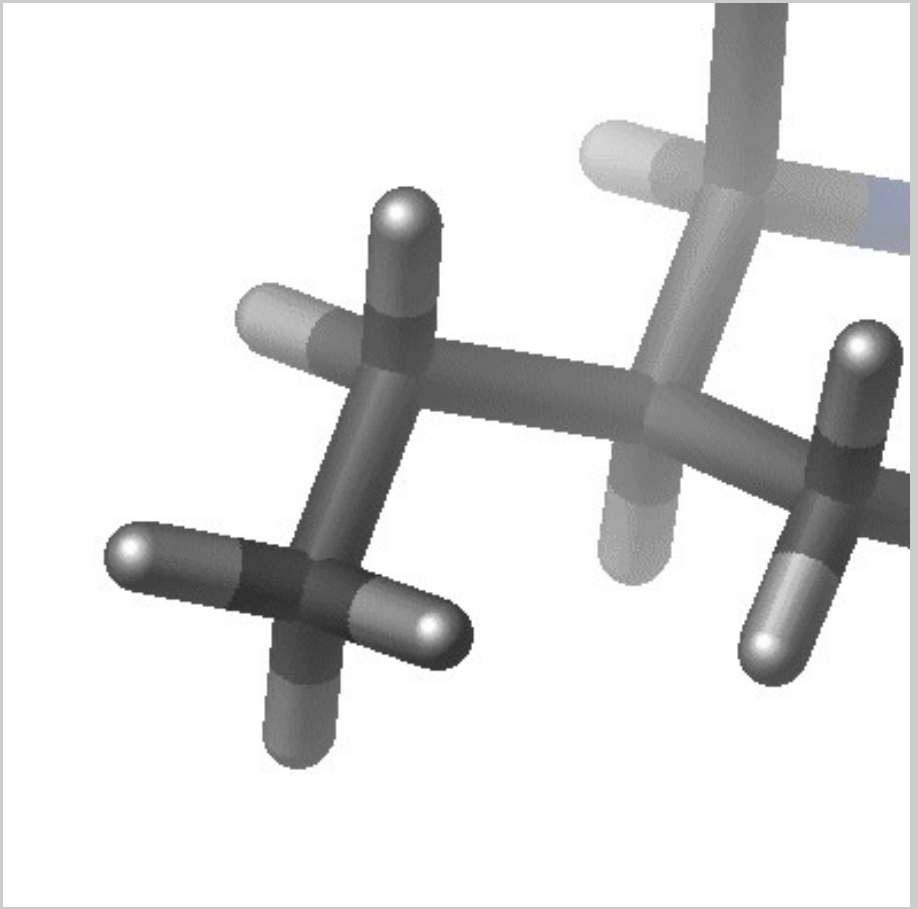
Total Potential Energy

- Mathematical expression of the potential function is necessary for simulation of protein fold.
- $E_{Total} = E_{Empirical} + E_{Effective}$
  - $E_{Empirical}$: energy of the molecule as a function of the atomic coordinates
  - $E_{Effective}$: restraining energy terms that use experimental information.
- Neglect $E_{Effective}$ term for true computational model.
- Select the structure with the lowest total energy is the final structure.

Potential Energy of Bond Lengths

- The bond length between each two atoms is known empirically
- Bond lengths should not exceed the expected values
- Requires atomic coordinates for two atoms

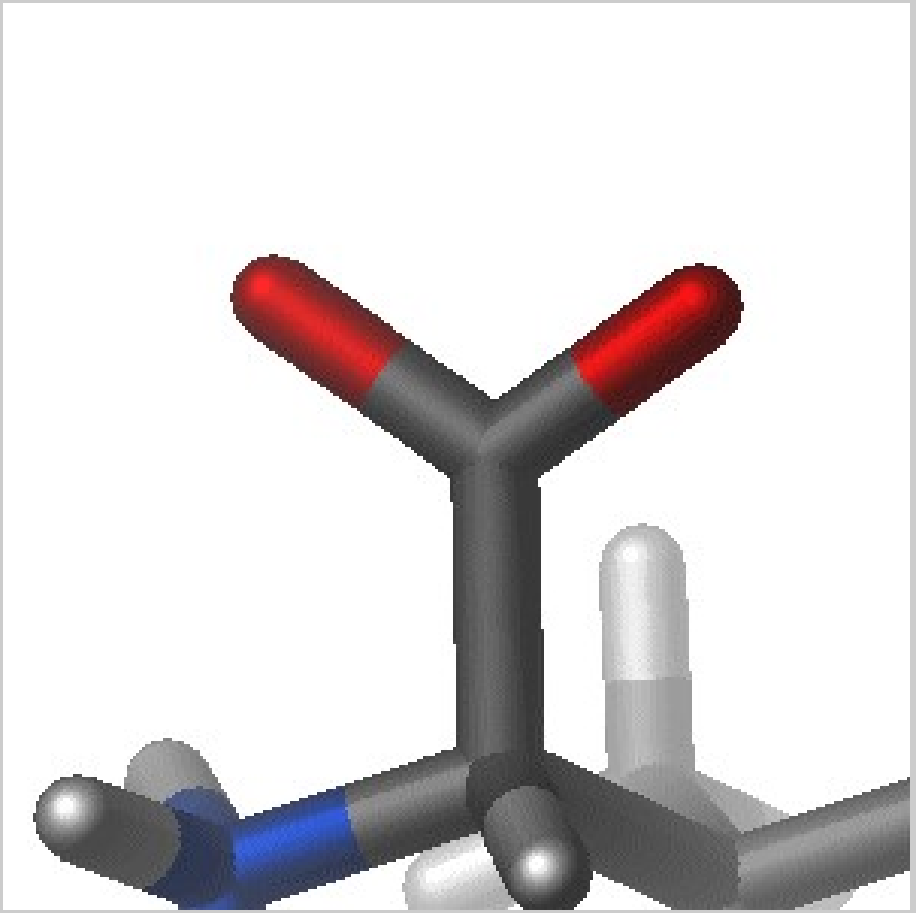
$$E_{BOND} = \sum_{bonds} k_b (r - r_0)^2$$



Potential Energy of Bond Angles

- Bond angles should not deviate from the known quantities
- Coordinates of three atoms is needed for this measure

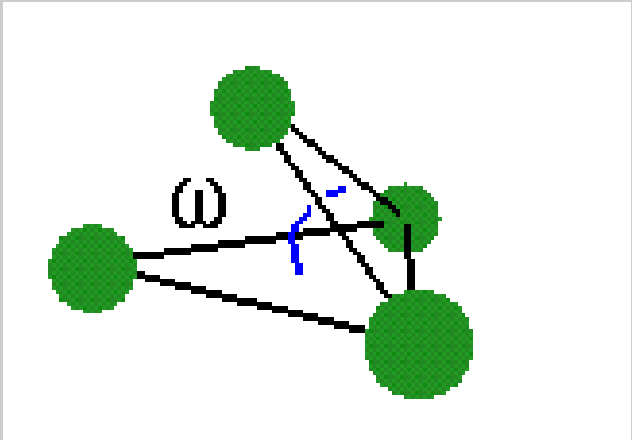
$$E_{ANGLE} = \sum_{angles} k_\theta (\theta - \theta_0)^2$$



P.E. of Improper Dihedrals

- Improper dihedrals represent the planarity of the peptide planes
- Four atoms are required for this measure

$$E_{IMPR} = \sum_{impropers} k_{\omega_i}(\omega_i - \omega_0)^2$$



Empirical Energy Terms

- All of the energy terms defined in terms of atomic coordinates of two, three and four atoms.
- Conformational Energy Terms:
  - E_{BOND} : describes the covalent bond energy over all covalent bonds
  - E_{ANGL} : describes the bond angle energy over all bond angles
  - E_{DIHE} : describes the dihedral angle energy over all dihedrals
  - E_{IMPR} : describes the improper angle energies (planarity and chirality)
- Nonbonded Energy Terms:
  - E_{VDW} : describes the energy of Van Der Waals terms
  - E_{ELEC} : describes the energy of electrostatic interactions

Other Potential Terms

- Hydrophobic and hydrophilic interaction.
  - Requires presence of water in the simulation.
  - Addition of water to the simulation is difficult.
  - Will require identification of cavities and calculation of movement of water molecules.
- Hydrogen bonds:
  - Also requires assessment of water accessibility.
  - Water interferes with formation of hydrogen bonds.
- Gas phase simulation
  - Absence of water.
  - Computationally much more convenient

Total Energy Term

- Force Field: A vector field representing the gradient of the total potential.

$$E_{Total} = \sum \left[w^p_{BOND} E_{BOND} + w^p_{ANGL} E_{ANGL} + w^p_{DIHE} E_{DIHE} + w^p_{IMPR} E_{IMPR} + w^p_{VDW} E_{VDW} + w^p_{ELEC} E_{ELEC} \right]$$

$$E_{BOND} = \sum_{bonds} k_b (r - r_0)^2$$

$$E_{ANGL} = \sum_{angles} k_\theta (\theta - \theta_0)^2$$

$$E_{DIHE} = \sum_{dihedrals} \sum_{i=1,m} \begin{cases} k_{\varphi_i}(1 + \cos(n\varphi_i + \delta_i)) & n_i ¿ 0 \\ k_{\varphi_i}(\varphi_i - \delta_i)^2 & n_i = 0 \end{cases}$$

$$E_{IMPR} = \sum_{impropers} \sum_{i=1,m} \begin{cases} k_{\varphi_i}(1 + \cos(n\varphi_i + \delta_i)) & n_i ¿ 0 \\ k_{\varphi_i}(\varphi_i - \delta_i)^2 & n_i = 0 \end{cases}$$

$$E_{VDW} = \sum_{vdw} \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}}$$

$$E_{ELEC} = \sum_{i,j} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}$$

Force Field

- Technically, the derivate of the potential energy.
  - A vector field of forces.
- Some currently existing force fields (forcefield):
  - Explor-NIH
  - AMBER
  - CHARMm
  - MM2, MM3 and MM4
  - Sybyl
  - Etc.

Minimization of Total Energy

- Theoretically, the structure with the minimum total energy is the structure of interest.
- A number of minimization algorithms can be utilized.
  - Gradient descent
  - Monte Carlo and Simulated Annealing
  - Newton's
  - Genetic Algorithm
  - Distributed Global Optimization
  - Branch and Bound
  - …

Complexity of The Problem

- Assuming a protein with 100 residues and in average 10 atoms per residue, what is the complexity of this problem?
- What are the variables of this problem? How many?
- How complex is the total energy landscape?
- How costly is each evaluation of the E_{Total} and its gradient?
- Beyond our computational capabilities.