

Stochastic models of Gene–Replicase–Translatase Systems

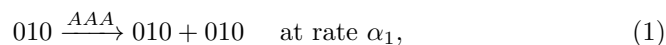
Alex Poppinga and Alexei Drummond

March 5, 2018

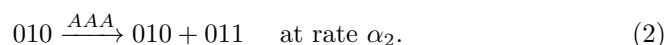
1 The GRT Model

Genes are described as binary sequences of length L . Then for a small system of $L = 3$, the set $G = \{000, 001, 010, 011, \dots, 111\}$ contains 8 possible genes. An individual gene is denoted $g \in G$. Proteins are also described as binary sequences of length L , but we distinguish them by using a unique alphabet such that $P = \{AAA, AAB, ABA, ABB, \dots, BBB\}$, and an individual protein is $p \in P$. In this primitive system we imagine a simple translation mechanism in which one ‘nucleotide’ position in a gene is translated to one ‘amino acid’ in a protein. In other words, a codon is also represented by a single bit, and therefore a gene and its translated protein are always the same length.

In this system, proteins have varying levels of catalytic activity for replication and translation. A replication reaction can proceed either without error:

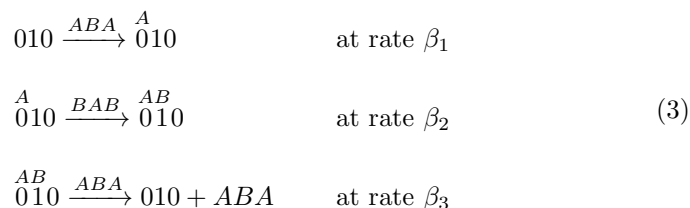


or with error:



(We assume the food set $F = \{0, 1\}$ contains an infinite supply of elements “0” and “1” and is an implicit component of the lefthand side of eq. 1,2.)

Translation reactions are modelled in a stepwise fashion so that partial translation yields intermediate complexes that are explicit species in the system:

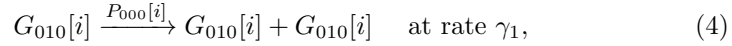


Above, ABA is a translatase that translates the codon ‘0’ to the amino acid ‘A’ and BAB is a translatase that translates the codon ‘1’ to the amino acid ‘B’

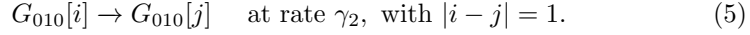
at the specified rates. As with the replication reactions, translation reactions proceed at higher rates for particular specifications but also have some rate at which an alternative outcome occurs (e.g., ABA translates the codon ‘0’ to the amino acid ‘ B ’ rather than ‘ A ’).

1.1 Diffusion representation

We make a small change of notation to introduce a spatial component for diffusion reactions, where space is described by a linear array of cells, indexed by i . The replication reaction introduced in eq. 1 is now represented in cell i as:

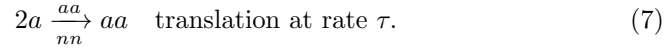
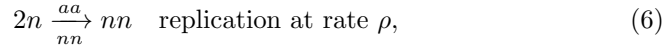


and now a diffusion reaction from cell i to cell j can also occur, such that:

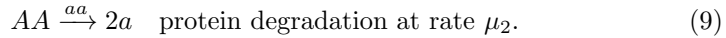
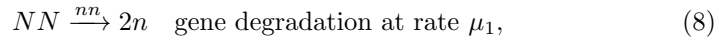


2 A simplified monopolymer model

To understand some of the basic dynamics we first consider a simplified monopolymer model of length $L = 2$. We will also assume that translation occurs perfectly in a single step, so that no intermediate species are represented in the system. Finally, we will explicitly describe food sources as nucleotide and amino acid monomers. We therefore have two symmetric replication and translation reactions, where in both the nucleotide polymer nn is the catalytic template and the amino acid polymer aa is the catalytic enzyme:



We additionally introduce two density-dependent degradation rates to prevent the system from growing without bound:



The change in concentration of the two polymer species ($m = nn$ and $p = aa$) and their two food sources (a and n) can be described by the following system of ordinary differential equations (ODEs):

$$\dot{m} = \rho m p n^2 - \mu_1 m, \quad (10)$$

$$\dot{p} = \tau m p a^2 - \mu_2 p, \quad (11)$$

$$\dot{n} = 2\mu_1 m, \quad (12)$$

$$\dot{a} = 2\mu_2 p. \quad (13)$$

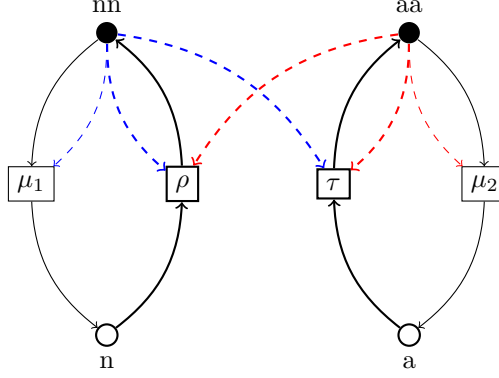
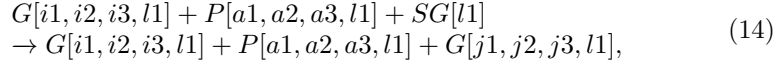


Figure 1: A replication-translation RAF. $\{a, n\}$ is the food set. $\{a, n, aa, nn\}$ is the set of species.

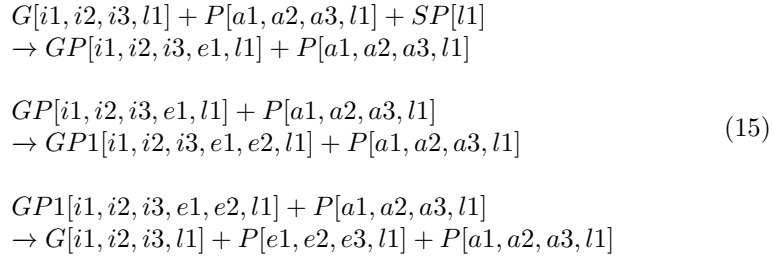
This model is a RAF in Steel's classification of chemical reaction systems and is depicted in his canonical form in Figure 1.

3 MASTER representation of full model

To input our model in MASTER [Vaughan and Drummond, 2013] we add indices to represent the sequence and location of each species. For example, the replication reaction in eq. 1 is now represented as:



with $i1, i2, i3$ (replicated with some small error to $j1, j2, j3$) denoting the sequence of the species and $l1$ denoting the cell that contains the species. Genes occupy space SG within the cell; proteins occupy SP . The translation reactions in eq. 3 are now represented as:



with additional indices $e1$ and $e2$ to denote either the 'A' or 'B' amino acid forming the partially translated gene-protein complexes GP and $GP1$. Each of these reactions has a rate specified within the reaction element. MASTER also

allows the option of applying a rate multiplier, which can be useful for adjusting the rate at which a particular reaction occurs based off of, for example, Hamming distances. To add a diffusion reaction in MASTER XML a predicate function is used to determine which species will be moved to which location based off of the location indices of the species. In the following example, the diffusion space is characterised by an annulus:

```
<reaction spec='Reaction' reactionName="Diffusion_G" rate="0.1">
  <predicate spec='Predicate' value="abs(l1-l2)==1 || abs(l1-l2)==G_dim[3]"/>
  G[i1,i2,i3,l1] + SG[l2] -> G[i1,i2,i3,l2] + SG[l1]
</reaction>
```

4 Preliminary simulation results in MASTER

For the symmetrical monopolymer model, the populations of amino acid and nucleotide food molecules and the 'NN' gene and 'AA' protein quickly reached an equilibrium, as shown in Figure 2. This also appeared to be true for the 'perfect GRT' model (with 3 binary $L = 3$ gene species and 3 corresponding binary proteins with $L = 3$), which behaves as a RAF in that no error is generated by either the replication or translation reactions. (In fact, it should be noted that this first implementation of the minimal GRT model in MASTER is actually a CAF, a stronger notion of RAF that has a food set - here including genes and proteins along with nucleotides and amino acids - which acts as catalysts immediately in the system).

However, when only nucleotides and amino acids were initially present in the system (and simulation time was extended), it was observed that one of the 3 gene species quickly stabilized while another became predominate and the third went to zero (occasionally being produced in small quantities by random food-generated-only reactions), as shown in Figure 3.

Replacing explicit food molecules (i.e., nucleotide and amino acid species) with spaces - in which the genes occupy a single space S , the gene-protein complexes each a single space S , and the proteins occupy three spaces ($3S$) - caused a similar behaviour but distinctive behaviour. In this case, one of the 3 gene species clearly began to dominate while the other two began to die out and that eventually the whole system collapsed (Figure 4). Increasing the simulation time again showed that all populations clearly went to zero.

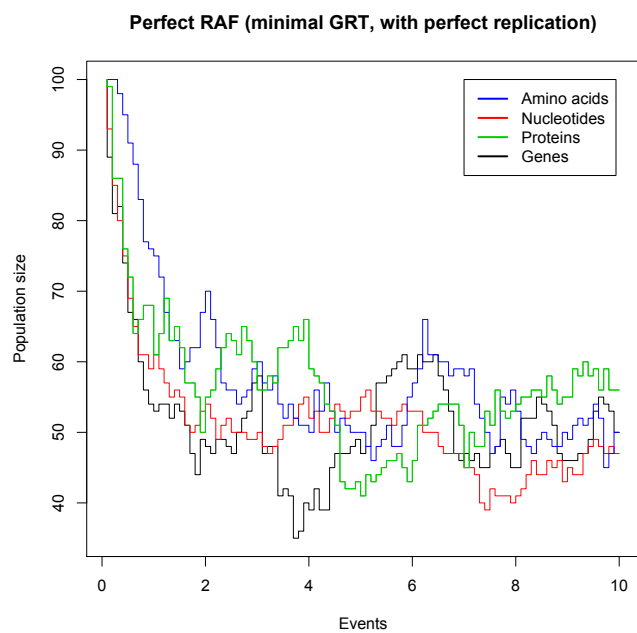
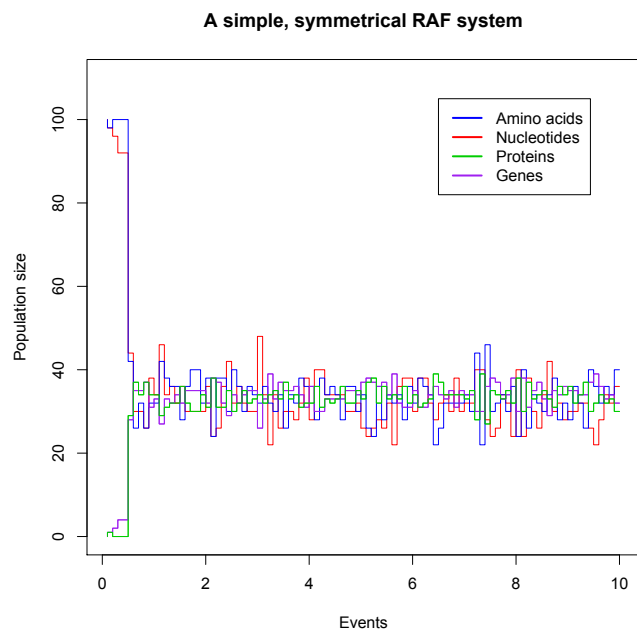


Figure 2: (top) The simplified monopolymer model with equal initial population sizes for the food set (amino acids and nucleotides). (bottom) The ‘perfect GRT’ model with no reaction errors and with all species present in the initial conditions.

'Perfect RAF' (Minimal GRT w/ 3 gene species & no rxn error)

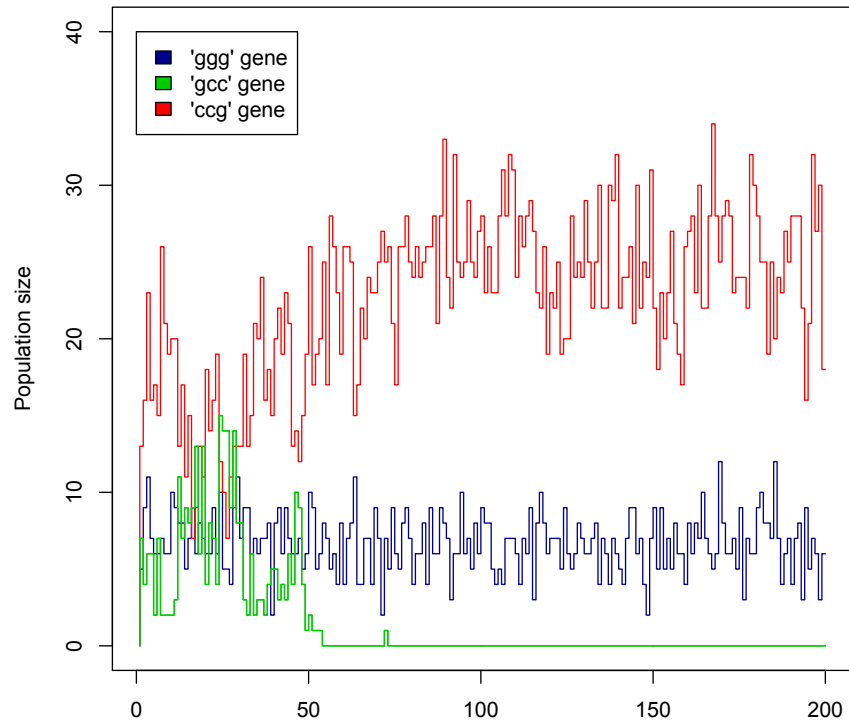


Figure 3: The 'perfect GRT' model with no reaction errors and with only the food set $F = \{g, c, a, b\}$ as part of the initial conditions, (where g and c are nucleotides and a and b are amino acids).

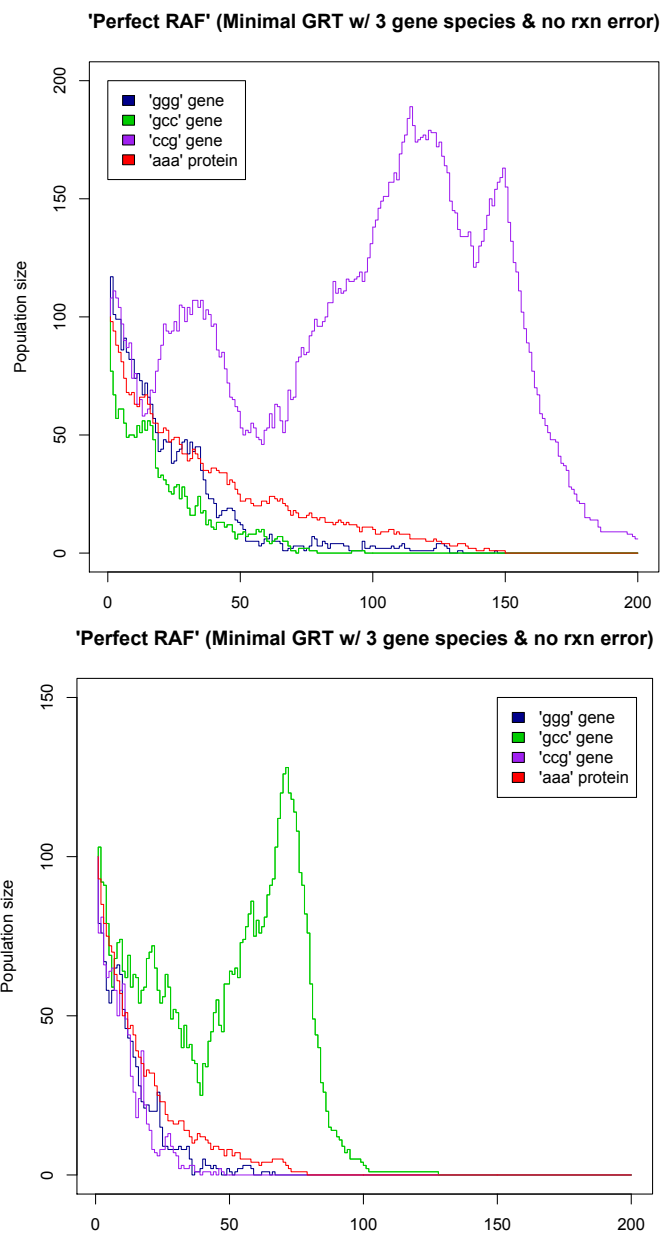


Figure 4: (top) The 'perfect GRT' model when the explicit food set was replaced with spaces S , with genes occupying $1S$ and proteins occupying $3S$.

References

- [Vaughan and Drummond, 2013] VAUGHAN, T. G. and A. J. DRUMMOND, 2013. A Stochastic Simulator of BirthDeath Master Equations with Application to Phylodynamics. *Mol. Biol. Evol.*