EE381V: Genomic Signal Processing

Lecture #14

The Course So Far

Gene finding
DNA
Genome assembly
Regulatory motif discovery
Comparative genomics
Sequence alignment
Gene finding
DNA

SEQUENCES

DONE (mostly)

INTERACTIONS

Gene expression analysis
Regulatory networks inference
Emerging network properties
Protein network analysis

Cluster discovery
Biomolecular Detection Systems

- Bio-molecular detection systems, in general, have the following sub-blocks:

**Molecular Biology Biochemistry**
- DNA Microarrays

**Fabrication/Synthesis Processes**
- Photodiode

**Circuit Design**
- Image Sensor

**Signal Processing**
- Analog to Digital Conversion

**Example:**
- DNA
- Microarrays

**Process:**
- Biological Sample
  - Biological Assay
  - Transducer/Interface
  - Detection Circuitry
  - Detection

**Uncertainties:**
- Biochemical Uncertainties
- Fabrication and Process Variation
- Electronic Noise
- Quantization Noise

**Interface:**
- ADC
Affinity-Based Biosensors

- Exploit the affinity of certain biomolecules for each other to capture and detect:

* There are different transduction methods for “counting” the binding events, e.g., fluorescence, electrochemical, chemi-luminescence ...

Parallel Affinity-Based Sensing

- For simultaneous detection of multiple targets, use affinity-based sensors in parallel:

Biological sample
Planar solid surface
Capturing sites
DNA Microarrays

- DNA microarrays are **massively parallel** affinity-based biosensor arrays:

**Applications of DNA Microarrays**

- Recall central dogma:

- DNA microarrays interrupt the information flow and measure gene expression levels
  - frequently, the task is to measure relative changes in mRNA levels
  - this gives information about the cell from which the mRNA is sampled (e.g., cancer studies)

- Other applications:
  - single nucleotide polymorphism (SNP) detection
  - simultaneous detection of multiple viruses, biohazard / water testing
Sensing in DNA Microarrays

- When complementary ssDNA molecules get close to each other, electrostatic interactions (in form of hybridization bonds) may create dsDNA

- Because of thermal energy, the binding is a reversible stochastic process
- Relative stability of the dsDNA structures depend on the sequences

Stability of dsDNA: Melting Temperature

- The *melting temperature* \( (T_m) \) of a DNA fragment: the temperature at which 50% of the molecules form a stable double helix while the other 50% are separated into single strand molecules

- Melting temperature is a function of DNA length, sequence content, salt concentration, and DNA concentration
- For sequences shorter than 18 ntds, there is a simple (Wallace) heuristic:

\[
T_m = 2(A + T) + 4(G + C)
\]
There are many different methods to approximate melting temperature:

**Wallace method** (DNA strands less than 18mers):

\[ T_m = 2(A + T) + 4(G + C) \]

**%GC method:**

\[ T_m = 81.5 + 16.6(\log 10[Na^+] + 0.4[\%GC] \cdot 625 / N) \]

**Nearest neighbor method [1]:**

\[ T_m = \frac{\Delta H - 3.4\text{cal}}{A + \Delta S + R \ln(C/4)} \approx 273.15 + 16.6 \log 10[Na^+] \]

- \( \Delta H \) is the sum of nearest neighbor enthalpy changes
- \( A \) is the initiation constant of -10.8 cal/K° mole for non-self complementary sequences, -12.4 cal/K°mole for complementary sequences
- \( \Delta S \) is the sum of nearest neighbor entropy changes
- \( R \) is the gas constant 1.987 cal/K°mole
- \( C \) is the concentration of DNA (generally fixed at 250 pM) [2]


Melting temperature is a function of DNA length, sequence content, salt concentration, and DNA concentration.
Nonspecific Binding (Cross-hybridization)

- Depending on sequences, non-specific binding may also happen:

\[ \Delta E_1 > \Delta E_2 \]

(Bond stability) \(_1 > (\text{Bond stability})_2 \)

- So, non-specific binding is not as stable as specific binding

Non-specific binding in microarrays

- Non-specific binding (cross-hybridization) manifests as interference:

- Interference may lead to erroneous conclusions
- Useful signal is affected by the interfering molecules, false positives possible
The kinetics of reactions depends on:

1. The frequency that the reactive species (e.g., molecules) get close to each other.
2. If in close proximity, can thermal energy “facilitate” microscopic interactions.

**Road to a Model: Underlying Physics**

**Arrhenius equation** represents the dependence of the rate constant \( k \) of a reaction on temperature \( T \):

\[
k = A e^{E_a/RT} \tag{1}
\]

In its original form the pre-exponential factor \( A \) and the activation energy \( E_a \) are considered to be temperature-independent.
The reaction kinetics of two molecules has the following Markov model:

\[
p_{0,0} = 1 - p_{0,2} \\
p_{0,2} = 2A_e^{\Delta E_2/k_B T} \\
p_{1,2} = 1/2 \\
p_{1,3} = 1 - p_{1,2} \\
\]

State 0: Molecules bind
State 1: Molecules close
State 2: Molecules activated

A more practical model is the two-state Markov model:

\[
p_{0,0} = 1 - p_{0,2} \\
p_{0,2} = 2A_e^{\Delta E_2/k_B T} \\
p_{2,1} = 1/2 \\
p_{2,2} = 1 \\
p_{1,2} = 1 - k_1 \Delta t \\
p_{1,3} = 1 - k_1 \Delta t \\
\]

State 0: Molecules bind
State 1: Molecules close
A more practical model is the two-state Markov model

\[
p_{0,1} = 1 - k_{1,0} \Delta t, \quad p_{1,0} = k_{1,0} \Delta t, \quad p_{0,0} = k_{0,0} \Delta t, \quad p_{1,1} = 1 - k_{1,1} \Delta t
\]

**State 0**
Molecules bind

**State 1**
Molecules close

Steady-state distribution:

\[
\begin{bmatrix}
    k_{1,0} \\
    k_{1,0} + k_{1,1}
\end{bmatrix} =
\begin{bmatrix}
    1 - k_{1,0} \Delta t & k_{1,0} \Delta t \\
    k_{1,0} \Delta t & 1 - k_{1,1} \Delta t
\end{bmatrix}^T
\begin{bmatrix}
    k_{1,0} \\
    k_{1,0} + k_{1,1}
\end{bmatrix}
\]

\[
\frac{k_{1,0}}{k_{1,0} + k_{1,1}} = \frac{1}{1 + Re^{-\Delta t/\tau}} \quad (\text{II})
\]

\[
\frac{k_{1,1}}{k_{1,0} + k_{1,1}} = \frac{1}{1 + Re^{-\Delta t/\tau}} \quad (\text{III})
\]

The number of captured targets forms a continuous-time Markov process
Microarrays

The steps involved in an experiment:

1. Sample Preparation
2. Array Fabrication
3. Incubation
4. Detection