Molecular Interactions and Biomolecular Structures; Computer Modeling

Main References: 1. Chapter 3 of van Holde 2. Second half of Chapter 9 of Tinoco (pages 493-516)

Copyright: Jianhan Chen

BIOCH 590: Biomacromolecules

Spring 2010

Hierarchical Organization of Proteins

Amino Acids

Peptides

Primary Sequence:

MTYKLILNGK TLKGETTTEA VDAATAEKVF KQYANDNGVD GEWTYDDATK TFTVTE

Protein G B1 (3gb1)

Myoglobin (1mbc)

tRNA transferase (1mxi)

Structure-Function Paradigm

Sequence

Folded 3D Structure

Functions

- catalysis
- transport
- signaling
- motors ...

• Structural Genomics (NMR, X-ray)
• Protein structure predictions

Intrinsic Disorder & Cellular Functions


• Intrinsically Disordered Proteins (IDPs): functional proteins that can exist as dynamic ensembles of disordered structures under physiological conditions.
• Over 30% of eukaryotic proteins are predicted to be disordered.
• Critical in cellular regulation and signal transduction.
• Allow high specificity with low affinity; structural plasticity for binding diversity
• Particular importance in cancer and protein misfolding diseases.
**Determinant of Structure (or Lack of It)**

- Probability of observing a particular structure (conformation) is determined by its stability (as defined by the free energy)
  - Thermodynamics and statistical mechanics!
- No single structure is the structure
  - It is all about probability (statistical mechanics!)
  - Motions and flexibility are important too
- The stability depends on a range of factors
  - Intramolecular interactions
    - Bonded: chemical bonds, angles, dihedrals etc
    - Nonbonded: “weak” interactions
      - Charged-charged, van der Waals (dispersion and repulsion)
  - Intermolecular interactions: nonbonded/weak interactions
    - Cellular environment: solvent (water), membrane, salt, pH etc
    - Association with other biomolecules, small molecules, ions, etc

---

**Quantum Mechanics vs. Molecular Mechanics**

- Quantum mechanics: “exact” and most applicable to understand chemical reactions
  - Separate nuclei and electrons
  - Too expensive, and not sufficiently accurate
  - Not relevant as many biological processes
- Molecular mechanics: classical mechanics at molecular level
  - Classical treatment of all atoms
  - No electron, no chemistry
  - Allows description of large molecules
  - Experimental methods available to determine the key parameters in a molecular mechanical treatment
- Hybrid QM/MM
  - QM for the active site (where reaction occurs) and MM for the rest
  - Accurate treatment of MM/QM Boundary is a problem

---

**Classical Mechanics**

- Total energy: \( E = K + V \)
  - Kinetic energy \( (K = \frac{mv^2}{2}) \), potential energy \( V \) (i.e., force field)
- Newton’s second law of motion: \( F = ma \)
  - Relation of force and potential energy: \( F = -\frac{\delta V}{\delta r} \)

---

**Molecular Potentials**

- Basic form: \( V = V_{\text{bonding}} + V_{\text{nonbonding}} \)
  - \( = (\sum V_{\text{bond}} + \sum V_{\text{angle}} + \sum V_{\text{dihed}}) + \sum (V_{\text{elec}} + V_{\text{vdw}}) \)
  - The potential energy is a function of all coordinates.
  - Additivity, empirical, transferability
Bonds and Angles

- $V_{\text{bond}} = k_{\text{bond}} (r - r_0)^2$
  - Harmonic approximation
  - OK for biomolecules
- $V_{\text{angle}} = k_{\text{angle}} (\theta - \theta_0)^2$

Dihedral Potentials

- $V_{\text{dihe}} = k_{\text{dihe}} \cdot [1 + \cos(n \phi - \delta)]$

Double Bonds: $n=2, \delta=-180$

- $V = k (1 + \cos(2 \phi - 180))$

Realistic Dihedral Potentials

- Actual dihedral potentials often have contributions with multiple periodicities
Electrostatic Interactions

- \( V_{\text{elec}} = \frac{q_1 q_2}{4 \pi \epsilon_0 r} \) Coulomb’s Law
  - \( \epsilon_0 \): permittivity constant of vacuum
- A simplified form: \( V_{\text{elec}} = 332 \frac{q_1 q_2}{r} \)
  - Where \( q \) is unit of electron charge, \( r \) is in Å and \( V \) in kcal/mol.
- Dielectric medium: \( V_{\text{elec}} = 332 \frac{q_1 q_2}{\epsilon r} \)
  - \( \epsilon \) is dielectric constant (relative permittivity).
  - \( \epsilon=78 \) for water under lab conditions (300K, 1atm)

\[
V \text{ or } \delta V/\delta r \text{ (kcal/mol)}
\]

\[
r \text{ (Å)}
\]

\[
\frac{1}{r}
\]

\[
\frac{1}{r^2}
\]

van der Waals Interactions

- London dispersion: attractive forces that arise from temporary dipoles (induced dipole-induced dipole interactions)
- van der Waals repulsion: all atoms repel at short distances
- A common function form: \( V_{\text{vdw}} = -A/r^6 + B/r^{12} \)
- Lennard-Jones potential function (12-6)

\[
V(r) = \epsilon \left[ \left( \frac{r_{\text{min}}}{r} \right)^{12} - 2 \left( \frac{r_{\text{min}}}{r} \right)^{6} \right]
\]

\[
\epsilon = 1, \ r_{\text{min}} = 1
\]

Dipole-Dipole Interactions

- Dipole moment: arise from charge separations
  - measure the “polarity” of a molecule (or fragment)
- Dipole-dipole interactions
  - Fully included if all charges treated explicitly
  - Offer simplifications (by reducing the number of terms)

\[
V_{\text{dd}} = \frac{\mu_A \cdot \mu_B}{r_{AB}^6} - \frac{2(\mu_A \cdot r_{AB})(\mu_B \cdot r_{AB})}{r_{AB}^{10}}
\]

- Decays faster: \( 1/r^3 \) dependence
- Special cases: parallel and perpendicular orientations

Lennard-Jones Potential

\[
V(r) = \epsilon \left[ \left( \frac{r_{\text{min}}}{r} \right)^{12} - 2 \left( \frac{r_{\text{min}}}{r} \right)^{6} \right]
\]
Hydrogen Bonds

- Very important in macromolecule structures
- Primarily a dipole-dipole interaction, but arguably with some covalent nature (electron sharing in so-called low barrier HBs)
- The strength of HBs vary greatly and depend on the environments (dielectric screening)
- The functional form for HB is unclear.
  - Often mimicked by Lennard-Jones potential
  - At present, often treated with electrostatic + vDW

(c) Jianhan Chen

What is a hydrogen bond worth?

<table>
<thead>
<tr>
<th>Secondary Structure</th>
<th>Stability per H-bond</th>
<th>Model</th>
<th>Reference State</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antiparallel β-sheet</td>
<td>-2.8</td>
<td>[Ac-ala-NHMe]</td>
<td>Infinite separation</td>
</tr>
<tr>
<td>Ala-gly Type II turn</td>
<td>-0.6</td>
<td>Ac-ala-gly-NHMe</td>
<td>Extended</td>
</tr>
<tr>
<td>Amide H-bond</td>
<td>-0.3</td>
<td>formamide</td>
<td>Infinite separation</td>
</tr>
<tr>
<td>1st helical H-bond</td>
<td>-0.2</td>
<td>Ac-(ala)-NHMe</td>
<td>Extended</td>
</tr>
<tr>
<td>2nd helical H-bond</td>
<td>-0.4/-1.0</td>
<td>Ac-(ala)-NHMe</td>
<td>extended</td>
</tr>
<tr>
<td>Ala-gly Type I turn</td>
<td>2.6</td>
<td>Ac-ala-gly-NHMe</td>
<td>extended</td>
</tr>
<tr>
<td>Pro-gly Type I turn</td>
<td>2.6</td>
<td>Ac-pro-gly-NHMe</td>
<td>extended</td>
</tr>
</tbody>
</table>

Tobias and Brooks CPL (1990)

β-sheets can have exceptional stability

- The property that nonpolar solutes aggregate in water
- Arise from a combination of elemental physical effects
  - Difference in strengths solute-water and water-water interactions
  - Difference in shapes (sizes) of solutes and water
  - Various entropic contributions
- One of the key driving forces for self-assembly in biology
  - Biological membrane, micelle formation, protein folding ...
  - Complex temperature dependence: cold denaturation of proteins
  - Very difficult to describe theoretically!

Small barrier and minimum associated with “naked” hydrogen bond, much more significant for beta-sheet model

Tobias and Brooks CPL (1990)

Hydrophobic Effects

(c) Jianhan Chen
A quantitative model of proteins need to be able to predict several kcal/mol differences in (free) energy!

This small energies result from a summation of many atom pairs (10^6-10^8)! Plus, entropic effects.

This is extremely difficult, if ever possible!
Two Excellent Readings

  - A good overview of current understanding of protein folding
  - The zipping and assembly hypothesis is interesting, but as a prediction method the success has been limited
  - One of the most readable and informative reviews on current understanding of hydrophobic effects

Basic Components of Modeling

Force Field
A set of basic model units and associated rules.

Sampling
The process of finding the optimal assembly of the basic model units.

Ethanol
http://www.sesame.org.jo/publication/NSLS.aspx
Molecular Models at Multiple Scales

increasing details and predicting power
increasing difficulty and computational cost

Lattice models
minimalist models
coarse-grained models
all-atom / simplified potentials
all-atom / molecular mechanics
quantum mechanics

provide the ultimate details necessary for understanding most biophysical processes.

CHARMM param22 Force Field

• Topology file: define the building blocks (atoms, connectivities)

atom types
residue blocks
atom compositions
connectivity

CHARMM param22 Force Field

• Parameter file: define the parameters of interactions

Bonds

\[
V_{\text{MM}} = \sum \frac{1}{2} k^b \cdot (b_i - b_j)^2 + \sum \frac{1}{2} k^\theta \cdot (\theta_i - \theta_j)^2 + \sum k^\phi \cdot [1 + \cos(n_i \phi_i - \delta_i)] + \sum \left[ \varepsilon_{\min} \left( \frac{r_{\min}^i}{r_{ij}} \right)^2 - 2 \frac{r_{\min}^i}{r_{ij}} + \frac{\varepsilon_{\min}^2}{r_{ij}} \right] + \sum q_i q_j \varepsilon_{ij} \]

Molecular Dynamics (MD)

\[m_i \ddot{r}_i = F_i = -\nabla V_i\]

Monte Carlo (MC)

\[P(\ddot{r}) = \exp(-\Delta V/kT)\]
Parameterization of Force Fields

- Bonded terms: spectroscopy or quantum mechanics
- Lennard-Jones: Small molecular crystals
- Electrostatic: quantum mechanics (fit monopoles to electrostatic potential)
- Many challenges in practice
  - which (model) molecules: availability, representative or not
  - how many atomic classes: transferability and tractability
  - Which properties to parameterize for?
  - Correlation of parameters
  - Higher order/new terms or not?
  - Electron polarization, non-additivity, ...
  - water, water, and water
- At the end, do they add up? (cancellation of errors)

Molecular Dynamics

- Objective: \{r_i(t), ..., r_N(t)\} → \{r_i(t+\Delta t), ..., r_N(t+\Delta t)\} \quad \text{f = ma}
- Basic idea: solve Newton’s equation of motion numerically
  - Given current coordinates (x), velocities (v)
    - Forces can be calculated based on coordinates \( f = -\text{d}V/\text{d}x \)
    - \( x(t+\Delta t) = x(t) + v(t) \Delta t \)
    - \( v(t+\Delta t) = v(t) + f(t)/m \Delta t \)
    - Repeat above operations
- More accurate integrators (better energy conservation)
  - Verlet Algorithm (Verlet J. Chem. Phys. 1967)
    - Consider Taylor’s expansions:
      \[
      x(t+\Delta t) = x(t) + v(t) \Delta t + 1/2 m f(t) \Delta t^2 + 1/6 d^3 x/dt^3 \Delta t^3 + O(\Delta t^4)
      \]
    - Adding expansion \( x(t+\Delta t) \) and \( x(t-\Delta t) \) and rearrange:
      \[
      x(t+\Delta t) = 2x(t) - x(t-\Delta t) + f(t)/m \Delta t^2 + O(\Delta t^4)
      \]
    - Subtracting expansion \( x(t+\Delta t) \) and \( x(t-\Delta t) \) and rearrange:
      \[
      v(t) = [x(t+\Delta t) - x(t-\Delta t)]/(2\Delta t) + O(\Delta t^2)
      \]

Energy Minimization

- Minimization follows gradient of potential to identify stable points on energy surface
  - Let \( V(x) = k(x-x_0)^2 \)
  - Begin at \( x' \), how do we find \( x_0 \) if we don’t know \( V(x) \) in detail?
    - How can we move from \( x' \) to \( x_0 \)?
  - Steepest descent (SD):
    - \( x' \rightarrow x = x + \delta \)
    - \( \delta = -\text{d}V/\text{d}x = -dx \quad k(x-x_0) \)
  - This moves us, depending on the step size \( \Delta x \), toward \( x_0 \).
  - On a simple harmonic surface, we will reach the minimum, \( x_0 \), i.e. converge, in a certain number of steps related to \( \Delta x \).

Solvent and Periodic Boundary Conditions

- Vacuum
  - Surface effects (surface tension)
  - No dielectric screening
- Droplets
  - Still surface effects (at water – vacuum interface)
  - Only partial dielectric screening
  - Evaporation of the solvent
- Periodic: system is surrounded by copies of itself
  - Advantage:
    - No surface effects
  - Disadvantage:
    - Artificial periodicity
    - High effective concentration

\[ \text{van Gunsteren Angew Chem Int Ed (2006)} \]
Controlling Thermodynamic Variables

- MD generate statistical ensembles that connect microscopic details to macroscopic/thermodynamic properties
- NVE (microcanonical - Entropy rules!)
- NVT (Canonical - Helmholtz free energy is relevant, A)
  - temperature \( T = \frac{\Sigma m<v^2>}{(3k_B)} \)
- NPT (Isothermal-isobaric - Gibbs free energy is relevant, G)
  - \( P \text{ kinetic + virial contributions} \)
- Thermostats, barostats, etc., allow one to choose appropriate ensembles
  - Following Nose’, Hoover, Evans and others...

Why is MD so slow?

<table>
<thead>
<tr>
<th>Simulated Time</th>
<th>1 ns (10^{-9} s) (500,000 MD steps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPU Time</td>
<td>~200 hours (10^6 s)</td>
</tr>
<tr>
<td>Wall Time</td>
<td>~1 days (10^5 s) / 8 CPUs</td>
</tr>
</tbody>
</table>

- very small time step required
  - \( \Delta t \sim 10^{-15} \text{ s} \)
- interactions between thousands of atoms need to be computed

Channel-forming peptides in a fully solvated membrane bilayer; Channel: 1795 atoms; All: 26254 atoms

Basic Flow of a MD Simulation

PSF: a computer model or representation of the protein given the force field, including all atoms, their connectivities and how they interact, etc.

Biological Time Scale

- Bond vibrations 1 fs (10^{-15} s)
- Sugar repuckering 1 ps (10^{-12} s)
- DNA bending 1 ns (10^{-9} s)
- Domain movement 1 ms (10^{-6} s)
- Base pair opening 1 ms (10^{-3} s)
- Transcription 2.5 ms / nucleotide
- Protein synthesis 6.5 ms / amino acid
- Protein folding ~ 10 s (speed limit: \( \mu \text{ s} \))
- RNA lifetime ~ 300 s

Simulation time should exceed the time scale of interest by ~10-fold!
**Gap in Timescales**

<table>
<thead>
<tr>
<th>Simulation Timescales</th>
<th>Biological Timescales</th>
</tr>
</thead>
<tbody>
<tr>
<td>s</td>
<td>10^6 s</td>
</tr>
<tr>
<td>ms</td>
<td>10^3 s</td>
</tr>
<tr>
<td>µs</td>
<td>10^6 s</td>
</tr>
<tr>
<td>ns</td>
<td>10^9 s</td>
</tr>
<tr>
<td>ps</td>
<td>10^-12 s</td>
</tr>
<tr>
<td>fs</td>
<td>10^-16 s</td>
</tr>
</tbody>
</table>

**Practical Considerations**

- **Long-range forces**
  - Using cut-off to reduce the number of nonbonded atom pairs (~12-15Å)
  - Electrostatic decays slowing (1/r) and cut off does not work well; Particle Mesh Ewald (PME) is needed.

- **Parallel execution**
  - Partition various regions of the system to different CPUs
  - Need to communicate information between nodes; this is a bottleneck

- **Simplifications of the model**
- **Enhanced sampling techniques**

**Implicit Solvent**

- Solvent increases the system size about 10-fold
- It is possible to describe the mean influence of water w/o explicitly including water

**Explicit solvent**
- Protein: 56 residues (855 atoms)
- Solvent: 5411 waters (16233 atoms)

**Implicit solvent**
- Hybrid macroscopic (solvent) / microscopic (solute)

**Coarse-Grained Models**

- Rely on reduced representation and/or simplified interaction schemes to access larger length and time scales

**Biomembrane sculpting by protein-BAR domains**

The simulation shown in the figure was carried out using a box with dimensions 100 x 16 x 50 nm and would correspond to a system of 10 million atoms. Using a shape-based CG model reduces the size to 3265 CG beads. The simulation showed that a concerted action of BAR domains arranged in a lattice results in the development of a global membrane curvature on a time scale of several µs, with the resulting curvature radius of ~30 nm that was observed experimentally.
Barriers, Temperature and Timescales

\[ \tau = \tau_0 \exp(\Delta G^\ddagger / kT) \]

\( \tau_0 \sim 10^{-12} \text{ s} \sim \text{ps} \)

\( T = 300 \text{ K} \)

\( \Delta G^\ddagger \): 1 kcal/mol, \( \tau \sim \text{ps} \)

5 kcal/mol, \( \tau \sim \text{ns} \)

10 kcal/mol, \( \tau \sim \mu \text{s} \)

Protein energy landscape is highly complex and rugged with numerous local minima.

Applications of Modeling

- Main advantages
  - Offer atomistic spatial resolution and femtosecond time resolution
  - Allow probing the system in many nontrivial ways that are not possible or too dangerous experimentally
  - Often much cheaper than doing the experiment itself
  - Can be applied at very large scales (computers are cheap)
  - Can provide theoretical frameworks for experimental studies

- A few prototypical areas
  - Protein structure prediction and calculation
  - Virtual screening and rational drug design
  - Simulation of important systems: mechanisms
  - Interpretation of (static) experimental data
  - Protein misfolding and aggregation
  - Biomolecular engineering: design of new enzymes etc
  - ...

Enhanced Sampling Techniques

Replica Exchange (REX)

Exchange criteria

\[ P_{\text{ex}} = \begin{cases} 1 & \Delta \leq 0 \\ \exp(-\Delta) & \Delta > 0 \end{cases} \]

\[ \Delta = (E_i - E_j) \cdot (1/kT_i - 1/kT_j) \]

Protein Energy Surface

Sugita and Okamoto, CPL (1999); MMTSB Tool Set: http://mmtsb.scripps.edu

NMR Structure Refinement

Initial Model from CNS

REX/GB Refinement

Refined Model

PDB: 1XJH

1 day later

1 month later

Chen et. al., JACS (2004); Chen et al., J. Biomol NMR (2004).
Functional Motions of Ribosome

“This movie depicts a ratchet-like rearrangement of the 70S ribosome. The rotation of the 30S ribosomal subunit relative to the 50S subunit shows high correspondence to motion captured in cryo-EM maps of the ribosome and postulated to be a key mechanical step in the translocation of the mRNA•tRNAs complex. “

http://brooks.chem.lsa.umich.edu/

Functional reorganization of the ribosome explored by theory and experiment

Florence Tama  Mikel Valle
Joachim Frank  Charles L. Brooks III

The Skirball Research Institute
Weinberg Center


What do we need to consider?

see also, Bionanotechnology D.S. Goodsell 2004 Wiley