Lecture 1: Introduction to molecular systems biology

Marko Djordjevic

The Chair for General Physiology and Biophysics, Faculty of Biology, University of Belgrade
Overview
(material for the school)

- **Lecture 1 (MD1):** Introduction to computational systems biology
- **Lecture 2 (MD2):** Modeling intracellular dynamics (introduction)
- **Working session 1:** Modeling intracellular dynamics (more advanced): bistability, bifurcations
- **Working session 2:** Biological rhythms (genetic oscillators): delay and relaxation oscillator
- **Lecture 3 (MD3):** Modeling gene expression regulation
- **Working session 3:** Examples of a biophysical approach to bioinformatics
Overview

• Basic molecular biology (brief introduction)
• General discussion: why quantitative biology?
• Non-linearity: feedback loops, cooperativity, allostery
I. Overview of basic molecular biology
Central dogma of molecular biology

**DNA** - medium for storing information, can be duplicated (during replication).

**RNA** - an intermediate (short lived), synthesized from DNA during transcription.

**Proteins** – translated from RNA, do all useful work in the cell.
DNA

ATGACCATGATTACGGGATTCACTGGCCGTCGT

From information point-1 dim sequence of letters

Double helix structure

DNA can be duplicated-replication
RNA
genome consists of DNA genes + intergenic regions

ATGACCATGATTACGGATTCACTGGCCGTCGT → DNA

transcription

AUGACCAUGAUUACCGGAUUCACUGGCCGUCGU → mRNA

RNA can assume quite complex 3D structure
From RNA to protein

Translation scheme

AUGACCAUGAUUACGGAUUCACUGC

Read mRNA three bases at time

Polypeptide sequence

Folds into protein

The Genetic Code
Regulation of gene expression

Pattern of gene expression has to change—i.e. it has to be regulated

In most cases transcription is regulated

Transcription process
Transcription factors (TFs)

TFs interact with DNA

a single amino-acid nucleotide interaction
Transcription regulation

The diagram illustrates the regulation of transcription in the lac operon. The CAP-binding site and the RNA-polymerase-binding site (promoter) are shown along the DNA. The start site for RNA synthesis is indicated.

- **+ GLUCOSE + LACTOSE**: OPERON OFF because CAP not bound.
- **+ GLUCOSE - LACTOSE**: OPERON OFF both because lac repressor bound and because CAP not bound.
- **- GLUCOSE - LACTOSE**: OPERON OFF because lac repressor bound.
- **- GLUCOSE + LACTOSE**: OPERON ON.
Gene networks

E. Coli gene network
II. Modeling in biology: why quantitative biology?
Traditional modeling

Scheme of traditional research in biology

No clear separation between theory and experiment
Scheme vs. model

RNA polymerase (RNAP) transcription scheme

This scheme is a model!

Can be described by even simpler conceptual model

General principle is complementarity

Alberts et al, Molecular Biology of the cell.
Information flow model

Central dogma of molecular biology

Emphases that information flows from DNA to RNA to proteins
Biological sequences

• Large amount of data is sequenced
  ~ 80000 sequenced genomes
  - Bacteriophage (genome ~50000 bps)
  - Bacteria (genome ~5 000 000 bps)
  - Human (chromosome ~50 000 000 bps)

• Informatics resources are necessary for repository and systematization

• Mathematical methods are necessary for sequence analysis

bioinformatics
Can have a close connection between biophysical modeling and bioinformatics

Quantitative (biophysical) understanding of biological processes

Improved bioinformatics methods

Specific examples in working session 3!
Measuring expression of all genes in genome

Mathematical modeling necessary for testing the hypothesis.
Research scheme in modern biology

Separation between theory and experiment

Falsifiable predictions → Quantitative measurements

Theoretical Models

Scheme reminds to physics
Understanding dynamics (complicated)

- Relatively easy to measure dynamics for macroscopic systems.
- Hard to measure dynamics of molecules “in-vivo” (but recent advances are making it possible).
- Systems typically employ a large number of components (degrees of freedom) with interactions that are highly non-linear.
- Advanced biophysical techniques from theoretical biophysics are necessary (statistical physics, nonlinear dynamics, stochastic modeling)
III. Role of non-linearity: Feedback loops, allostery and cooperativity
Central dogma of molecular biology

Classical formulation: Information flows from DNA to RNA to proteins

For (computational) systems biology feedback loops are crucial.
Intracellular regulation –
general properties

- Nonlinear relationship of output vs. input quantity – linear relationship is generally not a good approximation.
- Allostery and cooperativity often have a crucial role.
Input-output relationship

$A$ and $B$ are arbitrary molecules, $AB$ is their complex:

$$A + B \rightleftharpoons AB$$

Examples:

<table>
<thead>
<tr>
<th>$A$</th>
<th>$B$</th>
<th>$AB$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>Transcription factor</td>
<td>Complex which activates or represses transcription</td>
</tr>
<tr>
<td>$B$</td>
<td>DNA</td>
<td></td>
</tr>
<tr>
<td>$AB$</td>
<td>Small molecule (e.g. oxigen)</td>
<td>Ligand bound за рецептор</td>
</tr>
<tr>
<td>$A$</td>
<td>Receptor (e.g. hemoglobin)</td>
<td></td>
</tr>
<tr>
<td>$B$</td>
<td>$AB$</td>
<td></td>
</tr>
<tr>
<td>$A$</td>
<td>Antibody</td>
<td>Complex antibody-antigene</td>
</tr>
<tr>
<td>$B$</td>
<td>Antigene</td>
<td></td>
</tr>
<tr>
<td>$AB$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Input (ligand concentration) to output (binding probability) relationship

\[ K_D = \frac{[A][B]}{[AB]} \quad \text{dissociation const.} \]

"input" = [B]

"output" = \[ \frac{[AB]}{B} + [AB] = \frac{[A]}{[A] + K_D} \]

Receptor occupancy

(Fraction of the total receptors bound by ligand)
Cooperativity and allostery

Hemoglobin as a model system:

• Interaction of ligand with receptor
• Protein structure
• Molecular model of disease
• Allostery and cooperativity
Myoglobin – only one oxygen bound
Hemoglobin – four oxygens bound

Hemoglobin has much more rapid transition from “OFF” to “ON” state.

OFF state – small oxygen amount bound to receptors
ON state – almost all receptors bound by oxygen

Consequence of cooperativity in interactions (binding of oxygen).
Cooperativity in interactions is related with allostery – Ligand binding at one place in protein, influences binding properties at some other place.
Hill function

\( n \) molecules \( A \) are cooperatively (together) bound to molecule \( B \)

\[ nA + B \rightleftharpoons nAB \]

\[ K_d^n = \frac{[A]^n[B]}{[nAB]} \]

"output" = \[ \frac{[nAB]}{[B] + [nAB]} = \frac{([A]/K_d)^n}{1 + ([A]/K_d)^n} \]

\( \text{Hill constant} \)

Hill function - cooperative binding of ligand to receptor
Problem: hemoglobin vs. myoglobin binding

Hill curve, hemoglobin

Experimental data for fractional occupancy of hemoglobin receptors by oxygen are provided in the Excel file.

a) Show that Hill's function fits well this data
b) What is the Hill constant value?
c) What is biophysical interpretation of the second parameter in the fit?

Note1: The first column in Excel file is oxygen pressure (proportional to ligand concentration), the second column is fractional receptor occupancy.

Note2: You can do this in any software of your choice, where non-linear regression is available.
The same as for hemoglobin, but now data for mioglobin are provided.

a) Show that Hills function again fits well this data.

b) What is now the value of the Hill constant?

c) Based on this, discuss what is crucial difference between hemoglobin and mioglobin with respect to receptor binding?
Literature:

General:

Specific:
• Bruce Alberts et. al. (2014). *Molecular Biology of the Cell*. W. W. Norton & Company