COMS4761:
ASSIGNMENT 4.4: ezLife HMM Applications

This assignment continues to investigate Dr. No's ezLife, as defined in Assignment 2.1. To summarize: ezLife uses two nucleic acids, denoted {0,1}, which form stable ezRNA molecules, not requiring DNA. ezRNA molecules are translated to ezProteins, composed of 4 ezAmino-Acids (ezAA)–a,b,c,d-- using the genetic code: 00→a; 01→b; 10→c; 11→d.

Problem 1: Using Profile HMM to analyze ezRNA structures

Figure (1) below shows a Multiple Sequence Alignment (MSA) of ezRNA gene components recovered from 10 ezLife species.

```
1110101-00001-111111
1110000-00001-101111
1110101000001-110111
1110-1--00000-110111
1110101-00001-101111
1110-0100000011-111
1110110-0000--111111
1110001-0000101111111
111010--0000--11-111
1110101-00001-111111
```

Figure 1: An MSA of ezLife Genes

A. Create an HMM profile model for these gene components (hint: prune the HMM graph to aggregate single-symbol columns; explain your model.
B. What would be the impact on the HMM graph if sequences 4 and 9 had a 0 in column 7; what would be the impact if sequences 2,3,5,6 had indel in column 7?
C. Compute the transition and emission probabilities for the HMM based on these sample sequences.
D. What is the most likely HMM path to generate the sequence: 111011100001111111.
E. How likely is the HMM to generate this sequence (what is the probability)?

Problem 2: Computing Conformations of ezProteins

ezProteins are designed for a two-dimensional world; their folds may be entirely described in terms of planar conformations as follows:

- The ezAA {a,b,c,d} may be described as two dimensional unit vectors (i.e., they have equal length and their width/length ratio is very small).
- The conformation angles of ezPeptide-bonds are {0,π/2,-π/2}. That is, the bond angle formed by an ezAA to its predecessor in an ezProtein sequence is either 0 (laying on the same line as the predecessor), or is orthogonal to the predecessor in a counterclockwise direction (π/2) or in a clockwise direction (-π/2); we will use `{0,+,-}` to describe these bond angles
- Therefore, ezProteins conformations may be represented as grid paths.

Figure 2, below, depicts some basic ezConformations: zigzags (1,2), trains (3,4), line (5) and globules (6,7). Arrows represent ezAA molecules with colors representing the bond angle: black for 0, blue for π/2 (+) and red for –π/2 (-). For example, the small zigzag (2.1) corresponds to the sequence of bond angles S+-+-+-+- , the line (2.5) corresponds to `S0000`,the larger train (2.4) to S0+0-0-0+0+0-0-0+0, and the large globule (2.7) to S+0-0-0-0. Here `"S"` represents the selectable direction of the first ezAA of an ezProtein, typically set to 0.

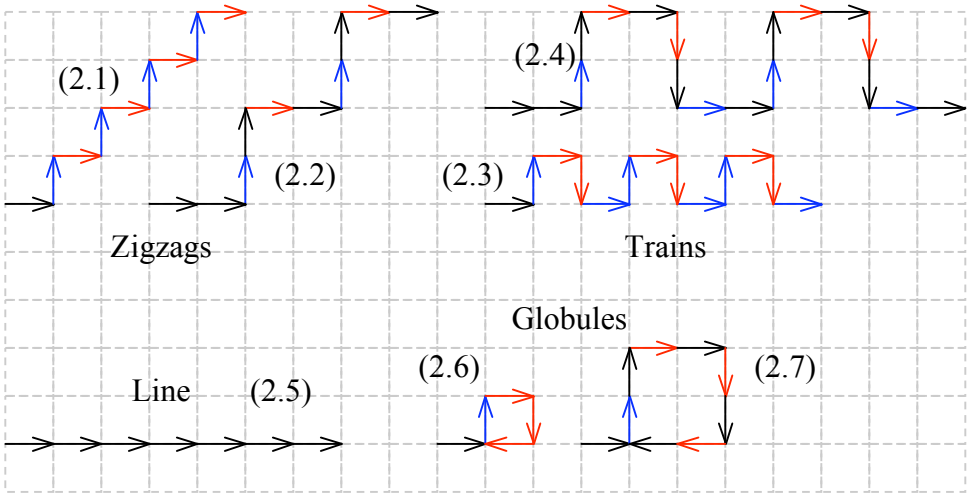

Figure 2: ezProtein Conformations

Consider an HMM model to analyze and predict ezProtein conformations. There are 5 hidden states, START (S) and End (E) states and one for each conformation angle `{0,+,-}`; these hidden states can emit the ezAA symbols {a,b,c,d}. Suppose the following database of ezProteins conformation has been obtained through crystallography:

ezProtein	Protein Sequence	Conformation Sequence
pyramidin	aabcbabdcabcba	S+-+-+-0-+-+-+
flagelin	bdddadbdbdcd	S000+0-0-0-0
cactuslin	cadcdbdbbaacbacbbdcdada	S+0+0-0--++--++--0-0+0+
Holin	dcddbcaabbddbbca	S+00--++--00--++
Snooplin	dcadbcabdbadddbbacbcabb	S++0--+-0-+000--+--++--

A. Draw the HMM with the transitions and emissions probabilities.

B. A newly discovered ezProtein colin, has been sequenced: **dddadacbbacd**
Use Viterbi decoding to compute its likely conformation and draw it.

C. (30 points) ezLife uses signaling pathways, depicted in figure 3 below, as follows. Z-proteins, conformed as short small Zigzag, on the left, are coupled with a transmembrane receptor ZPCR (Z-Protein Coupled Receptor). The ZPCR binds these Z-proteins to its own Zigzag-shaped coupling domain, protruding into the cell. The ZPCR receptor side consists of a small Globulin attached to a short small Train conformation, protruding outside the cell. Signaling proteins are organized as short Trains, which bind with the ZPCR receptor. Upon binding of a signaling molecule, the ZPCR switches the conformation of its Zigzag domain to a Line and releases the Z-protein at its tail. The Z-protein then activates a respective pathway. The goal of this part is to construct an HMM to detect ZPCR proteins.

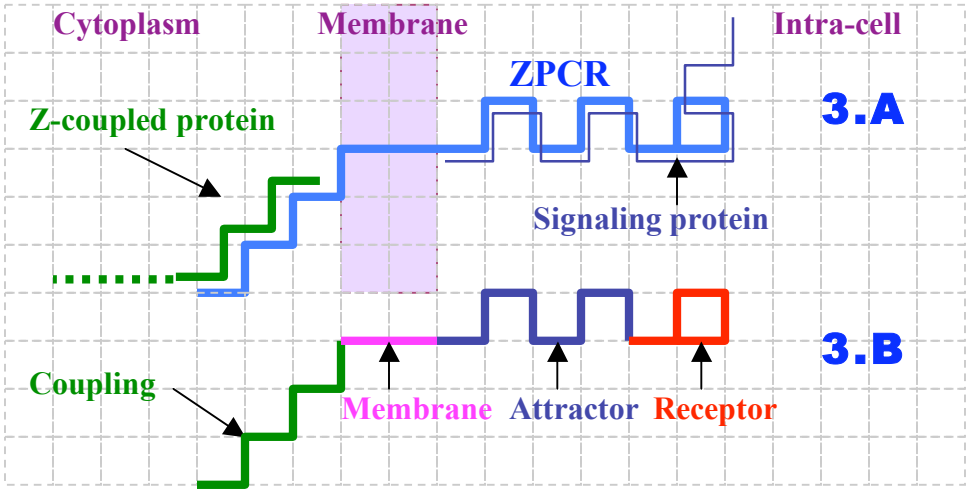

Figure 3: Operations (3.A) & Architecture (3.B) of ZPCR

(i) ZPCR proteins consist of 4 domains: (a) a Coupling domain, conformed as a small Zigzag (figure 2.1) of geometrically distributed length, averaging 8 ezAA; (b) a Membrane domain, conformed as a Line (figure 2.5) of 2 ezAAs; (c) a Signal Attractor domain, conformed as a small Train (figure 2.3), of geometrically distributed length, averaging 12 ezAAs; and (d) a Receptor domain, consisting of single small globulin (figure 2.6).

Design an HMM to detect ZPCR proteins; draw the HMM states, transition probabilities and emission probabilities (assume the emission probabilities are identical to those computed at part A). Explain your design choices and computations of probabilities.

(ii) Is the following ezProtein a ZPCR: **abcbdddabbaabcabcb** ?