

4.3.1 - RNA Hydrolysis

Parameters

Parameter Name	Variable	Default Value	Parameter Range	Description
FRAG_UR_D0	Unknown macro: 'mathinline'	1	>0	minimum length of fragme
FRAG_UR_DE LTA	Unknown macro: 'mathinline'	NaN ¹	Unknown macro: 'mathinline'	geometry of the fragment: 2=surface-diameter, 3=vo explicitly specified (NaN), depends logarithmically o
FRAG_UR_ETA	Unknown macro: 'mathinline'	NaN ¹	Unknown macro: 'mathinline'	intensity of fragmentation, breaks per unit length; if n Unknown m is determined by the corre Unknown m value and an expectation fragment size, if size sele

¹ NaN stands for "Not a Number" and marks the uninitialized state of a parameter

Algorithm

Frequencies

Unknown macro: 'mathinline'

 of fragment sizes

Unknown macro: 'mathinline'

 produced by a

Unknown macro: 'mathinline'

 uniform random fragmentation process have demonstrated to fall along Weibull distributions

Unknown macro: 'mathinline'

, if the fragmentation thermodynamics depends on the molecule size:

Unknown macro: 'mathblock'

Scale parameter

Unknown macro: 'mathinline'

 represents the intensity of fragmentation (i.e., breaks per unit length), and—as a determinant of the mean expected fragment size—is assumed to be constant across molecules of different lengths for fragmentation protocols where the number of produced fragments depends on the molecule length. Shape parameter

Unknown macro: 'mathinline'

 reflects the geometric relation in which random fragmentation is breaking a molecule (e.g.,

Unknown macro: 'mathinline'

 corresponds to uniform fragmentation on the linear chain of nucleotides,

Unknown macro: 'mathinline'

 splits uniformly the surface, and

Unknown macro: 'mathinline'

 the volume, etc.).

The Flux Simulator uses a 3-step algorithm to tokenize a molecule; first, geometry $\frac{L}{n}$ and the number of fragments that are obtained from the molecule are determined. We found empirically that parameter d depends logarithmically on L , the length of the molecule that is fragmented

The number of fragments produced from a specific RNA molecule is determined by n , where $\frac{L}{n}$ is the expectancy of the most abundant fragment size, computed from h and the gamma-function $\Gamma(h)$ of

$$n = \frac{L}{\Gamma(h)}$$

Second, n breakpoints are sampled uniformly from the interval $[0,1]$, resulting in relative length fractions $\frac{L_i}{L}$ for all i fragments. Third, relative fragment sizes $\frac{L_i}{L}$ are transformed from unit space to sizes L_i that follow a Weibull distribution of shape k by:

$$L_i = \frac{L}{n} \left(\frac{L_i}{L} \right)^{\frac{1}{k}}$$

where k is a constant of the transformation to ensure that the sizes of the fragments sum up exactly to the given molecule length

Latter transformation produces a slightly distorted Weibull distribution for the sizes L_i , however the deviation is sufficiently small to be neglected in our applications.