Solution: JC69 DNA Substitution Model

Problem Statement

In the simplest model of DNA substitution (JC69), each nucleotide can mutate to any other nucleotide with the same probability per unit time r. Knowing that at time $t = 0$, the base at the site is A:

$$P(A, t = 0) = 1$$

derive the general expression for the probability $P(G, t)$ of the site having a base G at time t.

Solution

1) Probability of a base change in Δt

Since each nucleotide can mutate to any other with equal probability per unit time r, the probability of a mutation occurring in a small time interval Δt is

$$P(\text{mutation in } \Delta t) = r\Delta t.$$

2) Expression for $P(G|t + \Delta t)$

The probability $P(G|t + \Delta t)$ can be expressed as:

$$P(G|t + \Delta t) = P(G|t) + F(t),$$

where $F(t)$ accounts for all possible events in Δt that change $P(G|t)$. These events include:

- A mutation from G to another base: this occurs with probability $r\Delta t P(G|t)$.
- A mutation into G from another base (A, T, or C): since each of these three bases mutates to G with probability $\frac{r}{3}\Delta t$, the total probability is:

  $$\frac{r}{4}\Delta t(1 - P(G|t)).$$

  Where we have used the fact that $P(A, T, \text{or } C|t) = 1 - P(G|t)$

Thus,

$$F(t) = -r\Delta t P(G|t) + \frac{r}{3}\Delta t(1 - P(G|t)).$$

3) Differential equation for $P(G, t)$

Rearranging the above equation, dividing by Δt, and taking the limit $\lim_{\Delta t \to 0}$ we obtain the differential equation, that describes the rate of change of the probability $P(G|t)$:

$$\lim_{\Delta t \to 0} \frac{P(G|t + \Delta t) - P(G|t)}{\Delta t} = \frac{dP(G|t)}{dt} = -rP(G|t) + \frac{r}{3}(1 - P(G|t)).$$

$$\Rightarrow \frac{dP(G, t)}{dt} = \frac{r}{3}(1 - 4P(G|t)).$$

4) Solution of the differential equation

The resulting equation is a first order linear differential equation. We can solve this differential equation in different ways. One of them is defining the change of variable

$$1 - 4P(G|t) \equiv u(t) \Rightarrow P(G|t) = \frac{1 - u(t)}{4}.$$

Substituting into the differential equation:

$$\frac{du}{dt} + \frac{4r}{3}u = 0.$$

This is a separable equation, and its general solution is:

$$u(t) = Ce^{-\frac{4r}{3}t}.$$

We can recover the expression for $P(G|T)$ by substituting the change of variable back

$$1 - 4P(G,t) = Ce^{-\frac{4r}{3}t}.$$

Applying the initial condition $P(G,0) = 0$, we find $C = 1$, so:

$$\boxed{P(G,t) = \frac{1}{4}\left(1 - e^{-\frac{4r}{3}t}\right)}.$$

Note that this expression has the right limit at long times $P(G|t \to \infty) \to 1/4$ i.e. all the nucleotides have the same probability since the memory from the initial state is erased.