

4.3.1 - RNA Hydrolysis

Parameters

Parameter Name	Variable	Default Value	Parameter Range	Description
FRAG_UR_D0		1	>0	minimum length of fragments produced by hydrolysis
FRAG_UR_DELTA		NaN ¹		geometry of the fragmentation process (1=linear, 2=surface-diameter, 3=volume-diameter, etc.); if not explicitly specified (NaN), the geometry of breakage depends logarithmically on the molecule length
FRAG_UR_ETA		NaN ¹		intensity of fragmentation, determining the number of breaks per unit length; if not explicitly specified (NaN), is determined by the corresponding value and an expectation of 200nt (or the mean filtered fragment size, if size selection is used) long fragments

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

Algorithm

Frequencies of fragment sizes produced by a uniform random fragmentation process have demonstrated to fall along Weibull distributions , if the fragmentation thermodynamics depends on the molecule size:

Scale parameter 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 reflects the geometric relation in which random fragmentation is breaking a molecule (e.g., corresponds to uniform fragmentation on the linear chain of nucleotides, splits uniformly the surface, and the volume, etc.).

The Flux Simulator uses a 3-step algorithm to tokenize a molecule; first, geometry and the number of fragments that are obtained from the molecule are determined. We found empirically that parameter d depends logarithmically on , the length of the molecule that is fragmented . The number of fragments produced from a specific RNA molecule is determined by , where is the expectancy of the most abundant fragment size, computed from h and the gamma-function of :

Second, breakpoints are sampled uniformly from the interval [0;1], resulting in relative length fractions for all fragments. Third, relative fragment sizes are transformed from unit space to sizes that follow a Weibull distribution of shape by:

where 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 , however the deviation is sufficiently small to be neglected in our applications.