Effect of the SOS Response on the Mean Fitness of Unicellular Populations: A Quasispecies Approach

Amit Kama, Emmanuel Tannenbaum*
Department of Chemistry, Ben-Gurion University of the Negev, Beer-Sheva, Israel

Abstract
The goal of this paper is to develop a mathematical model that analyzes the selective advantage of the SOS response in unicellular organisms. To this end, this paper develops a quasispecies model that incorporates the SOS response. We consider a unicellular, asexually replicating population of organisms, whose genomes consist of a single, double-stranded DNA molecule, i.e. one chromosome. We assume that repair of post-replication mismatched base-pairs occurs with probability θ, and that the SOS response is triggered when the total number of mismatched base-pairs is at least κ. We further assume that the per-mismatch SOS elimination rate is characterized by a first-order rate constant kSOS. For a single fitness peak landscape where the master genome can sustain up to l mismatches and remain viable, this model is analytically solvable in the limit of infinite sequence length. The results, which are confirmed by stochastic simulations, indicate that the SOS response does indeed confer a fitness advantage to a population, provided that it is only activated when DNA damage is so extensive that a cell will die if it does not attempt to repair its DNA.

Introduction
Genetic repair is an essential component of cellular genomes. Without mechanisms for repairing damaged and mutated DNA, genomes could not achieve sufficient information content to code for the variety and complexity of modern terrestrial life [1]. Genetic repair mechanisms fall into two main categories: Those that correct base mis-pairings during the replication cycle of a cell, and those that repair mutated and damaged DNA during the growth (G) phase of the cellular life cycle [1].

Two important examples of the first class of repair mechanisms are DNA proofreading and mismatch repair (MMR). DNA proofreading is a repair mechanism that is built into the DNA replicases themselves. During daughter strand synthesis, an erroneously matched base is excised, and a second attempt at a base pairing is made [1]. Mismatch repair also removes erroneous bases from the daughter strand, but does this shortly after daughter strand synthesis [1].

Two important examples of the second class of repair mechanisms are Nucleotide Excision Repair (NER) and the SOS response [1]. NER protects a cell from damage due to radiation, chemical mutagens, and metabolic free radicals by removing damaged portions of the DNA strand and using the other, presumably undamaged strand as a template for re-synthesis of the excised region [1].

The SOS response is a genomic repair mechanism that only activates when there is extensive damage to the cellular genome. When DNA damage is sufficiently extensive, the cell stops growing, and the SOS repair pathways attempt to restore complementarity to the genome [1]. The SOS response only takes effect when DNA damage is so extensive that it may be impossible to use undamaged template strands to correctly re-synthesize damaged portions of the genome. Thus, although this means that the SOS repair mechanism is highly error prone, it is evolutionary advantageous for the cell to repair the genome and risk fixing deleterious mutations, than it is to leave the damaged genome unrepaired [1–4].

In recent work with quasispecies models of evolutionary dynamics, quasispecies models [5–7] considering the first class of repair mechanisms have been studied [8–11]. In addition, semiconservative replication, including semiconservative replication with imperfect lesion repair (i.e. not all base-pair mismatches are eliminated), has been considered [12–15]. Additional effects, such as multiply-genomed genomes, as well as multiply chromosomed genomes, have been considered as well [16,17].

This paper continues the theme of incorporating various details characteristic of cellular genomes by developing a quasispecies model that takes into consideration the SOS repair mechanism. The model is highly simplified, and therefore only a first step in developing proper dynamical equations with SOS repair. Nevertheless, because our model is analytically tractable, we believe it is a useful and important initial approach to mathematically modeling the evolutionary aspects of the SOS repair pathway. A proper modeling of the SOS response is an essential component of developing a quantitative theory of mutation-propagation in cellular organisms, which is important for understanding phenomena such as the emergence of antibiotic drug resistance in bacteria, and cancer in multicellular organisms [2–4].

Materials and Methods
Definitions and Model Set-Up
We consider a unicellular population of asexually replicating organisms, whose genomes consist of a single DNA molecule, i.e.
one chromosome. The genome may then be denoted by \( \{ \sigma, \sigma' \} \), where \( \sigma, \sigma' \) denote the two strands of the DNA molecule. If the genome is of length \( L \), then we may write \( \sigma = b_1 \cdots b_L \), \( \sigma' = b'_1 \cdots b'_L \) where each base \( b_i, b'_i \) is chosen from an alphabet of size \( S \) (usually \( S = 4 \)). If \( b_i \) denotes the base complementary to \( b_i \) (for the standard Watson-Crick bases, the pairings are Adenine(A) \( \rightarrow \) Thymine(T), Guanine(G) \( \rightarrow \) Cytosine(C)), and \( s \) denotes the strand complementary to \( \sigma \), then \( \sigma = b_1 \cdots b_L \). This follows from the antiparallel nature of double-stranded DNA [1].

In our simplified model, the SOS response is triggered if a given genome \( s \) has more than \( l \) mismatches. The replication rate of all cells undergoing SOS repair is zero. We assume that removal of mismatches is catalyzed by an enzyme that binds to a mismatch and removes mismatches at a rate characterized by a first-order rate constant, denoted \( k_{\text{SSM}} \). The set of all \( k_{\text{SSM}} \) defines the fitness landscape. It should be emphasized that fitness in this model only refers to replication rate. In particular, this model does not consider cell death.

The semiconservative replication of the DNA genomes happens in three stages:

1. Strand separation, whereby each strand of the chromosome separates to act as a template for daughter strand synthesis.
2. Daughter strand synthesis. We assume a genome and base-independent mismatch probability \( \varepsilon \). This error probability \( \varepsilon \) includes all error correction mechanisms, such as proofreading and mismatch repair, that are active during the replication phase of the cell.
3. Lesion repair, where any post-replication mismatches are removed. Here, there is no longer the parent-daughter strand discrimination that was available during daughter strand synthesis, so in contrast to DNA proofreading and mismatch repair, lesion repair has a 50% chance of removing the mutation, and a 50% chance of communicating it to the parent strand and fixing the mutation in the genome (lesion repair can occur via either Base or Nucleotide Excision Repair) [1].

In our simplified model, the SOS response is triggered if a given genome has at least \( l^\text{SS} \) mismatches. The replication rate of all cells undergoing SOS repair is zero. We assume that removal of mismatches is catalyzed by an enzyme that binds to a mismatch and then eliminates the mismatch at a rate characterized by a first-order rate constant \( k_{\text{SSM}} \). Therefore, the probability that a given genome is viable is given by \( k_{\text{SSM}} \int d\tau \), if it has fewer than \( l \) mismatches, and if it does not differ from \( \{ \sigma_0, \sigma_0 \} \) by any fixed mutations. Otherwise, the genome is unviable, with a first-order growth rate constant of \( 1 \). We recognize that this terminology is somewhat inappropriate, since a genome with a first-order growth rate constant of \( 1 \) may still replicate. However, this is standard terminology from quasispecies theory, where "viable" and "unviable" are taken to be synonymous with "higher fitness" and "lower fitness" respectively.

The justification for this choice of fitness landscape is as follows: if a genome has a fixed mutation, then neither DNA strand corresponds to either of the master strands \( \sigma_0, \sigma_0 \). As a result, the genome does not contain all of the information corresponding to a viable organism, hence the organism is unviable. While this assumption is clearly extreme and oversimplified, it is the analogue of the single-fitness-peak landscape for single-stranded genomes.

However, in the case of a mismatch where one of the bases is the same as the corresponding base in one of the master strands, the information contained in the master genome is still preserved in one of the strands, so that the organism is assumed to remain viable if the total number of such mismatches does not exceed some cutoff value \( l \).

For convenience, Table 1 summarizes the main parameters of the model.

**Symmetrized Population Distribution**

We can develop the infinite sequence length equations for our model, assuming an initially prepared clonal population consisting entirely of the wild-type (mutation-free) genome \( \{ \sigma_0, \sigma_0 \} \), i.e. a population consisting entirely of the fastest replicating genotype. Because, during replication, only a finite number of mutations are possible, at any time time the population will consist of a population of genomes \( \{ \sigma, \sigma' \} \) where \( \sigma, \sigma' \) differ from each \( \sigma_0 \) and \( \sigma_0 \) in at most a finite number of spots. Thus, given two gene sequences \( \sigma_1, \sigma_2 \), if we let \( D_H(\sigma_1, \sigma_2) \) denote the Hamming distance [19] between \( \sigma_1 \) and \( \sigma_2 \) (i.e. the number of sites where \( \sigma_1 \) and \( \sigma_2 \) differ), then either \( D_H(\sigma, \sigma_0) \) and \( D_H(\sigma', \sigma_0) \) are finite, or \( D_H(\sigma, \sigma_0) \) and \( D_H(\sigma', \sigma_0) \) are finite.
The Hamming distance between two sequences \( \sigma_1 \) and \( \sigma_2 \) is simply equal to the number of positions by which they differ. It may be readily shown that the Hamming distance is a metric over the space of sequences [19], so that, in particular, the Hamming distance satisfies the Cauchy-Schwartz Inequality: Given three sequences \( \sigma_1, \sigma_2, \) and \( \sigma_3 \), we have \( D_H(\sigma_1, \sigma_2) + D_H(\sigma_2, \sigma_3) \geq D_H(\sigma_1, \sigma_3) \).

Now, in the limit of infinite sequence length, it may be shown that, with probability 1, the Hamming distance between \( \sigma_0 \) and its complement \( \sigma_0' \) is infinite [12]. Therefore, if \( D_H(\sigma, \sigma_0) \) and \( D_H(\sigma, \sigma_0') \) were both finite, we would obtain \( \infty = D_H(\sigma, \sigma_0) \leq D_H(\sigma, \sigma_0') + D_H(\sigma, \sigma_0) < \infty \Rightarrow \infty \), and so \( \sigma \) cannot simultaneously be of finite Hamming distance to \( \sigma_0 \) and \( \sigma_0' \). Similarly, \( \sigma' \) cannot simultaneously be of finite Hamming distance to \( \sigma_0 \) and \( \sigma_0' \).

As a result, we can define a strand ordering (\( \sigma, \sigma' \)) for a genome \( (\sigma_0, \sigma_0') \), where it is understood that \( \sigma \) is a finite Hamming distance from \( \sigma_0 \) and \( \sigma' \) is a finite Hamming distance from \( \sigma_0' \).

A given genome \( (\sigma_0, \sigma_0') \) may then be characterized by four parameters \( \ell_c, \ell_L, \ell_R, \) and \( \ell_B \). We let \( \ell_c \) denote the number of sites where \( \sigma \) and \( \sigma' \) are both complementary, yet differ from the corresponding bases in \( \sigma_0 \) and \( \sigma_0' \). We let \( \ell_L \) denote the number of sites where \( \sigma \) differs from \( \sigma_0 \), but \( \sigma' \) is identical to \( \sigma_0 \). We let \( \ell_R \) denote the number of sites where \( \sigma \) is identical to \( \sigma_0 \), but \( \sigma' \) differs from \( \sigma_0 \). Finally, we let \( \ell_B \) denote the number of sites where \( \sigma \) and \( \sigma' \) differ from \( \sigma_0 \) and \( \sigma_0' \), but are not complementary (for an illustration of these parameters, see [7,13]).

Note that the fitness landscape depends only on \( \ell_c, \ell_L, \ell_R, \) and \( \ell_B \), and hence the fitness of a given organism may be denoted by \( k_{(\ell_c, \ell_L, \ell_R, \ell_B)} \), where for our single-fitness-peak landscape we have \( k_{(\ell_c, \ell_L, \ell_R, \ell_B)} = k \) if \( \ell_c = 0 \) and \( \ell_L + \ell_R + \ell_B \leq l \), and 1 otherwise. The condition \( \ell_c = 0 \) means that there are no mutations fixed in the genome, while the condition \( \ell_L + \ell_R + \ell_B \leq l \) means that there are fewer than \( l \) lesions.

By the symmetry of the fitness landscape, and by the symmetry of the initial population distribution, we can group all genomes of identical \( \ell_c, \ell_L, \ell_R, \) and \( \ell_B \), and derive the dynamical equations of the symmetrized population distribution. We therefore let \( n_{(\ell_c, \ell_L, \ell_R, \ell_B)} \) denote the total number of organisms in the population whose genomes are characterized by the parameters \( \ell_c, \ell_L, \ell_R, \) and \( \ell_B \), and we let \( n_{(\ell_c, \ell_L, \ell_R, \ell_B)} \) denote the total number of organisms in the population undergoing the SOS response, whose genomes are similarly characterized by the parameters \( \ell_c, \ell_L, \ell_R, \) and \( \ell_B \). The corresponding population fractions are denoted \( z_{(\ell_c, \ell_L, \ell_R, \ell_B)} \) and \( z_{(\ell_c, \ell_L, \ell_R, \ell_B)} \), respectively.

### Dynamical Equations

To develop the dynamical equations for both the \( z_{(\ell_c, \ell_L, \ell_R, \ell_B)} \) and the \( z_{(\ell_c, \ell_L, \ell_R, \ell_B)} \) quantities, we begin by considering a genome \( (\sigma_0, \sigma_0') \), characterized by the parameters \( \ell_c, \ell_L, \ell_R, \) and \( \ell_B \).

We first consider the case where this genome is not undergoing the SOS response. Then, due to the semi-conservative nature of DNA replication, this genome is being destroyed at a rate given by \(-k_{(\ell_c, \ell_L, \ell_R, \ell_B)} n_{(\ell_c, \ell_L, \ell_R, \ell_B)} \). This genome, however, is produced by other genomes in the population, as a result of replication. So, consider some other genome \( (\sigma', \sigma'') \) which produces \( (\sigma_0, \sigma_0') \) upon replication. This can either occur via the \( \sigma' \) template strand, the \( \sigma'' \) template strand, or both.

If the \( (\sigma', \sigma'') \) genome is characterized by the parameters \( \ell_c', \ell_L', \ell_R', \) and \( \ell_B' \), then \( \sigma'' \) differs from \( \sigma_0 \) in \( \ell_c' + \ell_L' + \ell_R' + \ell_B' \) bases. Because sequence lengths are infinite, the probability of a mismatch in one of these bases during daughter strand synthesis is 0. In the remaining sites, let \( \ell_1' \) denote the number of mismatches that are not corrected, and \( \ell_2' \) denote the number of mismatches that are repaired, but fixed as a mutation in the genome. Then the resulting genome \( (\sigma', \sigma'') \) is characterized by:

1. \( \ell_c = \ell_c' + \ell_L' + \ell_R' + \ell_B' + \ell_2' \)
2. \( \ell_c = 0 \)
3. \( \ell_R = \ell_1' \)
4. \( \ell_B = 0 \)

The probability of a given set of mutations corresponding to \( \ell_1' \), \( \ell_2' \), is \( e^{\ell_1' + (1-\epsilon)/2} e^{-(\ell_1' + \ell_2')} \). The term \( 1 - e^{\ell_1' + \ell_2'} \) arises as a probability that the remaining \( l - \ell_c' - \ell_L' - \ell_R' - \ell_B' - \ell_1' - \ell_2' \) sites on \( \sigma' \) remain identical to \( \sigma_0 \), and the corresponding daughter strand sites are identical to \( \sigma_0' \). The per-site probability of this is the probability of error-free daughter strand synthesis, \( 1 - e^{-\epsilon} \), plus the probability of a mismatch, \( \lambda \), the probability that complementarity is restored, times \( 1/2 \), the probability that complementarity is restored correctly. It is assumed that complementarity is restored by various DNA repair mechanisms, such as Nucleotide Excision Repair (NER) and Base Excision Repair (BER) [1]. However, because NER and BER do not distinguish between parent and daughter strands, the probability of correctly removing a mismatch via these mechanisms is \( 1/2 \).

The degeneracy is given by \((L - \ell_c' - \ell_L' - \ell_R' - \ell_B')!/(\ell_1'!\ell_2'!(L - \ell_c' - \ell_L' - \ell_R' - \ell_B' - \ell_1' - \ell_2')!))\), in the limit of infinite sequence length the total probability becomes,

\[
\frac{(L - \ell_c' - \ell_L' - \ell_R' - \ell_B')!}{\ell_1'!\ell_2'!(L - \ell_c' - \ell_L' - \ell_R' - \ell_B' - \ell_1' - \ell_2')!} \times e^{\ell_1' + \ell_2'} \left(1 - \tilde{\lambda}\right) \frac{\tilde{\lambda}^{\ell_1'}}{\ell_1'} (1 - (1 - \epsilon)/2) e^{-(\ell_1' + \ell_2')} \to \frac{1}{\ell_1!\ell_2!} \left[\mu(1 - \tilde{\lambda})\right]^{\ell_1'+\ell_2'} \frac{\mu^{\ell_1'}}{2} e^{-(1 - 1/2)\mu}
\]

If \( (\sigma, \sigma') \) is generated by \( \sigma'' \), then we have,

1. \( \ell_c = \ell_c' + \ell_R' + \ell_B' + \ell_2' \)
We also obtain an overall transition probability of 

\[ \frac{1}{(l'_L l'_C)\mu/2} e^{-\mu(1-\lambda)/2}. \]

It is important to note from the \( \sigma' \) and \( \sigma'' \) results that genomes with \( l_R > 0 \) cannot be generated during replication. Since SOS repair eliminates mismatches, it follows that a population where \( l_R = 0 \) is initially 0 for all genomes will always have a population where \( l_R = 0 \). Therefore, we may assume in subsequent derivations that \( l_R, l'_R \) are 0.

Furthermore, note that strands \( \sigma' \) that are a finite Hamming distance away from \( \sigma_0 \) can only generate daughter genomes where \( l'_L = 0 \), while strands \( \sigma'' \) that are a finite Hamming distance away from \( \sigma_0 \) can only generate daughter genomes where \( l_R = 0 \).

Then for the genomes \( \{\sigma, \sigma'\} \) generated by \( \sigma'' \), we have \( l'_L = l'_C + l'_L + l'_R \), and \( l_R = l'_R \). Therefore, the restriction on \( \{\sigma', \sigma''\} \) is that \( 0 \leq l'_L \leq l'_C \), \( 0 \leq l'_R \leq l'_C - l'_L \), and \( l'_R = l'_C - l'_L - l'_R \). Note that there is no restriction on \( l'_R \).

Then for the population number \( n(l'_C, 0, l'_L, l'_R) \), we have a contribution from the \( \sigma' \) strands of

\[
\frac{1}{l'_R} \left[ \mu(1-\lambda) \right]^{l'_R} e^{-\mu(1-\lambda)/2} \times \sum_{l'_C=0}^{l'_C} \sum_{l'_L=0}^{l'_L} \sum_{l'_R=0}^{l'_R} \left( \frac{\mu l'_C}{2} \right)^{l'_C} \left( \frac{\mu l'_L}{2} \right)^{l'_L} \left( \frac{\mu l'_R}{2} \right)^{l'_R} \times K(l'_C - l'_L - l'_R, 0)^l(l'_C - l'_L - l'_R, 0) 
\]

A similar expression is obtained for the population number \( n(l'_C, l'_L, 0, l'_R) \), except \( l'_R \) is replaced with \( l'_L \), and the roles of \( l'_L \) and \( l'_R \) are exchanged.

It should also be noted that, by the symmetry of the fitness landscape, we have that \( n(l'_C, l'_L, l'_R) = n(l'_L, l'_C, l'_R) \). Another way to note this is that, for a given genome \( \{\sigma, \sigma'\} \), if we change the ordering of the strands so that the first strand is of finite Hamming distance to \( \sigma_0 \), and the second strand is of finite Hamming distance to \( \sigma_0 \), then the genome \( \{\sigma, \sigma'\} \) must be represented as \( \{\sigma', \sigma\} \), and is characterized by the parameters \( l'_C, l'_L, l'_R, l'_L \). If \( n(l'_C, l'_L, l'_R) \) denotes the number of genomes characterized by \( l'_C, l'_L, l'_R, l'_L \), with respect to the \( \{\sigma_0, \sigma_0\} \) strand ordering, then since there is a one-to-one correspondence between genomes \( \{\sigma, \sigma'\} \) with parameters \( l'_C, l'_L, l'_R, l'_L \) with respect to the first ordering, and genomes \( \{\sigma', \sigma\} \) with parameters \( l'_C, l'_L, l'_R, l'_L \) with respect to the second ordering, it follows that \( n(l'_C, l'_L, l'_R) = n(l'_L, l'_C, l'_R) \). However, since the fitness landscape is invariant under strand ordering, we have \( n(l'_C, l'_L, l'_R) = n(l'_L, l'_C, l'_R) \).

Taking into consideration the contribution to \( n(l'_C, 0, 0, l'_R) \), we may put everything together and obtain, after changing variables from population numbers to population fractions, the differential equations governing the time evolution of the various population fractions. These equations are,

\[
\frac{dz_{l'_C, 0, 0, l'_R}}{dt} = -(\kappa_{l'_C, 0, 0, l'_R}) z_{l'_C, 0, 0, l'_R} + \kappa_{l'_C, 0, l'_R} z_{l'_C, 0, l'_R} + 2e^{-\mu(1-\lambda)/2} \sum_{l'_C=0}^{l'_C} \sum_{l'_L=0}^{l'_L} \sum_{l'_R=0}^{l'_R} \left( \frac{\mu l'_C}{2} \right)^{l'_C} \left( \frac{\mu l'_L}{2} \right)^{l'_L} \left( \frac{\mu l'_R}{2} \right)^{l'_R} K(l'_C - l'_L - l'_R, 0) z_{l'_C - l'_L - l'_R, 0} z_{l'_C - l'_L - l'_R, 0} \]

Similarly, we obtain

\[
\frac{dz_{l'_C, 0, 0, l'_R}}{dt} = -(\kappa_{l'_C, 0, 0, l'_R}) z_{l'_C, 0, 0, l'_R} + \kappa_{l'_C, 0, l'_R} z_{l'_C, 0, l'_R} + 2e^{-\mu(1-\lambda)/2} \sum_{l'_C=0}^{l'_C} \sum_{l'_L=0}^{l'_L} \sum_{l'_R=0}^{l'_R} \left( \frac{\mu l'_C}{2} \right)^{l'_C} \left( \frac{\mu l'_L}{2} \right)^{l'_L} \left( \frac{\mu l'_R}{2} \right)^{l'_R} K(l'_C - l'_L - l'_R, 0) z_{l'_C - l'_L - l'_R, 0} z_{l'_C - l'_L - l'_R, 0} \]

where \( K(t) \) is the Kronecker delta function, so that \( \delta_{ij} = 1 \) if \( i = j \), and 0 otherwise.

Note that we do not write down the dynamical equations for \( z_{l'_C, 0, 0, 0} \) or \( z_{l'_C, 0, 0} \), since they are redundant.
The factor of 1/2 appearing in the SOS terms arises from the fact that when a mismatch is removed, it either corrects the daughter strand synthesis error, or it fixes the mismatch as a mutation in the genome. In the former case, the value of $l_c$ remains unchanged, while in the latter case it is incremented by 1.

It should be noted that this factor is missing in the contribution to $z(\ell_c,0,0,0)$ from SOS repair. The reason for this is that this contribution comes from $z(\ell_c,0,1,0)$ and $z(\ell_c,1,1,0)$ and $z(\ell_c-1,1,0,0)$. However, because $z(\ell_c,0,0,0)$, $z(\ell_c,1,0,0)$, and $z(\ell_c-1,0,1,0)$, we may combine identical terms and eliminate the factor of 1/2.

The factor of $l+1$ in the front of the $\kappa_{SOS}$ rate constant arises from the fact that the fraction of genomes whose SOS enzymes are bound to a mismatch is proportional to the total number of mismatches, hence the resulting SOS rate constant is proportional to the total number of mismatches.

**Results and Discussion**

**Steady-State Behavior**

**Definitions and basic equations.** To obtain the steady-state behavior of our model, we begin by introducing some definitions that will allow us to simplify the calculations.

1. $z_1 = z(0,0,0)$.
2. $z_2 = \sum_{l+1}^{\infty} z(l,0,0)$.
3. $z_3 = \sum_{l+1}^{\infty} z(1,l,0)$.
4. $z_4 = \sum_{l+1}^{\infty} z(l,1,0)$.
5. $z_5 = \sum_{l+1}^{\infty} z(1,0,l)$.
6. $z_6 = \sum_{l+1}^{\infty} z(0,1,l)$.
7. $z(0,0,0)$.
8. $z(l)$.
9. $z(l)$.
10. $z(0,0,0)$.

where we set $l = l_c - 1$ whenever $l$ was previously defined as $l \geq l_c$. The differential equations for $z_1$, $z_2$, $z_3$, $z_4$, $z_5$, and $z_6$ are readily derived. From the equations,

$$\sum_{l=0}^{\infty} \kappa_{SOS}(0,0,0) z(l,0,0) = k z_1 + k z_2 + z_3$$

and

$$\sum_{l=0}^{\infty} \sum_{h=0}^{l} \sum_{i=0}^{h} \frac{1}{h!} \mu^h \frac{1}{2^h} = \kappa_{SOS}(0,0,0) \sum_{l=0}^{\infty} \sum_{h=0}^{l} \sum_{i=0}^{h} \frac{1}{h!} \mu^h \frac{1}{2^h}$$

we obtain,

$$\frac{dz_1}{dt} = -(k + \kappa(t)) z_1 + 2 e^{-\mu (1-\lambda)/2} [k z_1 + k z_2 + z_3]$$

$$\frac{dz_2}{dt} = -(k + \kappa(t)) z_2 + \frac{1}{7} (f_l(1,0,0) - f_l(0,0,0)) e^{-\mu (1-\lambda)/2} [k z_1 + k z_2 + z_3]$$

$$\frac{dz_3}{dt} = -(1 + \kappa(t)) z_3 + \frac{1}{7} (f_l(1,0,0) - f_l(0,0,0)) e^{-\mu (1-\lambda)/2} [k z_1 + k z_2 + z_3]$$

$$\frac{dz_4}{dt} = -(1 + \kappa(t)) z_4 + \frac{1}{7} (f_l(1,0,0) - f_l(0,0,0)) e^{-\mu (1-\lambda)/2} [k z_1 + k z_2 + z_3]$$

$$\frac{dz_5}{dt} = -(1 + \kappa(t)) z_5 + \frac{1}{7} (f_l(1,0,0) - f_l(0,0,0)) e^{-\mu (1-\lambda)/2} [k z_1 + k z_2 + z_3]$$

$$\frac{dz_6}{dt} = -(1 + \kappa(t)) z_6 + \frac{1}{7} (f_l(1,0,0) - f_l(0,0,0)) e^{-\mu (1-\lambda)/2} [k z_1 + k z_2 + z_3]$$

We also have,

$$\frac{d z_l}{dt} = \kappa_{SOS} \frac{l + 1}{2} z_{l+1} - (k + \kappa(t)) z_l$$

for $l = 1, \ldots, l_S - 1$.
We can add these equations to obtain,
\[
\frac{dz_{(SOS)}}{dt} = -\kappa_{SOS}z_{(SOS)}^{(SOS)} - k(t)z_{(SOS)}^{(SOS)} + \left(1 - e^{-\mu(1/\lambda)}f_{S_{(-1)}}^{(SOS)}\right)\times [kz_1 + 2kz_2 + 2z_3 + z_4 + 2z_5 + 2z_6].
\] (8)

For the purposes of computing the mean fitness at steady-state, we can simplify the system of equations somewhat by defining \(z_4 = z_4 + 2z_5 + 2z_6\). We obtain,
\[
\frac{dz_{(SOS)}}{dt} = -(1 + k(t))z_4 + 2e^{-\mu(1/\lambda)}f_{S_{(-1)}}^{(SOS)}\times [e^{\mu/2}(kz_1 + 2kz_2 + 2z_3 + z_4) - (kz_1 + kz_2 + z_3)] + \kappa_{SOS} [2z_{(SOS)}^{(SOS)}]^{-1} \]
\] (9)

For consistency of notation, in what follows we shall simply let \(z_4\) denote \(z_4\).

**Determining the population fractions \(z_{(SOS)}^{SOS_{(SOS)}}\), \(z_{(SOS)}^{SOS_{(SOS)}}\), and \(z_{(SOS)}^{SOS_{(SOS)}}\).** To obtain the steady-state behavior of this system of equations, we begin by first solving for the steady-state of the population undergoing SOS repair.

For \(l = 0, \ldots, l_s - 1\) we have at steady-state that,
\[
\frac{SOS_{(SOS)}}{SOS_{(SOS)}} = \frac{2}{l + 1} \left(1 + \frac{k(t) = \infty}{\kappa_{SOS}}\right)z_{(SOS)}^{SOS_{(SOS)}}
\] (10)

which gives,
\[
z_{(SOS)}^{SOS_{(SOS)}} = \frac{2l_{s-1}^{(SOS)} - 1}{l_{s-1}} \prod_{i=1}^{l_{s-1}} \left(1 + \frac{k(t) = \infty}{\kappa_{SOS}}\right)z_{(SOS)}^{SOS_{(SOS)}}
\] (11)

For \(l \geq l_s\), we have,
\[
\frac{SOS_{(SOS)}}{SOS_{(SOS)}} = \frac{2}{l + 1} \left(1 + \frac{k(t) = \infty}{\kappa_{SOS}}\right)z_{(SOS)}^{SOS_{(SOS)}}
\]
\[
- \frac{2}{\kappa_{SOS}} \left(1 + l_{s} + 1\right) \mu_{SOS} (1/\lambda) \mu_{SOS} (1/\lambda) \times e^{-\mu(1/\lambda)}(kz_1 + kz_2 + z_3)
\] (12)

This expression has the form of the recursion relation, \(x_{n+1} = a_nx_n \rightarrow b_n\). Using mathematical induction, it is possible to prove that \(x_0 = a_0 \times \ldots \times a_0 = a_0 \times \ldots \times a_1b_0 - a_{n-1} \times \ldots \times a_2b_1 = a_0 = a_0 \times \ldots \times a_1b_0 - a_{n-1} \times \ldots \times a_2b_1 = a_0\). Therefore,
\[
z_{(SOS)}^{SOS_{(SOS)}} = \frac{2^{l_{s-1}}}{l_{s}} \prod_{l=1}^{l_{s-1}} \left(1 + \frac{k(t) = \infty}{\kappa_{SOS}}\right)z_{(SOS)}^{SOS_{(SOS)}}
\]
\[
- \frac{2}{\kappa_{SOS}} e^{-\mu(1/\lambda)}(kz_1 + kz_2 + z_3)
\]
\[
\times \left[\left(1 + l_{s} + 1\right) \mu_{SOS} (1/\lambda) \mu_{SOS} (1/\lambda) \right]
\] (13)

where we define \(\Pi_{l=1}^{l_{s-1}} a_l = 1\).

If we define \(g_t(\mu, \lambda; \bar{k}(t = \infty), \kappa_{SOS}) = \Pi_{l=1}^{l_{s-1}} \frac{\mu_{SOS} (1/\lambda)}{l + 1 + \kappa_{SOS}}\)
\[
\sum_{k=0}^{l_{s}} \frac{\mu_{SOS} (1/\lambda)}{l + 1 + \kappa_{SOS}} \]
\[
\frac{\mu_{SOS} (1/\lambda)}{l + 1 + \kappa_{SOS}}
\][limit to \(\gamma_{SOS}^{SOS_{(SOS)}} = 0\) gives, at steady-state, that,
\[
\kappa_{SOS}^{SOS_{(SOS)}} = 2e^{-\mu(1/\lambda)}(kz_1 + kz_2 + z_3) \times g_{SOS_{(SOS)}} (\mu, \lambda; \bar{k}(t = \infty), \kappa_{SOS})
\] (14)

Using a similar argument, we obtain,
\[
\kappa_{SOS}^{SOS_{(SOS)}} = e^{-\mu(1/\lambda)}(kz_1 + 2kz_2 + 2z_3 + z_4) \times g_{SOS_{(SOS)}} (\mu, \lambda; \bar{k}(t = \infty), \kappa_{SOS})
\] (15)

For the steady-state value of \(z_{(SOS)}^{SOS_{(SOS)}}\), we have, using the identity \(\bar{k}(t) = kz_1 + 2kz_2 + 2z_3 + z_4\),
\[
z_{(SOS)}^{SOS_{(SOS)}} = 1 - e^{-\mu(1/\lambda)} \times \left(1 + g_{SOS_{(SOS)}} (\mu, \lambda; \bar{k}(t = \infty), \kappa_{SOS}) \right)
\] (16)

**Computing the steady-state mean fitness \(\bar{k}(t = \infty)\).** Plugging our expressions for \(\kappa_{SOS}^{SOS_{(SOS)}}\) and \(\kappa_{SOS}^{SOS_{(SOS)}}\) into the steady-state population fractions equations, we obtain,
\[
0 = -(k + \bar{k}(t = \infty))z_1
\]
\[
+ 2e^{-\mu(1/\lambda)} \left(1 + g_{SOS_{(SOS)}} (\mu, \lambda; \bar{k}(t = \infty), \kappa_{SOS}) \right) \times (kz_1 + kz_2 + z_3)
\]
\[
0 = -(k + \bar{k}(t = \infty))z_2
\]
\[
+ (f_{S_{(-1)}}(\mu, \lambda) - f_{S}(\mu, \lambda)) e^{-\mu(1/\lambda)}(kz_1 + kz_2 + z_3)
\]
\[
0 = -(k + \bar{k}(t = \infty))z_3
\]
\[
+ (f_{S_{(-1)}}(\mu, \lambda) - f_{S}(\mu, \lambda)) e^{-\mu(1/\lambda)}(kz_1 + kz_2 + z_3)
\]
\[
0 = -(k + \bar{k}(t = \infty))z_4
\]
\[
+ 2e^{-\mu(1/\lambda)} \left(f_{S_{(-1)}}(\mu, \lambda) + g_{SOS_{(SOS)}} (\mu, \lambda; \bar{k}(t = \infty), \kappa_{SOS}) \right) \times (kz_1 + 2kz_2 + 2z_3 + z_4)
\]
\[
- 2e^{-\mu(1/\lambda)} \left(f_{S_{(-1)}}(\mu, \lambda) + g_{SOS_{(SOS)}} (\mu, \lambda; \bar{k}(t = \infty), \kappa_{SOS}) \right) \times (kz_1 + kz_2 + z_3)
\] (17)

From these equations we may derive the equality,
\[
k(z_1 + z_2) + z_3
\]
\[
= [k(z_1 + z_2) + z_3] e^{-\mu(1/\lambda)}
\]
\[
\times k + 2g_{SOS_{(SOS)}} (\mu, \lambda; \bar{k}(t = \infty), \kappa_{SOS}) + f_{S}(\mu, \lambda)
\]
\[
+ f_{S_{(-1)}}(\mu, \lambda) - f_{S}(\mu, \lambda)
\]
\[
\times \frac{1 + 2g_{SOS_{(SOS)}} (\mu, \lambda; \bar{k}(t = \infty), \kappa_{SOS}) + f_{S}(\mu, \lambda)}{k + \bar{k}(t = \infty)}
\] (18)

Below the error catastrophe, when \(z_1, z_2, z_3\) are not all 0, we may cancel \(k(z_1 + z_2) + z_3\) from both sides of the equation and re-arrange to obtain,
\[
\mathbf{\bar{K}}(t=\infty)^2 - A(\mu; \bar{K}, \mathbf{K}_{SOS}) \mathbf{\bar{K}}(t=\infty) - B(\mu; \bar{K}, \mathbf{K}_{SOS}) = 0 \tag{19}
\]

where,
\[
A(\mu; \bar{K}; \mathbf{K}(t=\infty), \mathbf{K}_{SOS}) = k\left[ e^{-\mu(1-\lambda)} \left( 1 + f_1(\mu, \bar{K}) + 2g_1 \left( \frac{k}{2}, \bar{K}(t=\infty), \mathbf{K}_{SOS} \right) \right) - 1 \right] + e^{-\mu(1-\lambda)} f_1(-1, \mu, \bar{K}) - 1 \tag{20}
\]

\[
B(\mu; \bar{K}; \mathbf{K}(t=\infty), \mathbf{K}_{SOS}) = k\left[ e^{-\mu(1-\lambda)} \left( 1 + f_1(\mu, \bar{K}) + 2g_1 \left( \frac{k}{2}, \bar{K}(t=\infty), \mathbf{K}_{SOS} \right) \right) - 1 \right] \]

Beyond the error catastrophe, the mutation rate is sufficiently high that the selective advantage for remaining localized about the \(k_c = 0\) genomes disappears, so that \(z_1, z_2, z_3\) drop to 0. The relevant steady-state equation is then,
\[
0 = -(1 + \mathbf{K}(t=\infty))z_4 + 2e^{-\mu(1-\lambda)} \left( f_{S-1}(\mu, \bar{K}) + g_{S}(\mu, \bar{K}(t=\infty), \mathbf{K}_{SOS}) \right)z_4
\tag{21}
\]

which may be solved for \(\mathbf{K}(t=\infty)\) to give,
\[
\mathbf{K}(t=\infty) = 2e^{-\mu(1-\lambda)} \left( f_{S-1}(\mu, \bar{K}) + g_{S}(\mu, \bar{K}(t=\infty), \mathbf{K}_{SOS}) \right) - 1
\tag{22}
\]

The error catastrophe occurs at the mutation rate for which the two expressions for the mean equilibrium fitness become equal. As with previous quasispecies models, the error catastrophe here also corresponds to a localization to delocalization transition over sequence space [5-7].

**Limiting Cases.** We now proceed to consider the behavior of the steady-state mean fitness for a number of limiting cases, in order to better understand our model. We consider the following cases: (1) \(\lambda = 1\), corresponding to perfect lesion repair, so that there are non-complementary genomes in the population. (2) \(I_S = \infty\), corresponding to the case where no genome ever undergoes the SOS response. (3) \(K_{SOS} \rightarrow \infty\), corresponding to the case where SOS repair happens rapidly, so that there is no negligible fitness penalty associated with undergoing the SOS response.

**Case 1:** \(\lambda = 1\). When \(\lambda = 1\), we get for \(I_S > 0\) that \(g_{S}(\mu, \bar{K}(t=\infty), \mathbf{K}_{SOS}) = 0\), and that \(f_{S-1}(\mu, \bar{K}) = 1\). Therefore, above the error catastrophe, we obtain \(\mathbf{K}(t=\infty) = 1\). Below the error catastrophe, we have \(A(\mu; \bar{K}, \mathbf{K}(t=\infty), \mathbf{K}_{SOS}) = k(2e^{-\mu/2} - 1)\), \(B(\mu; \bar{K}(t=\infty), \mathbf{K}_{SOS}) = k(2e^{-\mu/2} - 1)\), giving \(\bar{K}(t=\infty) = k(2e^{-\mu/2} - 1)\). These results are in agreement with the solution of the semiconservative quasispecies equations with perfect lesion repair [12].

**Case 2:** \(I_S = \infty\). When \(I_S = \infty\), then \(g_{S}(\mu, \bar{K}(t=\infty), \mathbf{K}_{SOS}) = 0\). Below the error catastrophe, we have \(A(\mu; \bar{K}(t=\infty), \mathbf{K}_{SOS}) = k(2e^{-\mu/2} - 1)\), \(B(\mu; \bar{K}(t=\infty), \mathbf{K}_{SOS}) = k(2e^{-\mu/2} - 1)\), and \(B(\mu; \bar{K}, \mathbf{K}_{SOS}) = k(2e^{-\mu/2} - 1)\), \(B(\mu; \bar{K}, \mathbf{K}_{SOS}) = k(2e^{-\mu/2} - 1)\), \(B(\mu; \bar{K}, \mathbf{K}_{SOS}) = k(2e^{-\mu/2} - 1)\). Above the error catastrophe, we have \(\bar{K}(t=\infty) = 1\). Both results are in agreement with axial lesion repair efficiency [13].
maximum selective advantage if it is activated when and only when the genome has sustained sufficient genetic damage so that it will be unviable without SOS repair. However, it should be emphasized that this only holds when \( \mu \) is not near the error catastrophe, so that \( \phi \) is sufficiently larger than 1 for large \( k \) that the above analysis holds.

**Stochastic Simulations**

We developed stochastic simulations of a unicellular population capable of undergoing the SOS response, in order to numerically test the analytical predictions of our model. We consider a constant population of genomes that is cycled over every time step. During each cycle, every genome is allowed to replicate with a probability \( k_{[\sigma,\sigma']} \Delta t \), where \( k_{[\sigma,\sigma']} \) is the first-order growth rate constant of genome \( \{\sigma,\sigma'\} \), and \( \Delta t \) is the length of the time step. We take \( \Delta t \) to be sufficiently small so that the probability of a given genome replicating more than once during a cycle is negligible.

We assume that the population initially consists of a clonal population of wild-type (mutation-free) genomes. The fitness of a given genome \( \{\sigma,\sigma'\} \) is determined by assigning \( l_C, l_L, l_R, l_B \) parameters to the ordered-pairs \( (\sigma,\sigma'), (\sigma',\sigma) \) with respect to the ordered-pair \( (\sigma_0,\sigma_0) \). For each set of \( l_C, l_L, l_R, \) and \( l_B \) parameters, a fitness is assigned based on the fitness landscape defined previously. The fitness of the genome is then taken to be the larger of the two calculated fitnesses. In the limit of infinite sequence length, this prescription for calculating fitnesses becomes identical to the method used in the analytical solution of our model.

If a genome replicates during a cycle, then it is removed from the population, and the two daughters are added to the population of genomes. To maintain a constant population size, another, randomly chosen genome is removed from the population as well. Because this approach is simply the stochastic implementation of the quasispecies dynamics of the system, it converges to the infinite population, continuous time result as the population size gets larger and the time steps get smaller.

If a daughter genome is produced that has at least \( l_S \) lesions, then it enters the SOS response, and is assigned a replication probability of 0. A genome that has initiated the SOS response continues to undergo SOS repair until all lesions have been removed, and a complementary genome has been restored. During every time step, a genome that is undergoing the SOS response has its lesions scanned, and each lesion is repaired with probability \( k_{SOS} \Delta t \). In addition to being chosen small enough so that the probability of a given genome replicating more than once during a cycle is negligible, we also choose \( \Delta t \) to be sufficiently small so that the probability that a given genome undergoing the SOS response has more than one lesion repaired during a cycle is also negligible.

The stochastic simulation is allowed to run for a sufficient number of time steps so that the mean fitness of the population does not change significantly, at which point the system is assumed to be at steady-state.

Figures 2 and 3 show plots comparing the mean fitness obtained from the analytical solution to the mean fitness obtained from the stochastic simulations. As can be seen from the figures, the agreement between the analytical solution and the stochastic simulation is excellent.

**Conclusions and Future Research**

This paper developed a quasispecies approach for describing the evolutionary dynamics of a unicellular population that incorporated a simplified model of the SOS response. The model was a generalization of the single-fitness-peak landscape that is often used in quasispecies theory to study various problems in evolutionary dynamics. The model was shown to be analytically solvable, and it was found that the solution led to a maximal selective advantage to the SOS response in a manner that is broadly consistent with the behavior of actual organisms. Specifically, we showed that the SOS response should only be activated in a cell with a sufficiently damaged genome that it will be unviable if the SOS response is not activated. In such a situation, the cell has “nothing to lose,” meaning that it is better to attempt to repair the genome and risk introducing deleterious mutations, than it is to leave a highly damaged genome alone.

![Figure 2. Comparison of the mean fitnesses obtained from both stochastic simulations (dots) and the analytical solution (solid line) of our model. Parameter values are \( k = 9, l = 4, l_S = 5, \lambda = 0.08, k_{SOS} = 100, L = 100 \). The population size was set at 1,000.](https://doi.org/10.1371/journal.pone.0014113.g002)
For future research, it will be important to consider more realistic models that will allow for quantitative models that can be used in collaboration with experiment. Because the SOS response is a genetic repair pathway that works in conjunction with other cellular repair pathways, a proper understanding of the SOS response is important for developing a coherent theory of mutation-propagation that will be useful for understanding the emergence of antibiotic drug resistance in bacteria, and cancer in multicellular organisms [2–4].

Author Contributions
Conceived and designed the experiments: EDT. Performed the experiments: EDT. Analyzed the data: AK EDT. Contributed reagents/materials/analysis tools: AK EDT. Wrote the paper: AK EDT.

References
1. Voet D, Voet JG (2004) Biochemistry, 3rd ed. Hoboken, NJ: John Wiley and Sons, Inc.
2. Reaber JW, Hochhut B, Waldor MK (2004) SOS response promotes horizontal dissemination of antibiotic resistance genes. Nature (London) 427: 72–74.
3. Miller C, Thomsen LE, Gaggero C, Mosseri R, Ingmer H, et al. (2004) SOS response induction by β-lactams and bacterial defense against antibiotic lethality. Science 305: 1629–1631.
4. McKenzie GJ, Harris RS, Lee PL, Rosenberg SM (2004) SOS response induction by
5. Bull JJ, Meyers LA, Lachmann M (2005) Quasispecies made simple. PLoS Computational Biology 1: e61.
6. Wilke CO (2005) Quasispecies theory in the context of population genetics. BMC Evolutionary Biology 5: 44.
7. Tannenbaum E, Shakhnovich EI (2005) Semiconservative replication, genetic repair, and many-genome paradigms. Proceedings of the Life Reviews 2: 290–317.
8. Tannenbaum E, Deeds EJ, Shakhnovich EI (2003) Equilibrium distribution of mutators in the single-fitness-peak model. Physical Review Letters 91: 130105.
9. Tannenbaum E, Shakhnovich EI (2004) Error and repair catastrophes: A two-dimensional phase diagram in the quasispecies model. Physical Review E 69: 011902.
10. Sasaki A, Nowak MA (2003) Mutation Landscapes. The Journal of Theoretical Biology 224: 241–247.
11. Kessler DA, Levine H (1998) Mutator dynamics on a smooth evolutionary landscape. Physical Review Letters 80: 2012–2015.
12. Tannenbaum E, Deeds EJ, Shakhnovich EI (2004) Semiconservative replication in the quasispecies model. Physical Review E 69: 061916.
13. Tannenbaum E, Shakhnovich EI (2004) Imperfect DNA lesion repair in the semiconservative quasispecies model. Derivation of the Hamming class equations and solution of the single-fitness-peak landscape. Physical Review E 70: 061915.
14. Brumer Y, Shakhnovich EI (2004) Host-parasite co-evolution and optimal mutation rates for semiconservative quasispecies. Physical Review E 69: 061903.
15. Brumer Y, Shakhnovich EI (2004) Importance of DNA repair in tumor suppression. Physical Review E 70: 061912.
16. Tannenbaum E, Shakhnovich EI (2004) Solution of the quasispecies model for an arbitrary gene network. Physical Review E 70: 021903.
17. Tannenbaum E, Sherley JL, Shakhnovich EI (2006) Semiconservative quasispecies equations for polysomic genomes: The haploid case. The Journal of Theoretical Biology 241: 791–805.
18. Nilsson M, Snaud N (2000) Error thresholds for quasispecies on dynamic fitness landscapes. Physical Review Letters 84: 191–194.
19. Ash RB (1990) Information Theory. Dover Publications.