The DNA sequence of a gene determines the amino acid sequence of the protein it produces, with every 3 DNA basepairs coding for one protein amino acid. In eukaryotes, predicting the protein sequence from a gene sequence is not as simple as taking the gene sequence and splitting it into 3's, because of *introns*, stretches of gene sequence that do not code for amino acids. (The regions of DNA in a gene that are not introns are called exons.)

A Markov chain model has been used to detect introns, assuming the 6 states shown in the diagram. The index set here does not represent time, but rather represents basepairs: Each increment is the transition from the previous basepair to the neighboring basepair.
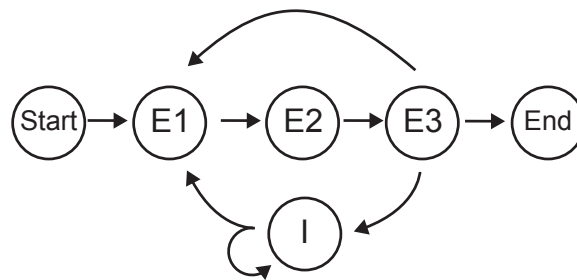
  i Suppose a newly-discovered species has the following properties:

- The average protein is $L_{\mathrm{g}} = 20000$ amino acids long (i.e., it is 60,000 basepairs long, not counting introns).

- The average protein-coding gene has $N_I = 10$ introns.

- The average intron is $L_I = 6000$ basepairs long.

Write the 6-state Markov transition matrix. The matrix should contain numerical values (e.g., expressed as a fraction or decimal).

 ii Under the assumptions of this model, what is the average length of a gene, including introns?