Introduction to Biological Macromolecules

General Principles and Key Concepts

Copyright: Jianhan Chen

Course Overview

• Scope: physical properties of biological macromolecules
  – Basic principles of structure at multiple levels
  – Focus on the quantitative aspects
  – Theoretical frameworks: thermodynamics, statistical mechanics
  – Physical biochemistry: molecular interactions and interactions of macromolecules with environments
  – Experimental techniques to measure certain properties, which can be then used to infer about the structural organizations

• For this course
  – Emphasize a qualitative understanding of the quantitative nature
  – Build an appreciation of structural complexity and principles of theoretical and experimental studies of molecular structures
  – In Part II, a survey of basic and practical principles of common spectroscopic techniques for studying molecular structures

General Introduction

• Course setup
  – Two parts: (I) physical biochemistry and (II) spectroscopy
  – Part I: from Jan 15 to Mar 6, by Jianhan Chen

• Lecture notes will be main study materials

• Textbooks:
  – two recommended
  – not required, but very useful resources for additional reading
  – Available from the Hale Library Reserve

• Two exams from Part I
  – Feb 9 and Mar 6 during lecture time (Mark these dates!)

• Grading
  – Both parts contribute equally
  – Part I: ~ 70% from exams and ~30% from homeworks

Tentative Topics

• Overview of macromolecular structure (1-2 lectures)
• Thermodynamics (4-5 lectures)
• Statistical Mechanics (3-4 lectures)
• Molecular interactions (3 lectures)
• Molecular motion and transport (3 lectures)
• Computer modeling (3 lectures)
• Inter-molecular interactions (3 lectures)
Biological Macromolecules

- Atom: smallest unit of an element (C, N, O, H etc)
- Molecule: smallest unit of an composite (H₂O, CO₂, C₆H₆ etc)
- Macromolecule: a large molecule
  - Smallest molecule: H₂
  - Largest molecules: DNAs (tens of billions of atoms!)
- Biological macromolecule: a large and complex molecule with specific biological functions
  - Typically biopolymers: monomers, oligomers
  - Hierarchical structures
    - Primary (1°): sequence of monomers (sugar, amino acid, nucleic acid, etc)
    - Secondary (2°): local regular structures (alpha helix, etc)
    - Tertiary (3°): global 3D fold
    - Quaternary (4°): organization of multiple subunits (polymers) in a functional complex

Stoichiometry and Geometry

- Stoichiometry: atomic composition
  - For biological macromolecules: compositions of (monomer) units
  - Relative easy to know
- Geometry: three-dimensional structure
  - More challenging to determine, especially for large, flexible ones
  - Protein folding problem

**ethanol**
a.k.a. alcohol, biofuel

\[ C₂H₆O \]

**hemoglobin**

\[ \alpha₂β₂ \]
Configuration and Conformation

- Configuration: arrange of atoms around non-rotating bonds or around chiral centers
- Conformation: spatial arrange around rotatable bonds

![](image1)

(a) The molecule 1,2-dilutrosene can rapidly interconvert between the two orientations shown, because a carbon-carbon single bond is not structurally rigid.

(b) The molecule 1,2-dilutrosene cannot interconvert between the two forms shown, because a carbon-carbon double bond is structurally rigid.

Chirality in Biology

- Most biomacromolecules consist of chiral monomers
- In biology, typically only one isomer is active
  - Virtually all sugars are D isomers
  - Most proteins are made of L amino acids

![](image2)

L- and D-alanine
D- and L-erythrose

Conformation/Struture

- Mainly defined by torsion angles
  - Also referred to as dihedral angle: the angle between two groups on either side of a freely rotating chemical bond
  - Defined by four consecutively linked atoms: $\psi (A-B-C-D)$

\[
\begin{align*}
A &\quad B \quad C \quad D \\
\text{(C lies directly under B)} &\quad \alpha &\quad \beta
\end{align*}
\]

- 48° − torsion angle

- Right-hand rule: positive sense is clockwise

\[
\begin{align*}
0^\circ &\quad 180^\circ &\quad 90^\circ
\end{align*}
\]

Determinant of Structure

- Probability of observing a particular structure (conformation) is determined by its stability (as defined by the free energy)
  - Thermodynamics and statistical mechanics!
- 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
- No single structure is the structure
  - It is all about probability (statistical mechanics!)
  - Motions and flexibility are important too
Water

- Solvent of life
- Many unique properties
  - Maximum density at 4 °C
    - Ice is lighter than liquid water
  - Polar molecule
    - hydrogen bonding network
    - High specific heat capacity
- Hydrophobicity and hydrophilicity
  - Solute polarity (carry partial charges or not)
  - Salts (e.g. NaCl) dissolve in water readily
  - Hydrocarbons (oil) do not mix with water
- Amphipathic molecules
  - Self-assembly to micelles, biological membranes

Natural Amino Acids

- All $\alpha$-amino acids in L-configuration
- Essential vs non-essential ones
- Co-exist in two forms
  - Ionic (zwitterionic form) and unionized
- Side chains vary

Hierarchical Organization of Proteins

Amino Acids

- Primary Sequence: MTKLILNGK TLKGETTTEA VDAATAEKVF KQYANDNGVD GEWTYDDATK TFTVTE

Peptides

- Protein G B1 (3gb1)
- Myoglobin (1mbc)
- tRNA transferase (1mxi)

Nonpolar, aliphatic R groups

- Glycine
- Alanine
- Valine

Aromatic R groups

- Phenylalanine
- Tyrosine
- Tryptophan

Polar, uncharged R groups

- Serine
- Threonine
- Cysteine

Positively charged R groups

- Lysine
- Arginine
- Histidine
Foldable Sequence Space is Limited

- Huge number of possible sequences: $20^N$
  - Only a small subset are explored in biology and most of them give rise to specific 3D structures under the proper conditions
    - Sequence determines structure (or lack of structure)
    - Protein structure-function paradigm: enzymes
    - Some sequences are “intrinsically disordered”! They play very important roles in signaling and regulation.

Important Torsions of Proteins

- Backbone torsions
  - $\phi$: C(O)-N-Ca-C(O)
  - $\psi$: N-Ca-C(O)-N
  - $\omega$: H-N-C(O)-O
    - Mostly 180° (trans)
- Ramachandran plot
  - 2D plot of ($\phi$, $\psi$) distribution
  - Allow assignment of secondary structures
  - Statistics of known protein structures reveal common regions of ($\phi$, $\psi$) distributions
  - An important way of structural evaluation

Three-Letter and Single-Letter Codes

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>3-Letter</th>
<th>1-Letter</th>
<th>Amino Acid</th>
<th>3-Letter</th>
<th>1-Letter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alanine</td>
<td>Ala</td>
<td>A</td>
<td>Leucine</td>
<td>Leu</td>
<td>L</td>
</tr>
<tr>
<td>Arginine</td>
<td>Arg</td>
<td>R</td>
<td>Lysine</td>
<td>Lys</td>
<td>K</td>
</tr>
<tr>
<td>Asparagine</td>
<td>Asn</td>
<td>N</td>
<td>Methionine</td>
<td>Met</td>
<td>M</td>
</tr>
<tr>
<td>Aspartate</td>
<td>Asp</td>
<td>D</td>
<td>Phenylalanine</td>
<td>Phe</td>
<td>F</td>
</tr>
<tr>
<td>Cysteine</td>
<td>Cys</td>
<td>C</td>
<td>Proline</td>
<td>Pro</td>
<td>P</td>
</tr>
<tr>
<td>Histidine</td>
<td>His</td>
<td>H</td>
<td>Serine</td>
<td>Ser</td>
<td>S</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>Ile</td>
<td>I</td>
<td>Threonine</td>
<td>Thr</td>
<td>T</td>
</tr>
<tr>
<td>Glutamate</td>
<td>Glu</td>
<td>E</td>
<td>Tyrosine</td>
<td>Tyr</td>
<td>Y</td>
</tr>
<tr>
<td>Glutamate</td>
<td>Glu</td>
<td>E</td>
<td>Valine</td>
<td>Val</td>
<td>V</td>
</tr>
</tbody>
</table>
Secondary Structures of Proteins

- α-helix: $(i, i+4)$ hydrogen bonds, 3.6 residues/1.8 Å per turn

Tertiary and Quaternary Structures

- A simple way of representing protein folds

Contact Map

Secondary Structures of Proteins

- β-strands: trans- backbone torsions, multiple hydrogen bonds
Nuclei Acids

- Polymer of nucleotides: highly flexible (compared to peptides)
- Deoxyribonucleic acid (DNA): A/G/C/T
- Ribonucleic acid (RNA): A/G/C/U

Watson-Crick Base Pairs

- G-C and A-T(U)
- DNA almost exclusively exist as duplex held together by forming Watson-Crick base pairs
- A robust mechanism for faithful replication and translation
- Other pairings possible but mostly in RNA to give rise to non-trivial structures

Helical Structures of DNA (RNA)

- Famous DNA double helix: stacking of base pairs, minor/major grooves, polymorphic (hydration dependent)

Tertiary Structure of RNA

- Single-strand RNA often fold into tertiary structures

(c) Jianhan Chen
Ribosome: the Protein Factory

- Prokaryotic ribosome consists of 42 proteins (~35% weight) and 3 ribosomal RNAs (rRNA) (~65% weight) and measures about 200 Å across.

Proteins shown in blue and RNA strands in orange and yellow.

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 Scripps Research Institute, La Jolla, CA

Homework Assignment 1

- 1. hand draw the backbone conformation of Alanine Dipeptide in (1) right-hand helx (phi,psi)=(-60,-60) and (2) left-hand helix (+60, 60)
- 2. hand draw a schematic contact map of the following beta-sheet protein fold.

(c) Jianhan Chen