Protein Energetics

Dr. Fayyaz ul Amir Afsar Minhas
PIEAS Biomedical Informatics Research Lab
Department of Computer and Information Sciences
Pakistan Institute of Engineering & Applied Sciences
PO Nilore, Islamabad, Pakistan
http://faculty.pieas.edu.pk/fayyaz/

Protein Energetics

- Why do proteins fold?
- What is the physics behind the behavior of proteins?
- How can we computationally model this physics?
- Meeting point of Physics, Chemistry and Biology

Concept

- Probability of finding a (mole of a) molecule in a particular state

$$p_A = w_A exp\left(-\frac{E_A}{RT}\right)$$

  - E_A: Total energy of the system (molecular kinetic, rotational, vibrational, interactions within and between molecules of the system)
  - w_A: Number of ways in which that total energy may be achieved or distributed, i.e., entropy
  - T: Temperature
  - R: Gas Constant

  - What states are more probable?
    - Based on energy
    - Based on entropy

https://en.wikipedia.org/wiki/Boltzmann_distribution

Observing a molecule in a state

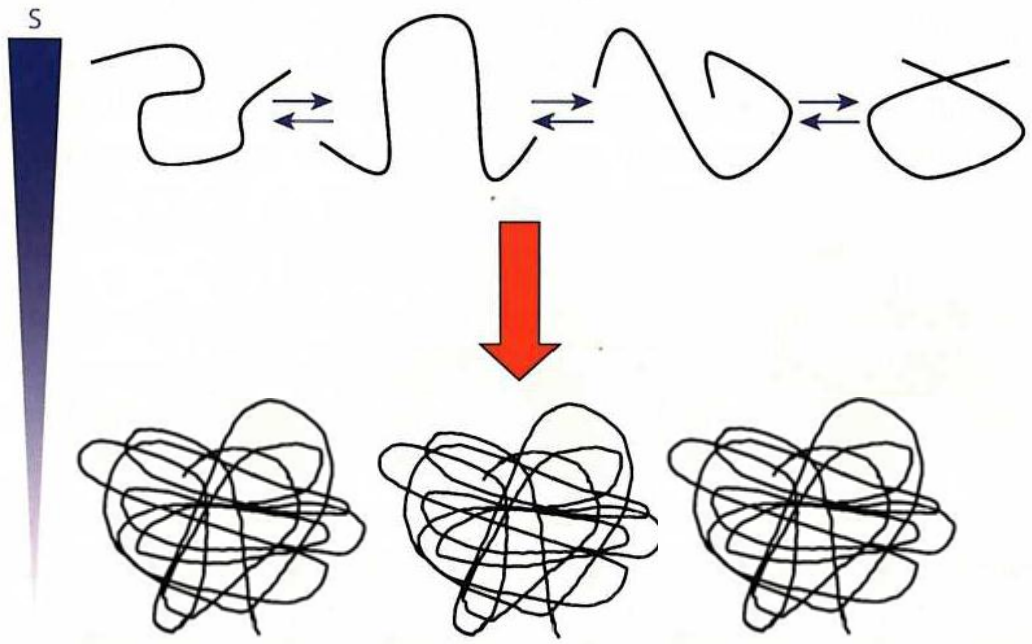
- Assume two states $A \rightleftharpoons B$
  - $p_A = w_A exp\left(-\frac{E_A}{RT}\right)$
  - $p_B = w_{\boldsymbol{B}} exp\left(-\frac{E_{\boldsymbol{B}}}{RT}\right)$
- The reaction constant is the ratio of the amount of B to that of A, i.e.,
  - $K = \frac{[\boldsymbol{B}]}{[\boldsymbol{A}]} = \frac{p_B}{p_A} = \frac{w_{\boldsymbol{B}} exp\left(-\frac{E_{\boldsymbol{B}}}{RT}\right)}{w_A exp\left(-\frac{E_A}{RT}\right)} = \frac{w_{\boldsymbol{B}}}{w_A} exp\left(-\frac{E_{\boldsymbol{B}} - E_A}{RT}\right)$
  - $\boldsymbol{\Delta G = \Delta H - T\,\Delta S}$
- For a biological system (which doesn't undergo changes in temperature or pressure)
  - $\boldsymbol{\Delta G = -RTln(K)}$
  - $\boldsymbol{\Delta E} = E_{\boldsymbol{B}} - E_A = \boldsymbol{\Delta H}$ = Change in internal energy
  - $\boldsymbol{\Delta S} = \boldsymbol{Rln}\left(\frac{w_{\boldsymbol{B}}}{w_A}\right)$ = Change in entropy

Spontaneity of a reaction

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

- $\Delta G = -RT ln\left(\frac{p_B}{p_A}\right) > 0$ **implies** $p_B < p_A$
  - Reaction would not be spontaneous
- $\Delta G = -RT ln\left(\frac{p_B}{p_A}\right) < 0$ implies $p_B > p_A$
  - Reaction would be spontaneous

Let's talk about proteins: entropy



Unfolded state
Ensemble of conformations existing in equilibrium which may be very different from each other

$$W_{unfolded}$$

Folded state
is a much smaller ensemble of conformations which exist in equilibrium. We can say that there is only one average conformation.

$$W_{folded}$$

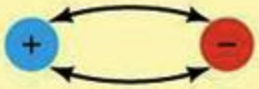
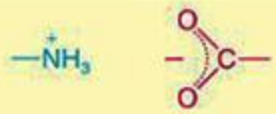
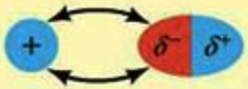
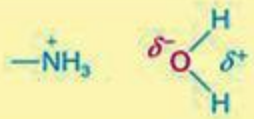
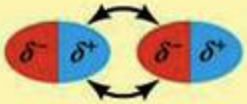
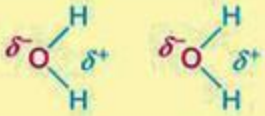
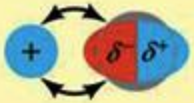
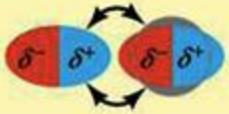
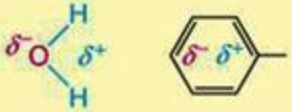
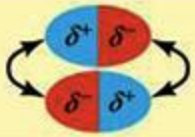
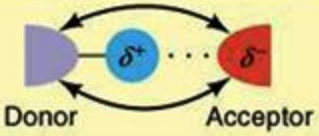
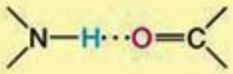
- Significant decrease in entropy
  - $\Delta S = Rln\left(\frac{W_{folded}}{W_{unfolded}}\right) < 0$

- Thus $\Delta S_{protein} < 0$
- For the process to be spontaneous
  - $\Delta G = \Delta H - T\Delta S < 0$
- Since $-T\Delta S > 0$, ΔH must be significantly negative
- Thus, we need to compute the change in energy of the system
  - For this, we need to calculate the energy of the system before and after folding

Calculating change in Energy of a protein

- Binding Energies
  - Disulfide
  - Bound ions
  - Etc,
- Non-binding energies
  - Electrostatic
  - Hydrogen bonds
  - Van der Waals Forces

Types of interactions

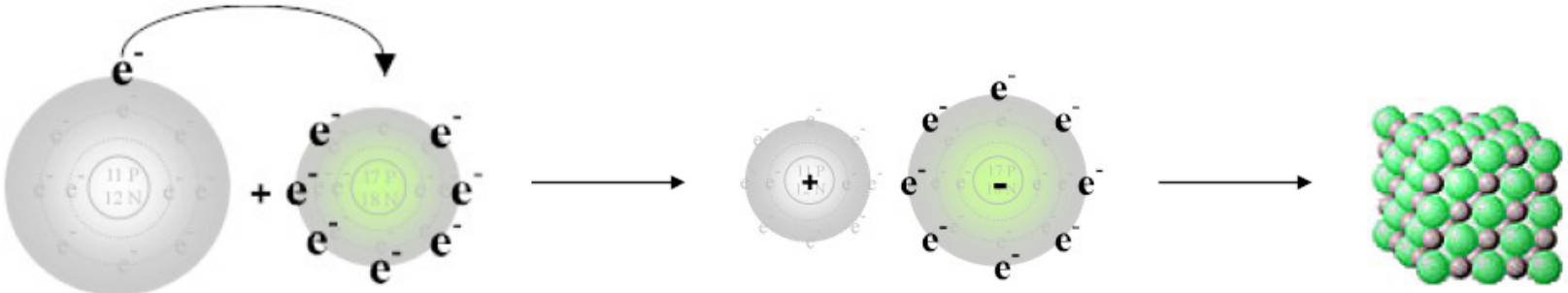
Type of Interaction	Model	Example	Dependence of Energy on Distance
(a) Charge–charge Longest-range force; nondirectional			$1/r$
(b) Charge–dipole Depends on orientation of dipole			$1/r^2$
(c) Dipole–dipole Depends on mutual orientation of dipoles			$1/r^3$
(d) Charge–induced dipole Depends on polarizability 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 (van der Waals) Involves mutual synchronization of fluctuating charges			$1/r^6$
(g) Hydrogen bond Charge attraction + partial covalent bond			Length of bond fixed

Electrostatic Interactions

Charged groups attract or repel each other. The force F of such an electrostatic interaction is given by Coulomb's law:

$$F = \frac{q_1 q_2}{D r^2}$$

q_1 and q_2 are the charges
r is the distance
D is the dielectric constant

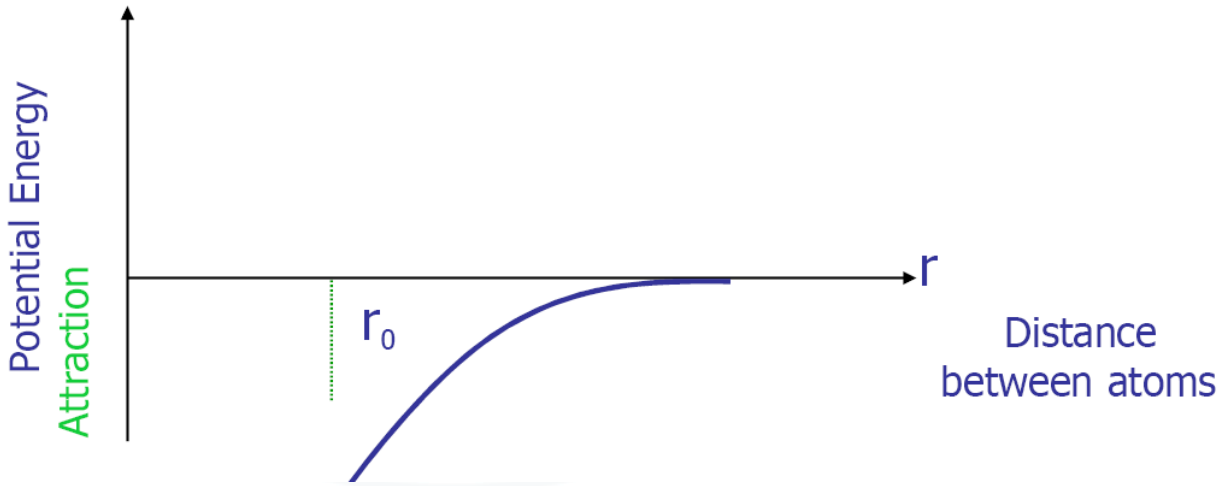


Coulomb's law is also used to determine interactions between uncharged, but polar atoms.

Van der Waals Interactions

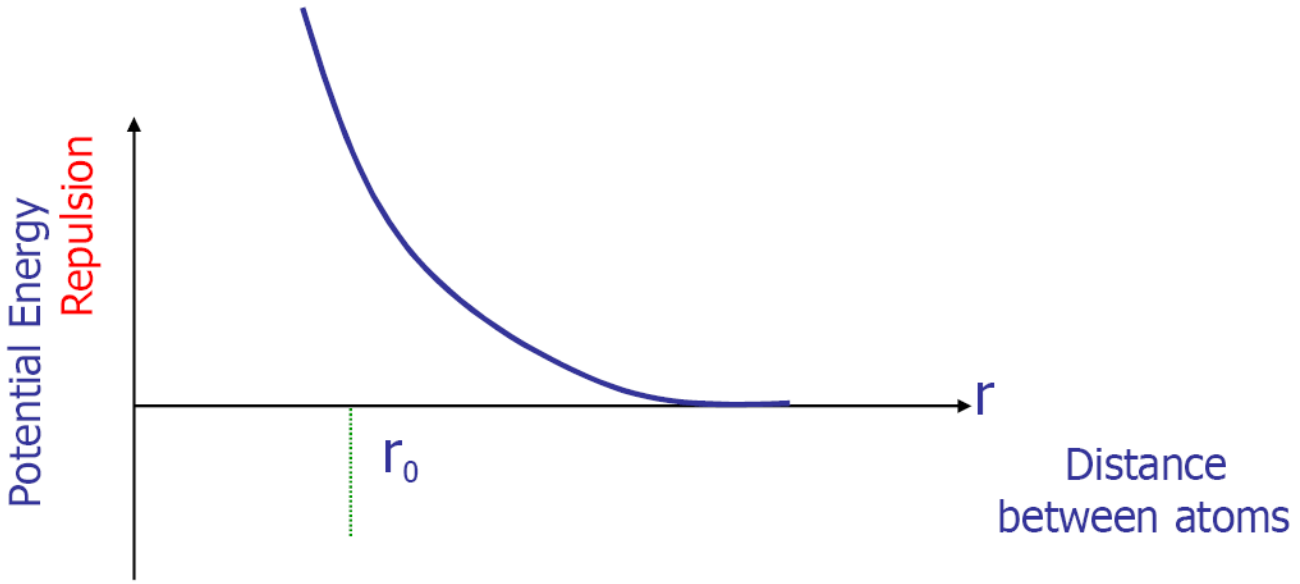
The distribution of electronic charges around an atom changes with time, and a transient asymmetry in the charges around one atom induces a similar asymmetry in the electron distribution around its neighboring atoms.

This is essentially an electrostatic interaction and results in a small distant-dependent (R^{-6}) attractive force.



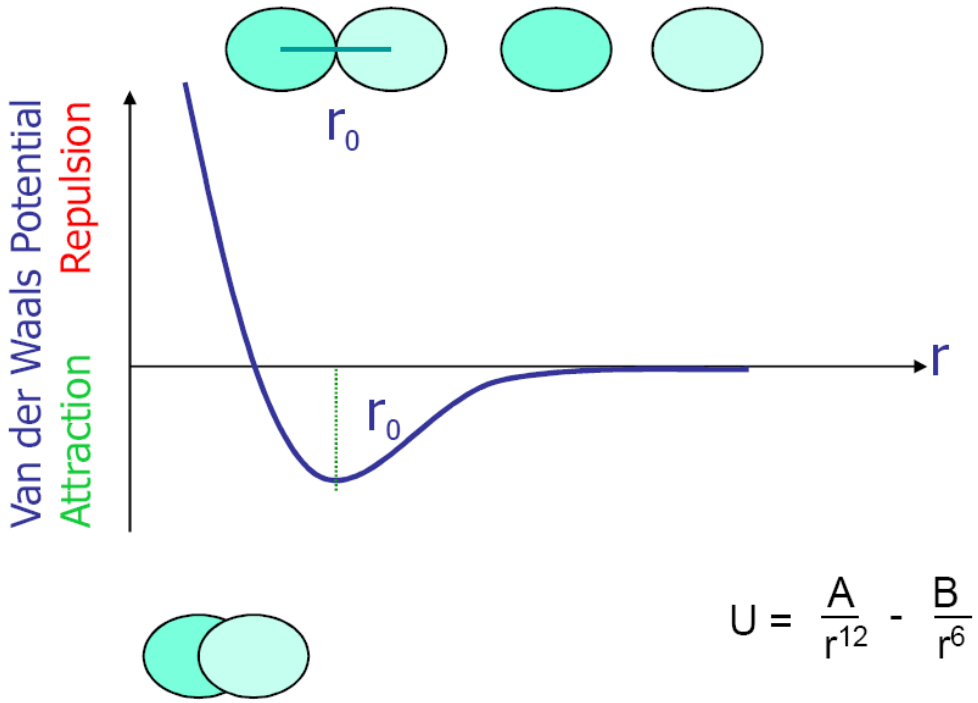
Van der Waals Interactions

As atoms get too close, their electron clouds will clash, resulting a distant-dependent (R^{-12}) repulsive potential energy.



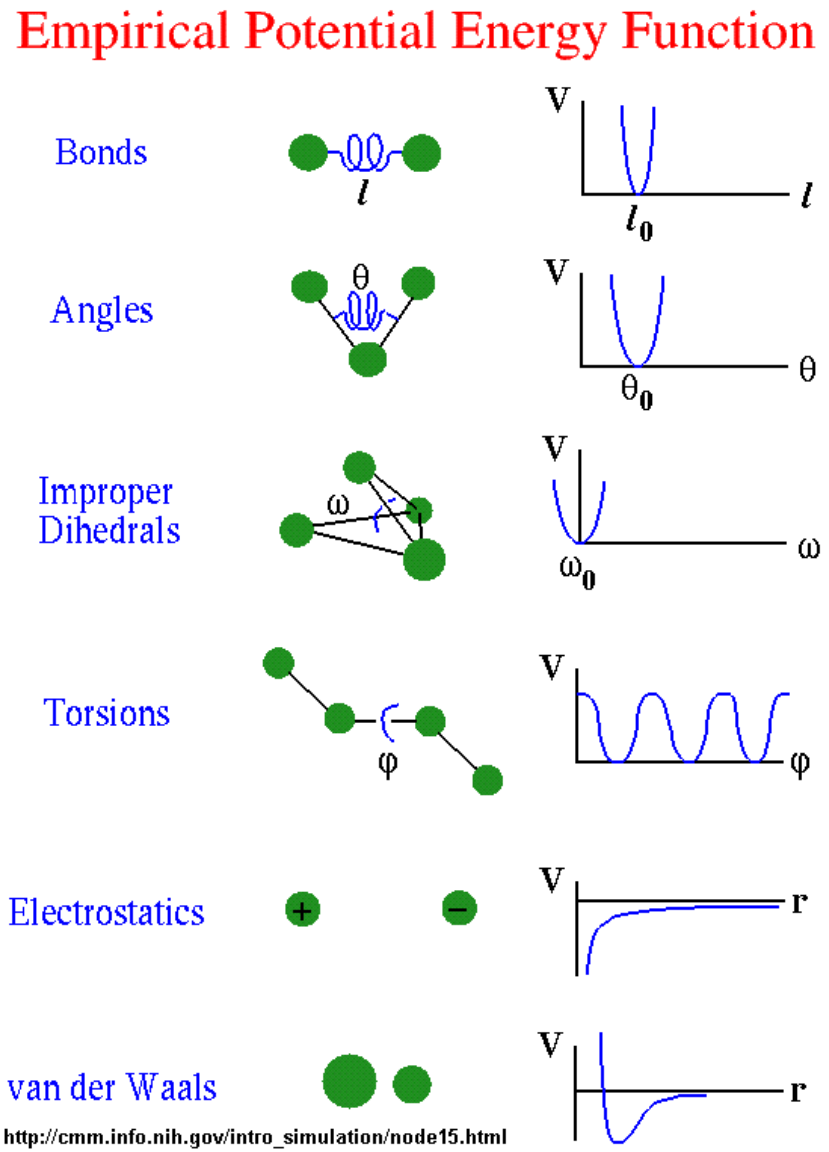
Lennard-Jones Potential

- The attractive and repulsive terms can be summed together to describe a distance-dependent interatomic potential energy.



Energy Calculations

- We can measure the potential energy of a molecule in molecular mechanics
  - $E_{total} = E_{bond} + E_{angle} + E_{torsion} + E_{electro} + E_{vdw}$
  - $\Sigma_{(i,j)\ E\ Sbond}\ k_{ij}^{b}(\alpha_{ijk}-\alpha_{ijk}^{0})^2$
    - Bond length
    - Bond Angle
    - Bond Torsion
  - van der Waals
    - $E_{vdw} = \Sigma_{(i,j)\ E\ Svdw}\ \varepsilon_{ij}\ [\ (\sigma_{ij}/r_{ij})^{12} - 2(\sigma_{ij}/r_{ij})^{6}]$
  - Electrostatics
    - $E_{electro} = \Sigma_{(i,j)\ E\ Selectro}\ (\ q_i q_j)/(e_{ij} r_{ij})$



http://en.wikipedia.org/wiki/Potential_energy_of_protein

Force Fields

- The energy function is a weighted summation of different contributing terms (electrostatics, VDW, …).
- These terms involve a set of parameters. The exact functional form of the energy function is called a “force field”.
- There are a number of different force fields available. For example:
  - AMBER, CHARMM, GROMOS, OPLS
- Example

$$V(r^N) = \sum_{\text{bonds}} k_b(l - l_0)^2 + \sum_{\text{angles}} k_a(\theta - \theta_0)^2$$

$$+ \sum_{\text{torsions}} \sum_{n} \frac{1}{2} V_n[1 + \cos(n\omega - \gamma)] + \sum_{j=1}^{N-1} \sum_{i=j+1}^{N} f_{ij} \left\{ \epsilon_{ij} \left[\left(\frac{r_{0ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{0ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right\}$$

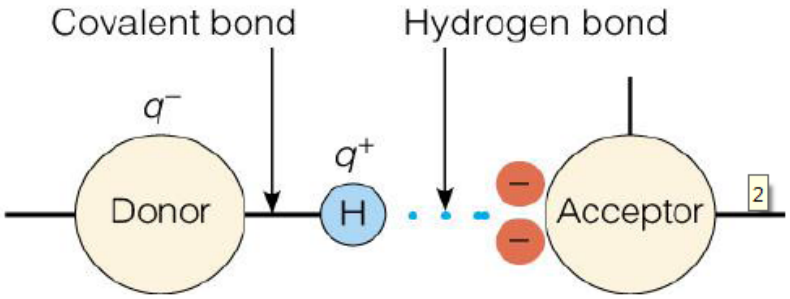
https://en.wikipedia.org/wiki/AMBER

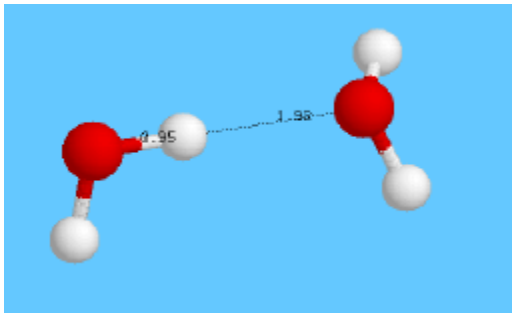
Discrepancies

- A number of people have experimentally calculated the change in energy for some proteins due to folding
- However, it was found that this change in energy is small and is, by itself, insufficient to cause
- There must be some other factor
  - Entropy
  - But that decreases
- So what is the factor?
  - Water

To understand the structure of individual water molecules in ice and liquid water, you first need to understand the hydrogen bond, one of the major types of noncovalent interactions.

H-bond definition: interaction between a covalently bonded electropositive hydrogen atom on a donor group and a lone pair of non-bonded electrons on an acceptor group





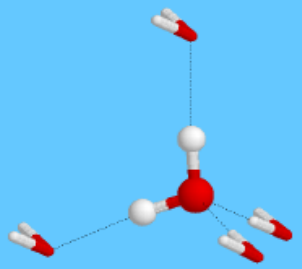
H-Bonding in water

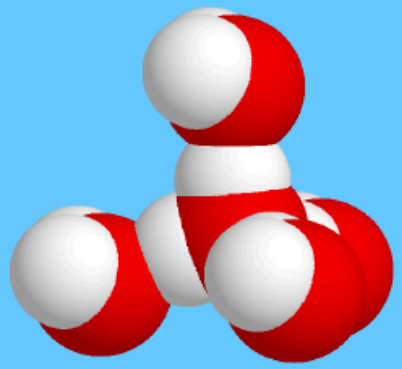
Water is both an H-bond donor and acceptor!!

The hydrogen bond has some features of covalent bonding: it is directional, strong, produces interatomic distances shorter than sum of van der Waals radii

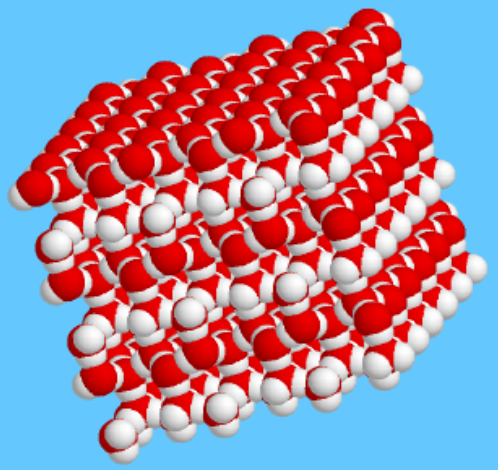
Water forms H-bonds which are near perfect in Ice

Water has tetrahedral geometry in ice

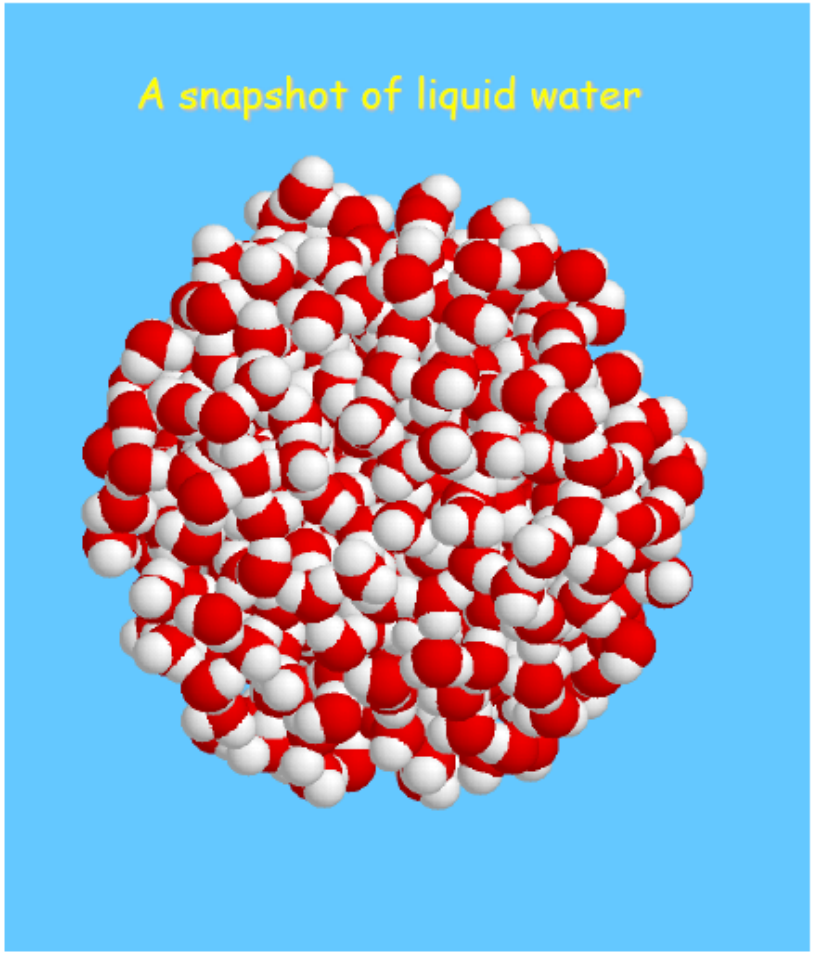




432 water molecules in ice



- Water is >99% H-bonded in ice
- When water freezes at 0°C, its volume increases by about 9% (and its density decreases). The packing density of ice at 0°C is 0.34. This means that 66% of the volume of ice is unoccupied because the packing is restricted by the hydrogen-bonding geometrical constraints.



Same 432 water molecules as liquid bulk water

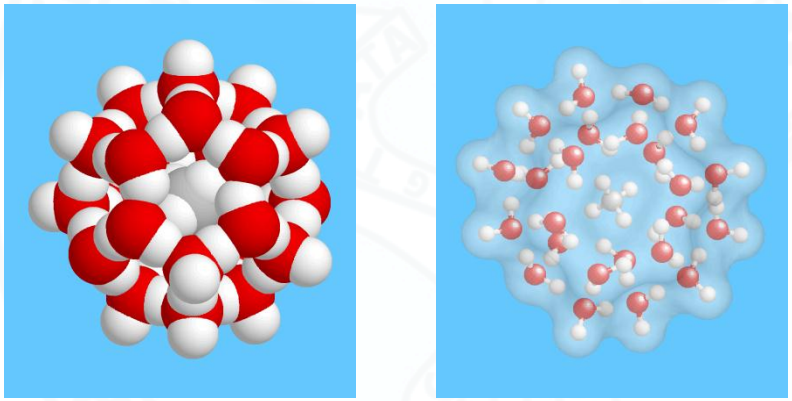
95% of the possible H-bonds are engaged in liquid water (and 85% in boiling water!)

In liquid water, hydrogen bonds are constantly stretching, bending, and breaking as the molecules rotate and jump around. **There is no regular stable tetrahedral geometry.** The average lifetime of a hydrogen bond is about one picosecond (10^{-12} sec) in liquid water at 25°C.

It is this network of "flickering" hydrogen bonds that gives liquid water its unique properties. This flickering also accounts for the fact that while water is more dense than ice **(packing density is 0.37 at 4°C)**, the 'collapsed' structure is still 'open' because of the highly directional character of the H-bonds.

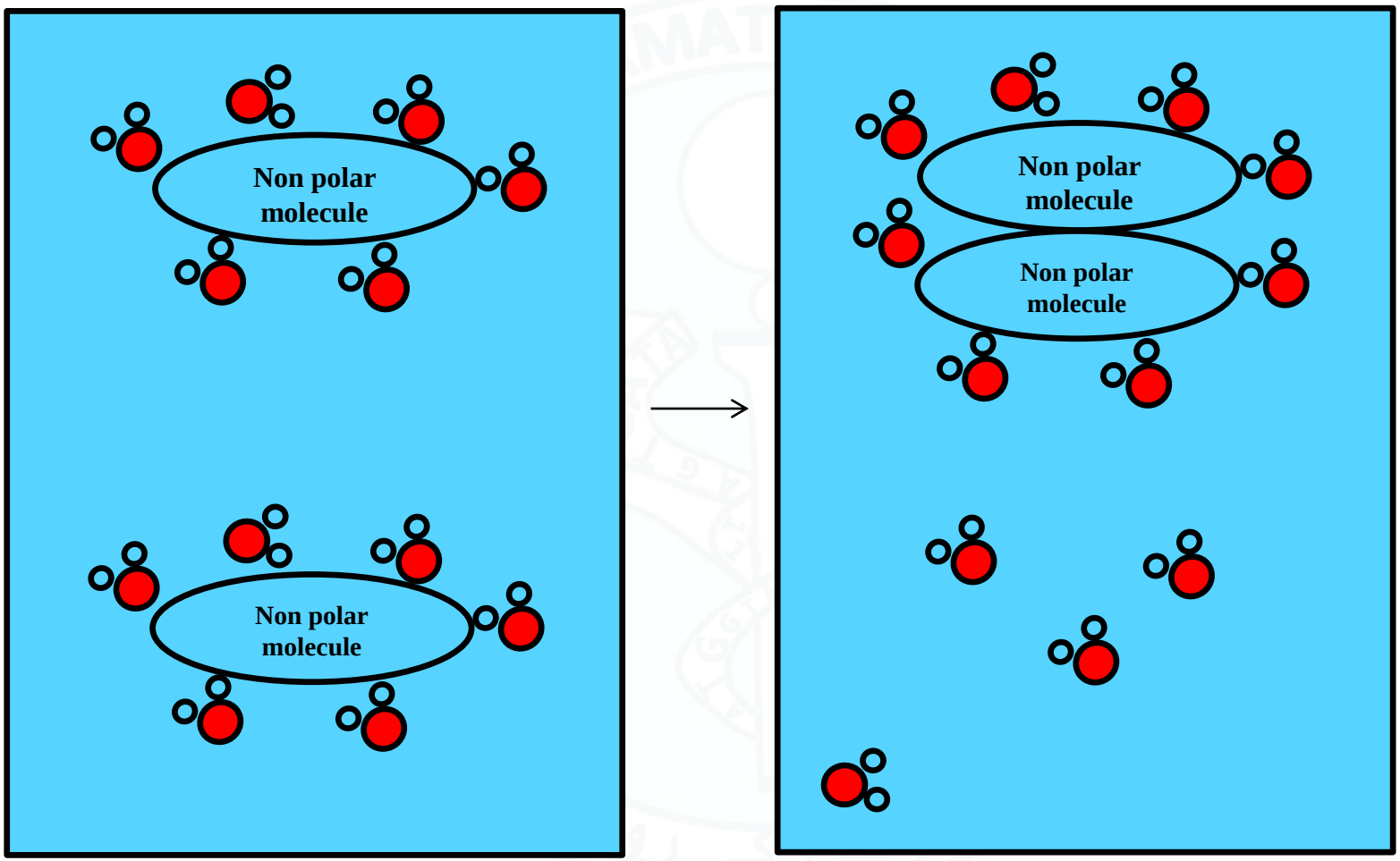
Behavior of water around hydrophobic molecules

- When you add a drop of hydrophobic residues to water (oil drop concept)
  - Water forms an ordered cage around that



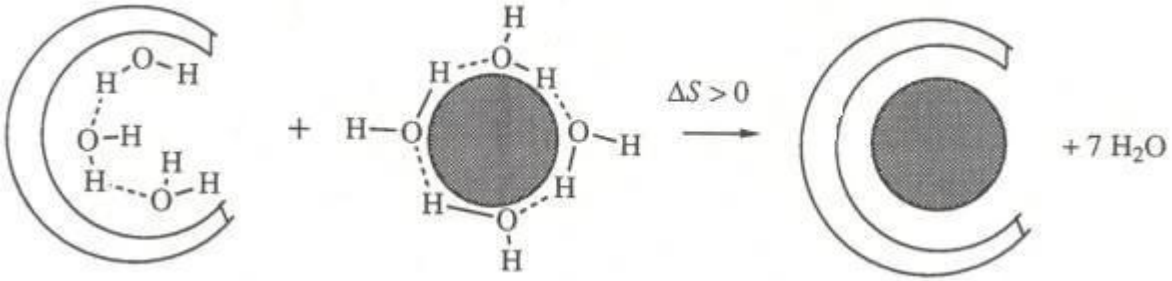
Methane in Water Mimics the Hydration of Hydrophobic Surfaces
Note that the water molecules avoid pointing their hydrogen-bonding groups toward the methane molecule. To do so would waste hydrogen bonds." In other words, they form and ordered cage around methane relative to the H-bonded waters in liquid water. The first-shell water molecules lose entropy to gain hydrogen bonding.

The Hydrophobic Effect



Water molecules have less degrees of freedom in the clathrate cage arrangements because some H-bonds cannot point inside toward the hydrophobic sphere

The Hydrophobic Effect

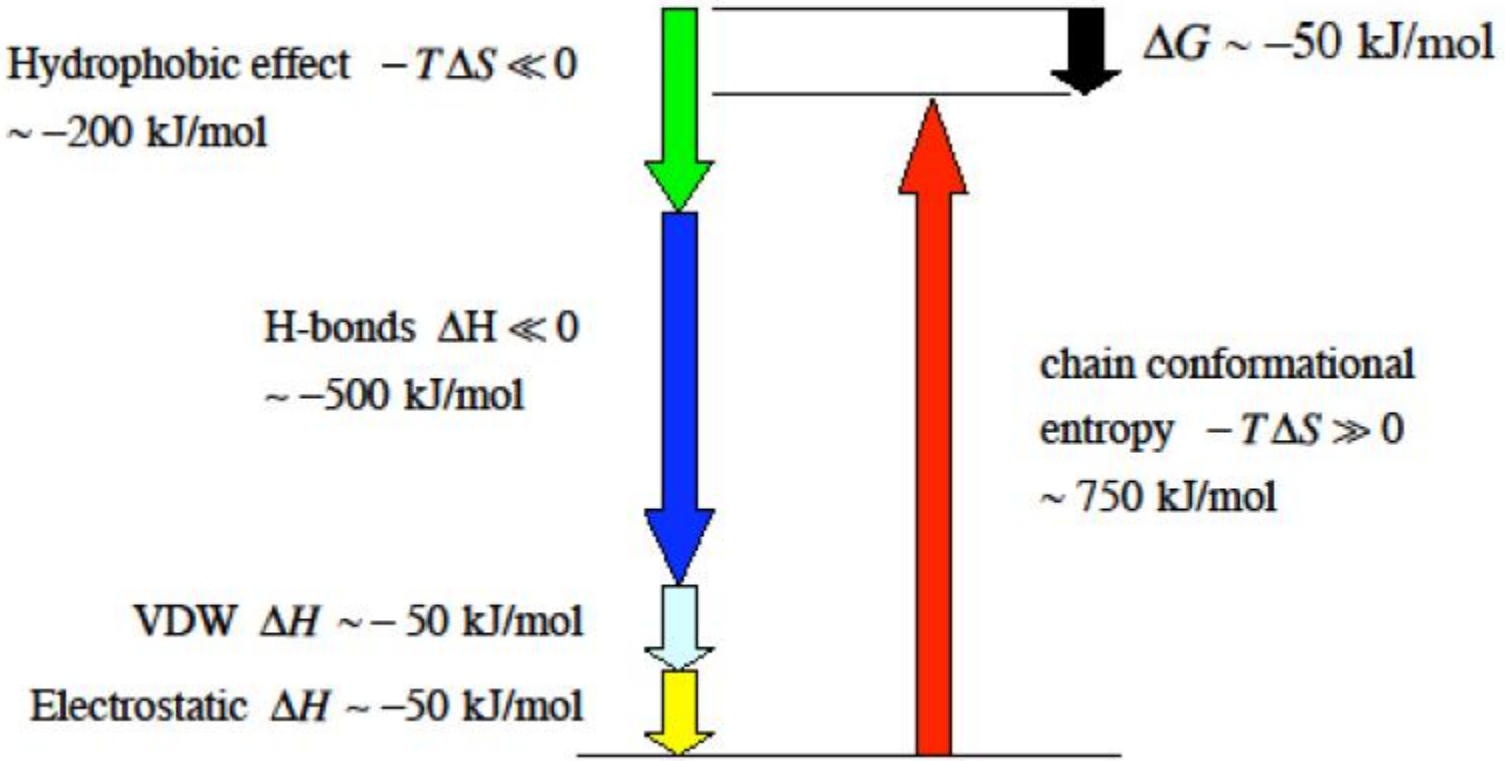


- There are “ordered water shells” around nonpolar groups and molecules in water. These hydration shells form to maximize H-bonding
- When nonpolar surfaces come in contact, the ordered water molecules are released into bulk solution, and the nonpolar surfaces are buried away from water in the process
- This leads to a large increase in ΔS, which drives association of the nonpolar surfaces

Dissecting the free energy of protein folding

$$\text{Unfolded} \overset{\Delta G}{\rightleftharpoons} \text{Folded}$$

$$\Delta G = \Delta H - T\Delta S < 0, \quad \Delta G = \sim -50 \text{ kJ/mol}$$



Energy Calculations

- Energy/Potential Function
  - Configuration of a protein (**X**)
    - For a sequence S of a protein, the location of the atoms in a protein with a given sequence can be determined if the torsion angles and rotamers are given.
  - For a given configuration of a protein, we can calculate its energy based on the interactions between the atoms by analyzing and combining the impact of different energy terms

Things that can change in a protein

- Sequence
  - Sequence of the protein determines its native structure
  - The change in the sequence of the protein is likely to cause a change in the structure of the protein
  - Most of the time, a particular sequence gives rise to a certain structure

Things that can change in a protein

- Torsion angles

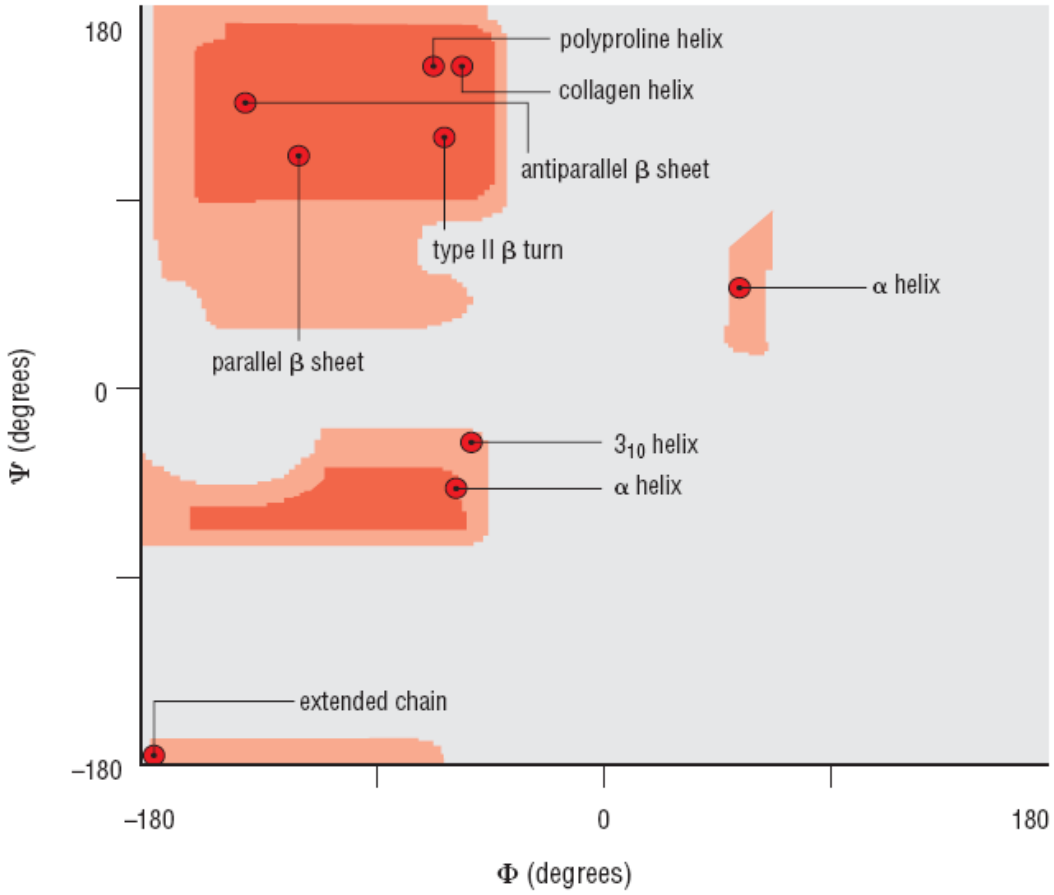
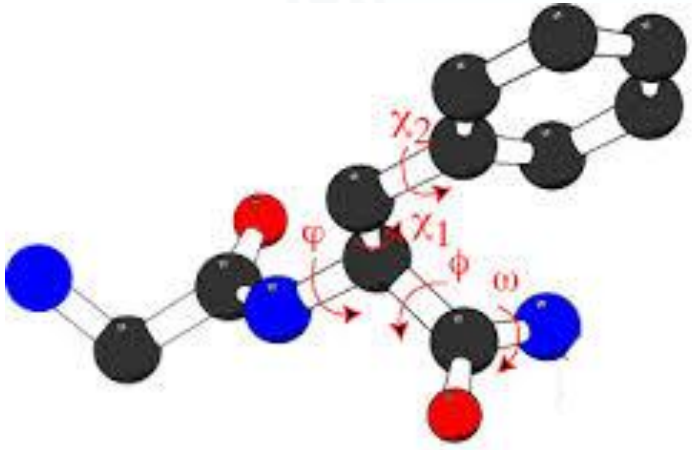
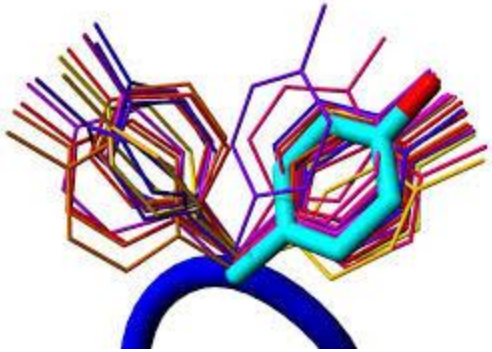


Figure 1-11 Ramachandran plot Shown in red are those combinations of the backbone torsion angles phi and psi (see Figure 1-9) that are "allowed" because they do not result in steric interference. The pink regions are allowed if some relaxation of steric hindrance is permitted. Common protein secondary structure elements are marked at the positions of their average phi, psi values. The isolated pink alpha-helical region on the right is actually for a left-handed helix, which is only rarely observed in short segments in proteins. The zero values of phi and psi are defined as the *trans* configuration.

Side Chain Conformations

- The side chains are also rotatable
- Impacts the shape of the protein (and interactions)
- And the number of possible shapes



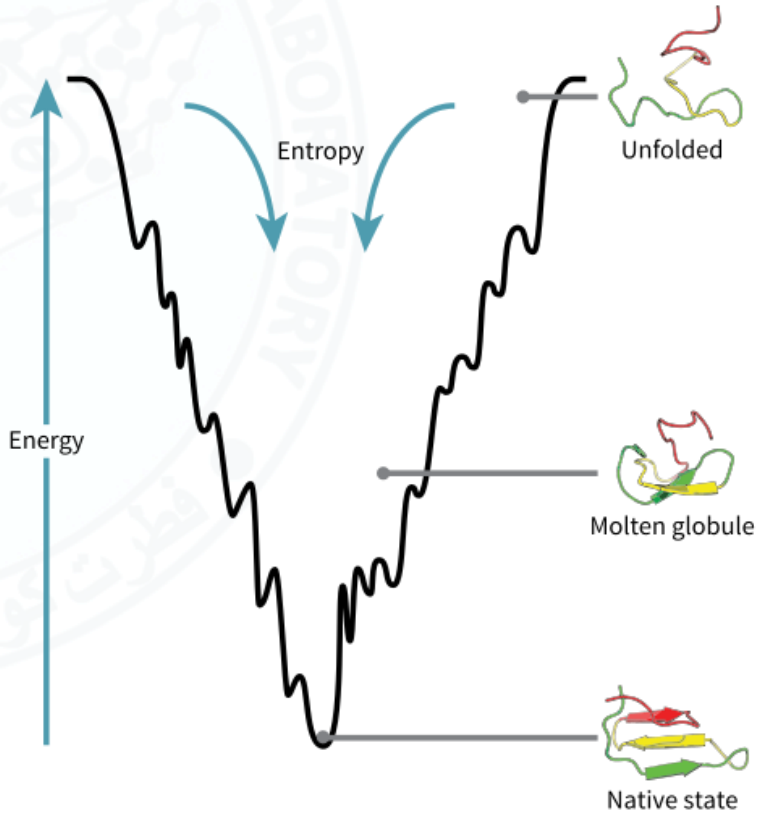


Interpretation of the Energy Function

- Abstractly, the energy function can be written as
  - E(S,$\mathbf{X}$;$\boldsymbol{\theta}$)
  - A mapping from the protein sequence and conformation spaces to its energy value
    - Parameters are denoted by $\boldsymbol{\theta}$
  - Given a sequence, the energy will be lowest for the native structure of the protein
- Physical interpretation
  - $\Delta \boldsymbol{G} = \Delta \boldsymbol{H} - \boldsymbol{T}\, \Delta \boldsymbol{S}$
  - Typically, the Entropy term is not explicitly modeled
  - A model of the energy of a protein

Physical Interpreation

- Energy Landscape of Protein Folding
  - The native state of the protein is the conformation of the protein with the lowest energy



http://en.wikipedia.org/wiki/Folding_funnel

Force Fields

- The energy function is a weighted summation of different contributing terms (electrostatics, VDW, …).
- These terms involve a set of parameters. The exact functional form of the energy function is called a “force field”.
- There are a number of different force fields available. For example:
  - AMBER, CHARMM, GROMOS, OPLS
- Example

$$V(r^N) = \sum_{\text{bonds}} k_b(l - l_0)^2 + \sum_{\text{angles}} k_a(\theta - \theta_0)^2$$

$$+ \sum_{\text{torsions}} \sum_{n} \frac{1}{2} V_n[1 + \cos(n\omega - \gamma)] + \sum_{j=1}^{N-1} \sum_{i=j+1}^{N} f_{ij} \left\{ \epsilon_{ij} \left[\left(\frac{r_{0ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{0ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right\}$$

https://en.wikipedia.org/wiki/AMBER

Points to Note

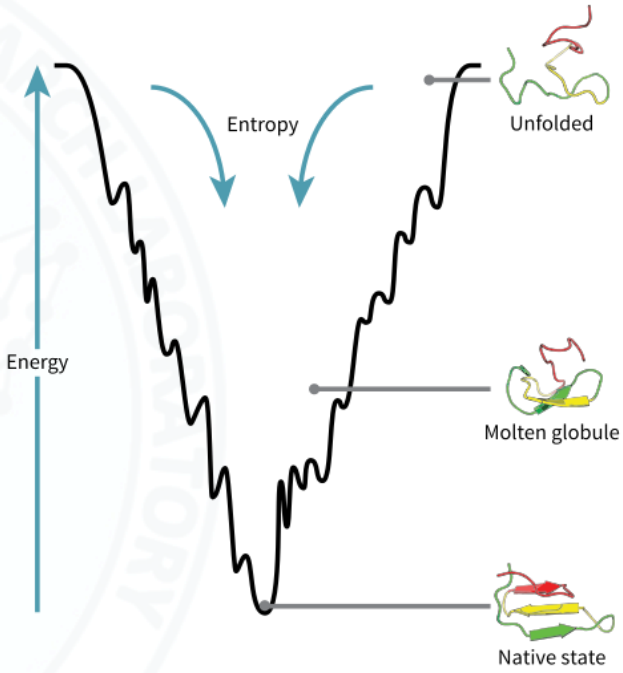
- This is only a model or an approximation.
  - Modeling molecular water
  - Empirical terms
  - Knowledge based
  - Quantum mechanical effects
    - http://www.sott.net/article/270402-Spooky-action-at-a-distance-Water-in-cells-behaves-in-complex-and-intricate-ways

- How to model what we cannot model?
  - Using machine learning aka statistical potentials
    - Given some training data, what are the chances of this particular conformation of the protein?
      - The higher the chances, the lower the energy and vice-versa

- Performing Energy Calculations
  - A number of software tools can calculate the energy of a protein
    - pyRosetta (http://www.pyrosetta.org/)
    - SHARPEN
      - https://www.engr.colostate.edu/~cdasnow/snowlab_software.shtml

What do these calculations help us do

- Protein Stability Calculations
  - Given: Structure(s) of a protein
  - Desired Output: An estimate of protein stability
    - The lower the energy of the protein the more stable it is. If two structures are given, we can calculate which one is likely to be more stable based on energy calculations
  - Computation
    - Simple Energy Calculations $E(S,\mathbf{X};\boldsymbol{\theta})$

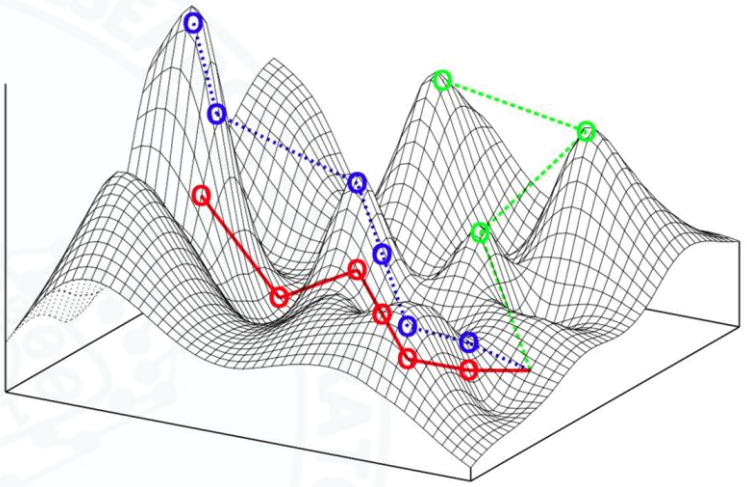


This is a consequence of the folding funnel model of the protein folding

What do these calculations help us do

- Protein Structure Prediction
  - Given: The unfolded structure/sequence of a protein
  - Desired output: The native structure of a protein
    - The native structure of the protein is the conformation of the protein at the lowest energy.
  - Thus, we are interested in solving the following optimization problem:

$$X^* = argmin_X \, \mathrm{E}(S, X; \theta)$$



Structure space

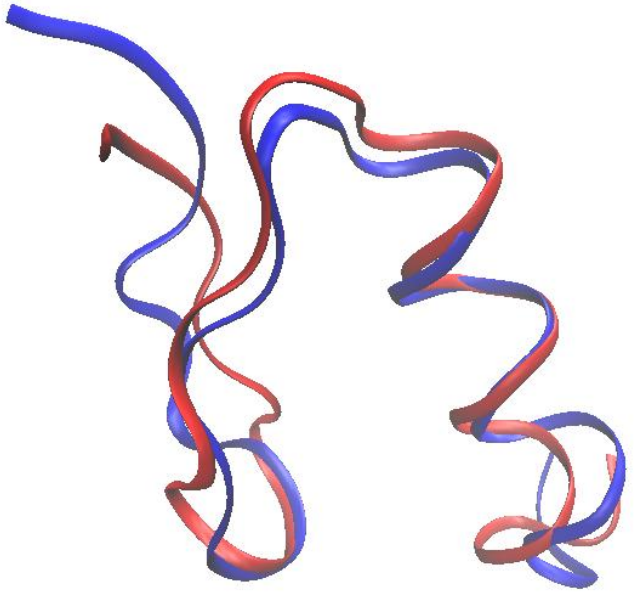
A change in, say, rotamer or torsion angles will cause the energy to change. This gives rise to the concept of a fitness landscape

https://en.wikipedia.org/wiki/Protein_structure_prediction

What do these calculations help us do

- Protein Design / Inverse Folding
  - Given: The desired structure of a protein ($X_{desired}$)
  - Desired output: A sequence of the protein whose native structure is as close to the desired structure as possible
  - Objective: Design a protein which has a certain function. For that function, we sculpt a structure of the backbone and possibly of some side-chains involved in the desired function. Now, we would like to have a sequence that generates the desired structure. We use computing to obtain that sequence and then develop the protein chemically.
  - Mathematically,

$$S^* = argmin_S \, \mathrm{E}(S, X_{desired}; \theta)$$



FSD-1 (shown in blue, PDB id: 1FSV) was the first de novo computational design of a full protein.[2] The target fold was that of the zinc finger in residues 33-60 of the structure of protein Zif268 (shown in red, PDB id: 1ZAA). The designed sequence had very little sequence identity with any known protein sequence.

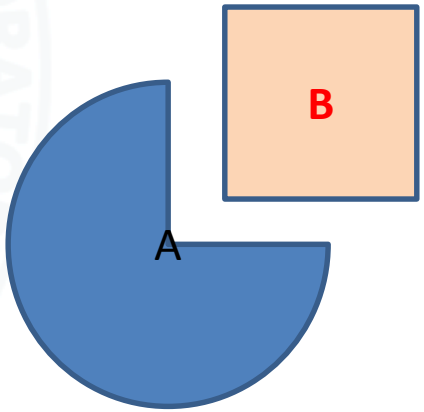
https://en.wikipedia.org/wiki/Protein_design

What do these calculations help us do

- Protein Interactions Studies
  - Given: Two proteins A and B
  - Output:
    - Do these proteins interact
    - What parts of the two proteins interact
  - Two proteins will interact only if this leads to a decrease in the energy
    - Energy before binding:
      $$\Delta \boldsymbol{G}_A + \Delta \boldsymbol{G}_B$$
    - Energy before binding
      $$\Delta \boldsymbol{G}_{AB}$$
    - The proteins will bind only if the binding free energy:
      $\Delta\Delta \boldsymbol{G} = \Delta \boldsymbol{G}_{AB} - (\Delta \boldsymbol{G}_A + \Delta \boldsymbol{G}_B) < 0$
    - Stability of the interaction

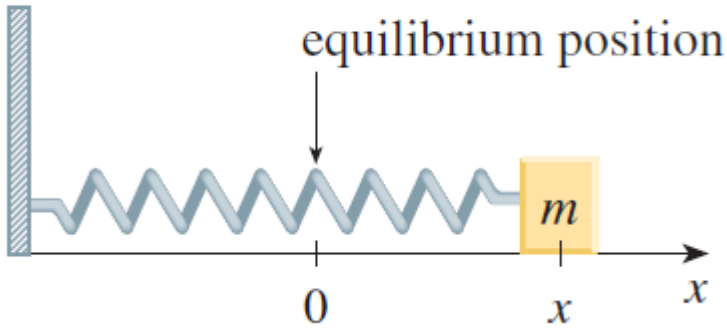
      Given by how negative the binding free energy is
    - Interaction Sites

      The two proteins will produce a joint conformation X_{AB} which produces lowest $\Delta \boldsymbol{G}_{AB}$



What do these calculations help us do

- Protein Dynamics
  - Given: Protein Structure
  - Desired output: How does this protein move or behave dynamically
  - If we know the energy function, the derivative of the energy function gives us the force. We can then use Newtonian mechanics to model the impact of that force on individual atoms to see how they move.



$$E = -kx^2$$

$$F = \frac{dE}{dx} = -kx$$

$$m\frac{d^2x}{dt^2} = -kx \quad \text{or} \quad m\frac{d^2x}{dt^2} + kx = 0$$

$$x(t) = A\cos(\omega t + \delta)$$

A simple example of dynamics simulation based on energy calculations.

https://en.wikipedia.org/wiki/Molecular_dynamics

Protein Dynamics

$$F_i = -\nabla_i E$$

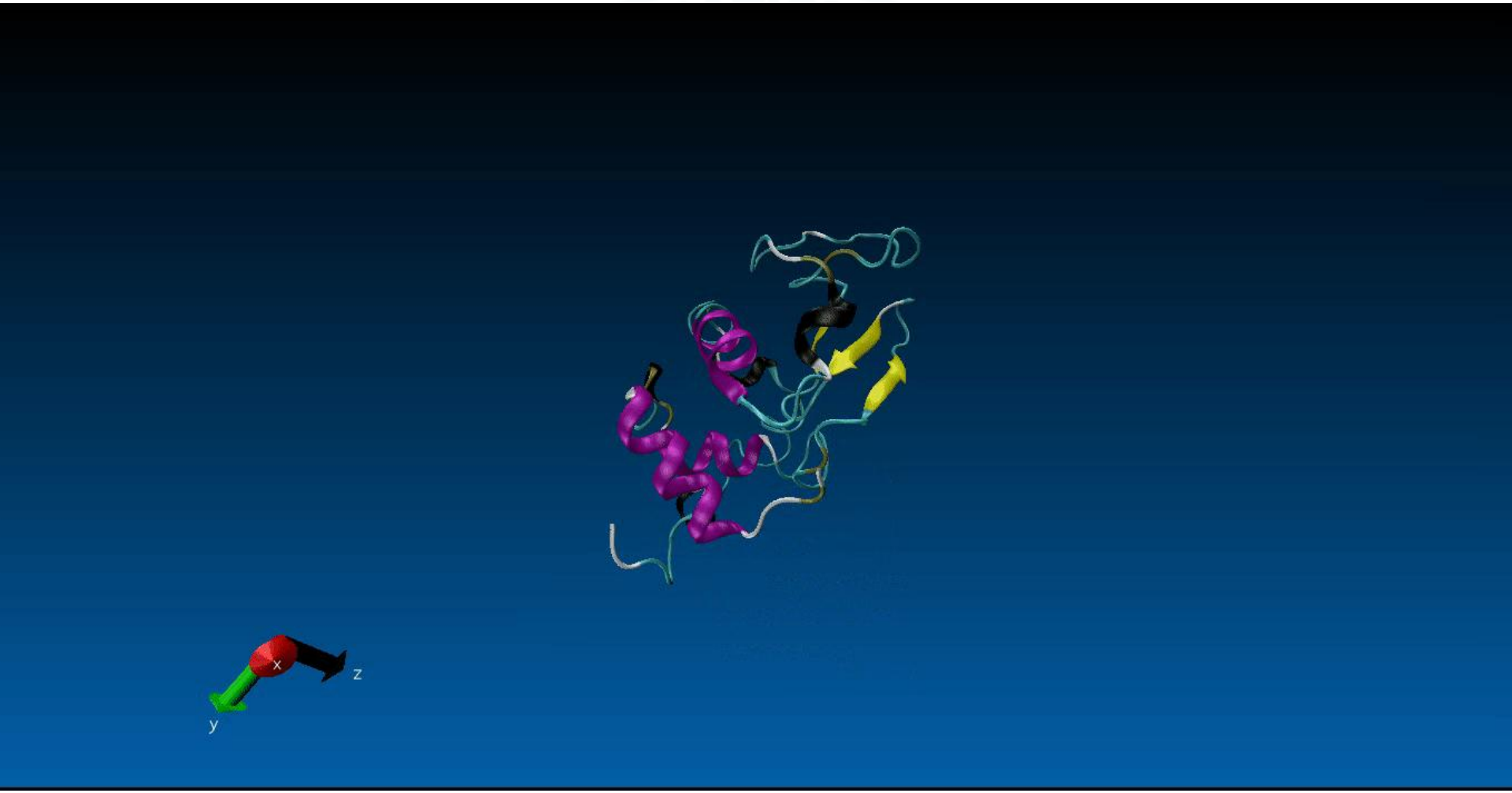
$$F_i = m_i a_i$$

$$-\frac{dE}{dr_i} = m_i \frac{d^2 r_i}{dt^2}$$

E is the potential as described by the force field

To calculate a trajectory of motion of an atom, one only needs the initial positions of the atoms, an initial distribution of velocities and the acceleration, which is determined by the gradient of the potential energy function. The equations of motion are deterministic, e.g., the positions and the velocities at time zero determine the positions and velocities at all other times, t. The initial positions can be obtained from experimental structures, such as the x-ray crystal structure of the protein or the solution structure determined by NMR spectroscopy.

10ns Molecular Dynamics Simulation of a protein (1AKL) with GROMACS



Protein Folding Simulation Video