Error Threshold in Pre-biotic Evolution

In order to encode enzymes the RNA has to exceed a certain length. However, at the same time there is an error threshold which determines an upper bound for successfull replication IF no enzymes are present to support the copying process. Let x_i denote the concentration of molecules with sequence i, we can write down a simplified replication equation as follows:

$$\frac{\partial x_i}{\partial t} = (A_i\,Q - E)x_i + \sum_{j=1}^{n} w_{ij} x_j, \tag{4}$$

$$\sum_{i=1}^{n} x_i = 1 \tag{5}$$

where A_i is the replication rate, n is the number sequences, Q is the replication fidelity, w_{ij} is the mutation rate and E is a normalization factor. The replication fidelity can be written as $Q = q^N$ for a sequence of length N and replication fidelity q per base. The error rate per base is then given by $(1 - q)$.

In the following, we will simplify equation (4) by combining all sequence concentrations which are not the master sequence, i.e. the sequence with the highest fitness, into one:

$$X_m = x_k, \quad \text{sequence } k \text{ has the highest fitness,} \tag{6}$$

$$X_r = \sum_{i=1, i\neq k}^{n} x_i. \tag{7}$$

Thus, we end up with three equations:

$$X_m + X_r = 1, \tag{8}$$

$$\frac{\partial X_m}{\partial t} = A_m\,Q\,X_m - E\,X_m \tag{9}$$

$$\frac{\partial X_r}{\partial t} = A_r\,X_r + A_m\,(1 - Q)\,X_m - E\,X_r \tag{10}$$

Where has the mutation term gone? Since we have only two sequences left, every mutation of the master results in a sequence belonging to X_r. Thus, the term $A_m\,(1 - Q)\,X_m$ represents the effect of the mutation of the master sequence. Nearly all mutation of a sequence in X_r will again result in a sequence in X_r. The very small probability that a mutation will result in the master sequences X_m is omitted. Now, we derive the condition for $X_m > 0$ at equilibrium, meaning

$$\frac{\partial X_m}{\partial t} = \frac{\partial X_r}{\partial t} = 0. \tag{11}$$

Using equation (8), we can eliminate X_r in equation (10):

$$\begin{aligned} 0 &= A_r\,(1 - X_m) + A_m\,(1 - Q)\,X_m - E\,(1 - X_m) && (12)\\ &\Leftrightarrow \\ E - A_r &= X_m\,(-A_r + A_m\,(1 - Q) + E) && (13) \end{aligned}$$

Using $E = A_m\,Q$ from equation (9), we get:

$$\begin{aligned} A_m\,Q - A_r &= X_m\,(-A_r + A_m\,(1 - Q) + A_m\,Q) && (14)\\ &\Leftrightarrow \\ X_m &= \frac{A_m\,Q - A_r}{A_m - A_r}. && (15) \end{aligned}$$

Therefore, $X_m > 0$ if

$$A_m\,Q \;>\; A_r \quad \Leftrightarrow \quad q^N \;>\; \frac{A_r}{A_m} \tag{16}$$

Defining the selective superiority of the master by $s = A_m/A_r$ and using $q^N \sim \exp(-N(1-q))$, we get

$$\begin{aligned} -N(1-q) &> \ln\left(\frac{1}{s}\right) && (17)\\ &\Leftrightarrow \\ N &< \frac{\ln(s)}{1-q}. && (18) \end{aligned}$$

Equation (18) results in $N \sim 100$ for $s \sim e$ and $q < 0.99$.