

BE.011/2.993J
Spring 2003
QUIZ I SOLUTIONS

1.) (20 points) DNA sequence

You have a genome that is exactly 1×10^9 bases long. You would like to choose a DNA sequence out of the genome that is unique. What is the minimum length of the sequence such that it could be unique (i.e., it is possible that it does not occur anywhere else in the genome)?

Basically, we want the probability of the sequence occurring to be less than $1/(1 \times 10^9)$. That means it has the probability of occurring only once in the genome, and thus could be unique. Another way of looking at it is that if its probability were greater than $1/(1 \times 10^9)$, such as $2/(1 \times 10^9)$, then it has the possibility of occurring **twice** in the genome and may not be unique.

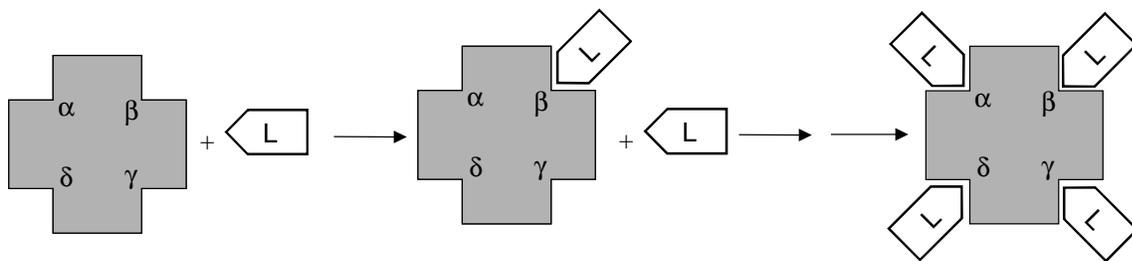
Now, for a sequence, the probability that it occurs is $1/4^n$, where n is the length of the sequence (remember the coin flips from Chapter 1? In that case, as well as here, order matters). The 4 comes from the 4 nucleotides we have to choose from (A,T,G,C).

For a 14mer:
 $4^{14} = 2.68 \times 10^8$
 $1/4^{14} > 1/(1 \times 10^9)$
 so not long enough.

For a 15mer:
 $4^{15} = 1.07 \times 10^9$
 $1/4^{15} < 1/(1 \times 10^9)$
 so long enough to be unique!

By the way, this is based on an experimental technique called antisense gene regulation. In that case, a short “antisense” oligo is used to hybridize to mRNA over the start codon, prohibiting the gene from being translated. In this experiment, you’d like to choose an antisense oligo long enough so that it hybridizes to only one spot in the gene, and not anywhere else. If you look in the literature, these antisense oligos are almost always 15-20 nucleotides long.

2.) The protein below has four binding sites (α , β , γ , δ) for the ligand L. We would like to find its equilibrium binding population. For now assume that the association and dissociation constants are equal.



a) (15 points) Calculate W and the entropy (in units of k) for the situation in which

i) 0 ligands are bound ($N_L = 0$)

Only 1 way to arrange this.

$$W = 4!/4!0!$$

$$S = k \ln(1) = 0$$

ii) 1 ligand is bound ($N_L = 1$)

4 ways to arrange: on $\alpha, \beta, \gamma, \delta$ (by "hand")

or $W = 4!/3!1!$

$$W = 4$$

$$S = k \ln(4) = 1.386k$$

iii) 2 ligands are bound ($N_L = 2$)

6 ways to arrange (by "hand"):

$\alpha\beta, \alpha\gamma, \alpha\delta$

$\beta\gamma, \beta\delta$

$\gamma\delta$

or $W = 4!/2!2! = 6$

$$S = k \ln(6) = 1.792k$$

iv) 3 ligands are bound ($N_L = 3$)

4 ways to arrange (vacant spot is on $\alpha, \beta, \gamma, \delta$)

$$W = 4!/1!3! = 4$$

$$S = k \ln(4) = 1.386k$$

v) 4 ligands are bound ($N_L = 4$)

$$W = 4!/0!4! = 1$$

$$S = 0k$$

b) (5 points) Which states have the highest entropy?

$$N_L = 2$$

c) (5 points) Which states have the lowest entropy?

$$N_L = 0 \text{ and } N_L = 4$$

d) (5 points) Which state will you find it in at equilibrium?

$N_L = 2$ (because it has the highest entropy)

e) (20 points) Let's say that the binding constants are not equal, i.e., ligand L has a higher probability of being bound:

$$p_{\text{bound}} = 0.75$$

$$p_{\text{unbound}} = 0.25$$

Plot the probability distribution for all of the states, $p(i)$.

Now we have to use the formula for the Bernoulli trials (Dill, equation 1.28). Basically this multiplies W by a term which accounts for the nonequal probabilities of a protein being bound:

$$P(N_L, M) = M!/N_L!(M-N_L)! p^{N_L} (1-p)^{M-N_L}$$

for N_L ligands in M sites, which have a probability p of being bound (and $1-p$ of being unbound). Note: this is a little more like real life because proteins often have unequal association-dissociation constants.

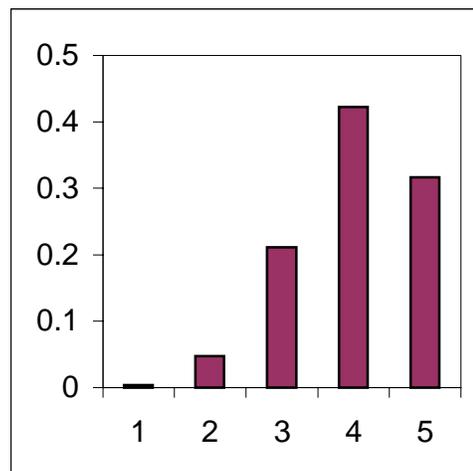
for $N_L = 0$:

$$P = 4!/0!4! (0.75)^0(0.25)^4 = 3.91 \times 10^{-3}$$

for $N_L = 1$:

$$P = 4!/(1!3!) (0.75)^1(0.25)^3 = 4.69 \times 10^{-2}$$

for $N_L=2$:



$$P = 4!/(2!2!) (0.75)^2(0.25)^2 = 0.211$$

for $N_L=3$:

$$P = 4!/(3!1!) (0.75)^3(0.25)^1 = 0.422$$

for $N_L=4$:

$$P = 4!/(4!0!) (0.75)^4(0.25)^0 = 0.316$$

Note: the sum of all $p_i = 1.00$, which is important!

3.) DNA has 3 different possible configurations: A, B, and Z: Let the scores for each configuration ϵ_i be : $\epsilon_A = 1$, $\epsilon_B = 2$, $\epsilon_Z = 3$.

a) (10 points) Write down the partition function, q . (you may substitute $x = e^{-\beta}$)

$$q = \sum_i e^{-\beta \epsilon_i}$$

$$q = x + x^2 + x^3$$

b) (10 points) Write down the expressions for p_A and p_B and p_Z .

$$p_A = x^1/q = x/(x + x^2 + x^3)$$

$$p_B = x^2/q = x^2/(x + x^2 + x^3)$$

$$p_Z = x^3/q = x^3/(x + x^2 + x^3)$$

c) (10 points) What is the variance of the distribution, $\langle \sigma^2 \rangle$, if all three configurations are equally possible?

If all 3 configurations equally possible, then $p_A = p_B = p_Z = 1/3$

Then we need the formula for the variance:

$$\langle \sigma \rangle = \langle \epsilon^2 \rangle - \langle \epsilon \rangle^2$$

where

$$\langle \epsilon \rangle = \sum \epsilon_i p_i = 1(1/3) + 2(1/3) + 3(1/3) = 2$$

$$\langle \epsilon^2 \rangle = \sum \epsilon_i^2 p_i = 1(1/3) + 4(1/3) + 9(1/3) = 4.67$$

$$\langle \sigma \rangle = 4.67 - 2 = 2.67$$