Where does Physical Biochemistry/Biophysics fit into ‘molecular biology’?

Central Dogma of Molecular Biology:

DNA → RNA → PROTEIN

- **Structural** (e.g. ribosome, spliceosome)
- **Ribozymes**
- **Other:** e.g. siRNA
- **Metabolic:** Polysaccharides, Lipids, Other molecules (e.g. ATP)
- **Enzymes**
- **Non-enzymes**
  - **Structural**
  - **Functional** (e.g. DNA-binding transcription factors)

Gene: Unit of heritable information
Codes for protein(s) or RNA(s)

**GENOME SEQUENCING**

<table>
<thead>
<tr>
<th>Year</th>
<th>Genome</th>
<th>Size (base pairs, bp)</th>
<th>Number of genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1982</td>
<td>Bacteriophage λ</td>
<td>48,500</td>
<td>71</td>
</tr>
<tr>
<td>1986</td>
<td>HIV virus</td>
<td>~10,000</td>
<td>9</td>
</tr>
<tr>
<td>1997</td>
<td><em>Escherichia coli</em> K12</td>
<td>4.6 Mbp</td>
<td>4,288</td>
</tr>
<tr>
<td>1997</td>
<td><em>Saccharomyces cerevisiae</em></td>
<td>12 Mbp</td>
<td>6,000</td>
</tr>
<tr>
<td>2002</td>
<td><em>Schizosaccharomyces pombe</em></td>
<td>13.8 Mbp</td>
<td>4,824</td>
</tr>
<tr>
<td>1998</td>
<td><em>Caenorhabditis elegans</em></td>
<td>97 Mbp</td>
<td>19,000</td>
</tr>
<tr>
<td>2000</td>
<td><em>Arabidopsis thaliana</em></td>
<td>125 Mbp</td>
<td>25,498</td>
</tr>
<tr>
<td>2000</td>
<td><em>Drosophila melanogaster</em></td>
<td>180 Mbp</td>
<td>13,600</td>
</tr>
<tr>
<td>2001</td>
<td><em>Homo sapiens</em></td>
<td>3,000 Mbp</td>
<td>30 – 40,000</td>
</tr>
</tbody>
</table>
What does a simple gene look like?

ATG

TAA

enhancer

promoter

transcription

“AUG

translation (genetic code)

UAA

AAAAAA

term.

mRNA

Protein

MAKSDYPILL............STVAGPQKQ*

Further levels of diversity of protein product from a single ‘gene’:

genomic

A

mRNA splicing

B

C

D

Protein isoforms

A

B

C

D

Post-translational modifications:

e.g. phosphorylation, acetylation, ubiquitination, glycosylation...

proteolysis

Greatly expands the total number of proteins possible from a single genome
PROTEOME (PROTEOMICS)

Proteome: Complement of expressed proteins produced by an organism or tissue/cell/organelle...

Within an organism:

Different cell type, different proteome.

Some proteins common to more than one cell type

Others specific to one cell type

Proteome can change during development/differentiation and in response to stimuli

2D-GEL ELECTROPHORESIS OF HUMAN HEART PROTEINS

Ventricle

Atrium

SDS-PAGE

IEF

Proteins found in atrium not ventricle

Need to identify and characterize molecules (mass spectrometry)
How do molecules work?

Anfinsen expt. (1955): Ribonuclease

Native fold
Active

Denature

Unfolded (denatured)
Inactive

Renature

Native fold
Active

1) Molecular sequence sufficient to determine 3D-fold (structure)
2) Activity (interactions) depends on correctly folded structure

Every cellular process depends on molecular interactions:
- Enzyme-substrate interactions
- Receptor-ligand interactions
- Protein-nucleic acid interactions: transcription, translation...
- Protein-protein: e.g. actin-myosin...

Permanent interactions (obligate)
- stable interaction (strong binding energy)

Temporary/transient interactions (non-obligate)
- weaker, reversible association (weaker binding energy)
- molecules exist as complex or as free components

Molecular structure can change upon association/dissociation and play a central role in functional output of interaction
What are the underlying driving forces for molecular structure and interactions?

Weak, interatomic interactions:

i. Electrostatic interactions  
ii. Van der Waal’s interactions (dipole moments)  
iii. Hydrogen bonds

Directional component to magnitude - so shape is important!

Stability of folded structure depends partly on interplay of 100s - 1000s of weak interatomic forces (contribute to favorable enthalpy of folding)

Stability of complex depends on interplay of 10-100s of weak interatomic forces (contribute to favorable enthalpy of interaction) and shape of interacting surfaces (stericchemical complementarity):
Can we predict protein structure and function?

Similar sequence → Similar structure? → Similar interactions?? → Similar function???

Generally, for two protein sequences:
If greater than ~30% sequence identity will have similar structure

Families/superfamilies of proteins/structural folds that share similar sequence/structure - often have similar biochemical properties/function

Query sequence → BLAST → sequence database matches

Databases:
Pfam (pfam.wustl.edu) - protein sequence families
Protein Data Bank (www.rcsb.org) - molecular structures
SCOP (http://scop.berkeley.edu/) - structural folds classification

Can we predict protein structure and function?

Similar sequence → Similar structure? → Similar interactions?? → Similar function???

Generally, for two protein sequences:
If greater than ~30% sequence identity will have similar structure

Families/superfamilies of proteins/structural folds that share similar sequence/structure - often have similar biochemical properties/function

Query sequence → BLAST → sequence database matches

Databases:
Pfam (pfam.wustl.edu) - protein sequence families
Protein Data Bank (www.rcsb.org) - molecular structures
SCOP (http://scop.berkeley.edu/) - structural folds classification
GH5 vs. CAP
13.4% sequence identity

GH5 vs. HNF3
12.5% sequence identity

TATA-Box Binding Protein (TBP)

Human
40% sequence identity
75% sequence similarity

Extreme thermophile
(Pyrococcus furiosus)
What does physical biochemistry/biophysics seek to do?

What do biophysics techniques allow us to do?

1. Identify and characterize macromolecules

   Mass spectrometry

2. Identify, characterize and **quantify** molecular interactions

   \[ K_d = \frac{[P]_f [A]_f^n}{[PA_n]} = \frac{k_{off}}{k_{on}} \]
   \[ \Delta G = \Delta H - T \Delta S = RT \ln K_d \]

3. Determine atomic resolution structures of macromolecules and macromolecular complexes

   X-ray Crystallography
   NMR Spectroscopy
   Cryo-electron microscopy
CRITICAL PARAMETERS DESCRIBING MOLECULAR INTERACTIONS

Unbound or free components (the interacting species)

Complex (bound state)

MOLECULAR INTERACTION PARAMETERS

1. STOICHIOMETRY composition of complex composition of interacting species

2. AFFINITY strength of interaction(s)

3. CONCENTRATION OF INTERACTING SPECIES

4. ENERGY $\Delta H, \Delta S, \Delta G, \Delta C$

   “thermodynamic signature”

5. INFLUENCE OF SOLUTION CONDITIONS temperature, pH, ionic strength ...

6. INFLUENCE OF COMPETING FACTORS specificity, regulation

7. KINETICS rate(s) of association/dissociation

8. CELLULAR LOCALIZATION

   \textit{In vitro} and \textit{in vivo}
What factors can influence degree of complex vs. free components at any instant in time?

Why does $P$ interact with $A$ but not $B$, $C$, $D$, ....?
**Fraction complex (PA)**

\[
K_d = \frac{[P][A]}{[PA]}
\]

Units: M

**Concentration of components**

\[
\text{Fraction complex} = \frac{(K_d + [P]_T + [A]_T) - \sqrt{(K_d + [P]_T + [A]_T)^2 - 4[P]_T[A]_T}}{2[P]_T}
\]

\[
[P]_T = [P]_F + [PA]
\]

\[
[P]_T = [A]_F + [PA]
\]

\([P]_T \leq 0.1 K_d\)

[Diagram showing the relationship between fraction complex and concentration of components]

- Never interact
- Interaction can be regulated
- Permanent complex
a) Hormone circulating in blood

\[ K_d = 1 \times 10^{-6} \, \text{M?} \]

If signal \( \propto \% \) complex
then if \( K_d = 10^{-6} \, \text{M} \), will need \([\text{hormone}] \geq 10^{-6} \, \text{M}\)

In reality \( K_d \) lower than this (< \( nM \))

\[ \text{Membrane} \]

Hormone Receptor

Signals to cell

b) Hemoglobin (\( \alpha_2\beta_2 \) heterotetramer)

\[ K_d = 1 \times 10^{-6} \, \text{M} \]

\[ \alpha \beta \text{ heterodimers} \]

binds and releases \( \text{O}_2 \) efficiently

But \([\text{Hemoglobin}]_{\text{blood cell}} > 1 \times 10^{-3} \, \text{M (mM)}\) so mostly \( \alpha_2\beta_2 \)
Important contributions of selected biophysical methods for identifying and characterizing molecules and molecular interactions
(i) Absorption or Excitation: Molecule absorbs photon of energy equal to difference between ground and excited states ($\Delta E_1, \lambda_1$)

(ii) Molecule loses some energy via non-radiative transitions (transitions between vibrational levels)

(iii) In a fluorophore, due to more rigid, inflexible molecular structure, vibrational levels of excited state do not overlap with ground state

Emission: Molecule returns to ground state vibrational level via RADIATIVE TRANSITION, emitting a photon of energy ($\Delta E_2, \lambda_2$)

(iv) Molecule can lose remaining energy via non-radiative transitions to return to ground state

Excitation energy $\Delta E_1 >$ Fluorescence or Emission Energy $\Delta E_2$

therefore excitation wavelength $\lambda_1 <$ emission wavelength $\lambda_2$
A SPECTROFLUORIMETER

Fluorescence parameters:

1) Fluorescence lifetime (excited state lifetime) - $\tau$
   
   length of time from absorption to emission of photon (1 - 10 ns)

2) Quantum yield, $Q = \frac{\text{number of photons emitted}}{\text{number of photons absorbed}} \leq 1.0$

Factors affecting fluorescence:

1) Molecular structure (electronic and non-electronic transitions)
2) Quenching: Internal (due to structure)
   External (interactions with other molecules)

Static Quenching

Collisional/Dynamic Quenching

Both involve NON-RADIATIVE loss of energy from excited fluor
a) Reduces Quantum yield
b) Changes Absorption wavelength
c) Changes Emission wavelength
FLUOROPHORES

<table>
<thead>
<tr>
<th>Fluorophore</th>
<th>$\lambda_{\text{exc}}$ (nm)</th>
<th>$\lambda_{\text{em}}$ (nm)</th>
<th>Q</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tyrosine</td>
<td>274</td>
<td>303</td>
<td>0.14</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>280</td>
<td>348</td>
<td>0.2</td>
</tr>
<tr>
<td>Fluorescein</td>
<td>494</td>
<td>518</td>
<td>~0.9</td>
</tr>
</tbody>
</table>

GREEN FLUORESCENT PROTEIN

<table>
<thead>
<tr>
<th>Protein</th>
<th>$\lambda_{\text{exc}}$ (nm)</th>
<th>$\lambda_{\text{em}}$ (nm)</th>
<th>Q</th>
</tr>
</thead>
<tbody>
<tr>
<td>GFP</td>
<td>395</td>
<td>510</td>
<td>0.79</td>
</tr>
<tr>
<td>Engineered</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyan (CFP)</td>
<td>434</td>
<td>476</td>
<td>0.42</td>
</tr>
<tr>
<td>Yellow (YFP)</td>
<td>513</td>
<td>528</td>
<td>0.6-0.8</td>
</tr>
<tr>
<td>Blue (BFP)</td>
<td>384</td>
<td>448</td>
<td>0.25-0.3</td>
</tr>
</tbody>
</table>
FLUORESCENCE RESONANCE ENERGY TRANSFER (FRET)
A “SPECTROSCOPIC RULER”

Two different fluorophores separated by distance $R$

**Fluorophore A**
- Absorbance or Emission

**Fluorophore B**
- Absorbance or Emission

If $R < \sim 10$ nm ($\sim 100$ Å) ‘resonance energy transfer’ can occur between A and B.

Non-radiative transfer of energy

Efficiency of transfer $\propto \frac{1}{R^6}$

- **Excite at $\lambda_1$**
  - $R > 10$ nm
  - Emission at $\lambda_2$
  - No FRET

- **Excite at $\lambda_1$**
  - $R < 10$ nm
  - FRET
  - Emission at $\lambda_3$
Some applications of FRET

1. Measuring interactions

Donor fluorophore (D) covalently attached to protein X, acceptor fluorophore (A) to protein Y

\[ \lambda_{\text{exc}}^D \rightarrow \lambda_{\text{em}}^D \rightarrow \lambda_{\text{exc}}^A \rightarrow \lambda_{\text{em}}^A \]

Fusion proteins in vivo


2. Structural information e.g. conformational change

A fluorescence polarimeter (similar to a spectrofluorimeter except for polarizers)

Polarization:
Measure of orientation(s) of plane of oscillation of light wave

Polarization, \( P = \frac{I_{\|} - I_{\perp}}{I_{\|} + I_{\perp}} \)

- \( I_{\|} = \) intensity of parallel component
- \( I_{\perp} = \) intensity of perpendicular component

Normal unpolarized light

Selects one plane of polarization of incident light

Polarizer A

Polarizer B

or

I_{\|}

parallel to plane of incident light

I_{\perp}

perpendicular to plane of incident light

Unpolarized or depolarized

P = 0

Polarized

P = 1
Molecular rotation and rotational relaxation time (\( \rho \))

Molecules tumble!

![Diagram of a molecule with axes x, y, and z with \( \rho_x \), \( \rho_y \), and \( \rho_z \) labels]

For a spherical molecule

\[
\rho_0 = \rho_x = \rho_y = \rho_z
\]

Rotational relaxation time, \( \rho \) (seconds) is a measure of rate of tumbling (= average time take to rotate 90°)

\[
\rho_0 \propto \text{Molecular Volume, } V \propto M_r
\]

A + B

\( M_r A \ll M_r B \)

Tumbles quickly

Tumbles slowly

Tumbles slower than free B and much slower than free A
Fluorophore emits photon

Fluorescence lifetime ($\tau$) 1 - 10 ns

Fluorophore absorbs photon

Plane polarized incident light

Covalently attach fluorophore to molecule A

Molecular rotation changes orientation of plane of polarization of emitted light

larger complex rotates more slowly

Emitted light remains more polarized

Emitted light depolarized

Covalently attach fluorophore to molecule A

- used in Jacobs et al. (Discussion paper for Oct 5th)
THERMODYNAMICS
- systematic description of POPULATION of molecules at EQUILIBRIUM in terms of MACROSCOPIC phenomena

Thermodynamic parameters: temperature (T), pressure (P) volume (V)

enthalpy (H) entropy (S) Heat Capacity (C)

Gibbs free energy (G) $K_d$

Laws of Thermodynamics:
1) Conservation of matter: Heat change, q (constant T, P) = $\Delta H \approx \Delta E$
2) Entropy and order: System always tends towards maximum entropy

Gibbs free energy G: $\Delta G = \Delta H - T\Delta S$

For molecular interactions

$$\Delta G = \Delta G^0' - RT \ln K_d$$

$$\Delta G^0' = + RT \ln K_d$$

$$K_d = e^{\left(\frac{\Delta G^0'}{RT}\right)}$$

Van’t Hoff analysis:
Measure $K_d$ at different T to determine $\Delta H^0', \Delta S^0'$ and $\Delta G^0'$:

$$\Delta H^0' = R \ln \left(\frac{K_d^1}{K_d^2}\right) \left(\frac{T_2 T_1}{T_2 - T_1}\right)$$

$$\ln K_d = \frac{\Delta H^0'}{RT} - \frac{\Delta S^0'}{R}$$
ISOTHERMAL TITRATION CALORIMETRY

Principle:
1. Add small amount of molecule Y ($\Delta Y$) to sample cell in successive injections (titrate)
2. Heat absorbed or evolved ($q$) as a result of interaction causes temperature difference ($\Delta T$) between sample reference cells
3. After each injection, feedback heaters add heat energy to sample or reference cell to bring $\Delta T$ back to zero

Each spike represents one injection of molecule Y. Integrating area of each peak gives total heat released ($\Delta q$) at each titration point (as total concentration of Y increases)

Can obtain information on $\Delta H$, $K_d$ (and $\Delta G$ and $\Delta S$)

Determine $\Delta H$ at different T gives $\Delta C_P = \frac{\Delta H_1 - \Delta H_2}{T_2 - T_1}$
**Thermodynamics of Interactions**

1. **Enthalpy:**
   - Individual components interact with water molecules (and ions)
   - Some interactions with water molecules at interaction surfaces are broken (unfavorable).
   - New interactions between X and Y at interface (favorable).
   - Lower heat capacity.

2. **Entropy:**
   - Entropy of rotation and translation of free X and Y.
   - Relative rotation/movement of X and Y reduced so entropy reduced.
   - Entropy of bound water molecules low.
   - Entropy higher (displaced water).

3. **Gibbs Free Energy:**
   - Positive (unfavorable).
   - Negative (favorable).
   - \(-T\Delta S\) rotation/movement.
   - \(\Delta H\) displaced water (hydrophobic effect).
   - \(-T\Delta S\) net \(\Delta G < 0\).
PROPERTIES OF INTERACTION SURFACES, INTERACTION STRENGTH AND ENTHALPY vs. ENTROPY DRIVEN INTERACTIONS

1) Size and nature (properties) of interface determines magnitude of \( \Delta G^0 \) and therefore \( K_d \) (strength of binding)

2) Relative contribution of enthalpy (\( \Delta H^0 \)) and entropy (\( -T \Delta S^0 \)) to a negative (favorable) \( \Delta G^0 \) also depends on size and properties of interface

Small surface area of interaction (\( Y = \) small molecule e.g. glucose, ATP)

\[ \Delta G^0 = \Delta H^0 - T \Delta S^0 \]

ENTHALPY DRIVEN (\( \Delta H^0 > -T \Delta S^0 \))

fewer displaced water molecules means smaller contribution to favorable entropy

Small molecules also have higher entropy (greater movement) so larger entropic cost when bound in complex

Large surface area of interaction (protein:protein, protein:nucleic acid interactions)

\[ \Delta G^0 = \Delta H^0 - T \Delta S^0 \]

ENTROPY DRIVEN (\( \Delta H^0 \leq -T \Delta S^0 \))

More interactions between X and Y means greater favorable enthalpy

More water molecules displaced so larger favorable entropy contribution

Experimentally observed that magnitude of change in heat capacity (\( \Delta C_p \)) is related to size (surface area) of interaction surface

Large \( \Delta C_p \) can also result from conformation change in molecules upon interaction (proportional to number of residues involved)
PROTEIN:PROTEIN COMPLEXES

Diagram showing interaction surfaces exposed to solvent and inaccessible to solvent.

Non-obligate (non-permanent) complex vs. permanent (obligate) complex:

- **Properties of Interaction Surface**
  - **SIZE**
    - Average ~1600 Å² (~800 Å² from each protein)
    - Average ~3400 Å² (up to 5000 Å²)
  - **SHAPE**
    - Relatively flat
    - More convoluted
  - **CHEMISTRY**
    - Up to ~40 amino acids (20 from each protein)
    - ~700 Å² non-polar groups (modest contribution from hydrophobic effect to TΔS)
    - ~900 Å² polar groups (enthalpy contribution: hydrogen bonds, electrostatic)

- **Surface Area (Å²)**

- **NON-OBLIGATE COMPLEX**
  - Weaker interaction
  - Relatively high $K_d$
  - Modest, negative $\Delta G^{0'}$

- **PERMANENT (OBLIGATE) COMPLEX**
  - Very strong interaction
  - Very low $K_d$
  - Large, negative $\Delta G^{0'}$

- Polar groups important since when proteins X and Y are *not* interacting with each other they have to interact favorably with water.
A NON-OBLIGATE COMPLEX
Cytochrome c : Cytochrome c peroxidase

A PERMANENT COMPLEX
Histone H2A:H2B heterodimer
STEREOCHEMICAL COMPLEMENTARITY - SHAPE MATTERS!

Complementary shape and chemistry (spatial arrangement of chemical groups)
So maximize possible number of favorable interactions to contribute to best $\Delta G$ possible

Different shapes and chemistry
Few, if any interactions possible
Therefore no binding

If the shape and/or chemistry of the interaction surface changes (e.g. by natural or artificial mutation, post-translational modification) the interaction strength will change

Some interactions but shape and chemistry does not entirely fit well so weaker binding

Greater complementarity so more favorable interactions can be made giving greater $\Delta G$ which means stronger binding
INTERACTION OF THE CHROMODOMAIN OF HP1 WITH N-TERMINAL TAIL OF HISTONE H3


Cage of hydrophobic side chains surrounds end of Me₃K9 residue

Hydrogen bonds between main chain of HP1-CD and Me₃K9-H3 main chain


Useful literature for lecture 1

**General:**


**Fluorescence:**


**Isothermal Titration Calorimetry:**