Abstract


Introduction

DNA and RNA are fundamental to biogenesis, nanotechnology, and artificial life. The RNA World hypothesis posits that self-replicating RNA emerged early from the primordial soup (Robertson and Joyce, 2012; Pressman et al., 2015), concentrated within amphiphilic micelles or hydrothermal vents (Joyce and Szostak, 2018). Many naturally-occurring ribozymes, riboswitches, and regulatory sequences provide circumstantial evidence for the RNA world (Kalvari et al., 2018). Further, experiments in vitro demonstrate the plausibility of RNA-like self-replicators (Tjhung et al., 2020), while simulations in silico recapitulate aspects of micelles (Boghosian et al., 1996; Coveney et al., 1996; Boghosian et al., 1999) and ribo-cells (Mavelli, 2012; Hutton, 2004, 2007). Nucleic acids are also powerful tools for nanotechnological AI, having been used to build molecular logic circuits (Penchovsky and Breaker, 2005) including a basic perceptron (Qian et al., 2011), data storage devices (Goldman et al., 2013), and programmable self-assembling materials (Rothemund, 2006).

In bioinformatics, simulations of RNA folding kinetics typically model secondary structure as an abstract graph of base-pair contacts (Flamm et al., 2000; Aviram et al., 2012; Xayaphoummine et al., 2003; Senter and Clote, 2015). Many biophysical parameters of these simulations can be measured experimentally (Lu et al., 2006) or calculated by statistical mechanics arguments (Isambert and Siggia, 2000). In general these approaches generally rely on approximate treatments of spatial phenomena such as the entropic cost of loop closure (Jacobson and Stockmayer, 1950). Since coordinates are not included, any local concentration must explicitly be built into the model.

Figure 1: A simple helix-hairpin RNA structure (left) and a corresponding configuration on the CARNAVAL lattice (right). Each node of the lattice corresponds to a single base or a base-pair; only complementary Watson-Crick (A:U, G:C) and wobble (G:U) base-pairs are allowed. Each step of the path corresponds to either to a link in the RNA polymer chain (for unpaired bases), or two antiparallel links (for paired bases). Each such link connects a node to an adjacent horizontal, vertical, or diagonal node (that is, the Moore neighborhood). Bases and base-pairs can diffuse by making single steps in the Moore neighborhood that preserve the above constraints. Diffusion moves are allowed to form (or break) base-pairs, with a Metropolis-Hastings rejection probability based on the change to the free energy. For example, the node labeled “CG” can diffuse left or right, as shown by arrows; alternatively, either the “C” or the “G” could split off to the left or right, forming an internal loop within the helix.

An alternative, more amenable to spatial structure while avoiding the cost of an all-atom simulation, is a coarse-grained lattice-based molecular dynamics that tracks the positions of individual bases (Bundschuh and Hwa, 1999; Leoni and Vanderzande, 2003; Zara and Prett, 2007; Pincus et al., 2008; Gillespie et al., 2009; Ding et al., 2008; Everaers et al., 2007; Jost and Everaers, 2010). Several such approaches model an RNA molecule as a “two-tolerant random walk” (Leoni and Vanderzande, 2003; Zara and Prett, 2007; Jost and Everaers, 2010). Bases occupy individual sites on the lattice, with two bases allowed to occupy the same site if (and only if) they form a basepair. Previous work has in-
vestigated square (Leoni and Vanderzande, 2003), triangular (Zara and Pretti, 2007) and face-centered cubic lattices (Jost and Everaers, 2010).

Here we describe a model of template-directed RNA self-replication based on a simplified two-tolerant RNA model on a cubic lattice, allowing diagonal steps (Figure 1). An advantage of this lattice is that it permits implementation via cellular automata where each cell is influenced by (and affects) only its Moore neighborhood. A useful benefit for pedagogic, artistic, and entertainment purposes is that this allows for accurate recapitulation of RNA folding dynamics in two dimensions (on a square lattice). The primary drawback of the Moore neighborhood (compared to the other lattices mentioned) is its weaker symmetry, which complicates calculations of bond rigidity and stiffness.

For the present experiments, we used a simplified energy-like function for basepair $x-y$ with the form $E = E_{xy} + sE_s$ where $s$ is the number of adjacent stacked basepairs with antiparallel linkage (0, 1 or 2). $E_{xy}$ is a base-pairing energy-like parameter that depends on the bases $x$ and $y$, and $E_s$ is a stacking energy-like parameter. A move is accepted with probability $\exp(-\Delta E/T)$ where $T$ is a temperature-like parameter. For the preliminary experiments described here, these parameters were set to $E_{AU} = -2, E_{GC} = 2, E_{GU} = -3, E_s = 4, T = 1$ except where otherwise specified. We omitted rigidity or stiffness terms, involving (respectively) bond angles or extension, although these can readily be introduced (albeit requiring new parameters, more so for weakly-symmetric lattices).

Early computational experiments on the base-pairing and drift properties of this model suggest that, despite its considerable simplicity compared to earlier models of RNA folding kinetics, the model efficiently explores conformational space and reproduces features of RNA folding (Figure 2).

To develop this model for RNA replication, rather than just folding, we add to the diffusion rules a reaction rule that allows covalent bond formation between adjacent bases bound to the same complementary strand. Seeded with a population of monomeric units and a single template strand, this allows for template-driven polymerization (Figure 3).

As proofs of concept we implemented several versions of this model, including two prototyped reaction-diffusion game engines, and one implementation optimized for scientific investigation of folding and self-replication kinetics, named CARNAVAL (Cellular Automata RNA for Virtual A-Life). The results reported here use the latter, available at https://github.com/evoldoers/carnaval.

Preliminary experiments with CARNAVAL reveal that the fidelity of RNA self-replication under this model is highly sensitive to the energy model (G-U wobble basepairs induce mismatches); replication is more accurate in the limit $E_{GU} \to -\infty$. The model naturally generates plasmid-like circularized RNAs, without this being explicitly specified in the design. Our energy model’s lack of a bond rigidity term means that some of these circular RNAs are implausibly short (the shortest are 3 bases) suggesting that bond angles should be an important feature of future experiments.

When combined with lattice gas models of amphiphilic micelles (Boghosian et al., 1996), models such as this may deepen the connection to physical first principles of A-Life simulations of the primordial ribo-cell, as described by Hutton (2004, 2007).

Acknowledgements

This work was supported by NIH grant GM080203 and inspired by Richard Evans, Chris Mungall, and Rudy Rucker.
References


