NON-EQUILIBRIUM IMPACTS ON EVOLVABILITY

AIDAN ETTERER

Abstract. New developments in non-equilibrium thermodynamics have applications to questions in evolutionary biology. One of these developing ideas involves the occasional tendency for a biological structure to reset its development in order to save time and maximize entropy, called dynamic instability. Here we extensively analyze dynamic instability in a toy model of protein folding. This process could help create less diversity within a population such that mutated, advantageous traits spread more easily. In this paper, mathematical methods are used to create and explore possible models of dynamic instability, and how it modifies the impact of mutations on the evolvability of certain traits or structures.

Contents

1. Biophysical Motivation

In evolutionary biology, there have long been open questions about the exact processes by which complex structures come about. Every necessary protein has many dozens of inter-working components such that, if even a single one were lost, it would drastically hinder the overall function. How could something like this, where having only most (but not all) of the parts is worthless, evolve organically? Even if an organism is able to manifest part of the folding through mutations, the half-folded-protein would not benefit the organism in any way, and so it would be no more likely to pass on its genes to offspring than any other individual of its same species that also lacks the completed correct protein.

Alternatively, if an organism already has the information to create said protein, but not necessarily the information on how to create it well, the same problem arises. If some components are correct and others are not, then the execution on the folding may suffer significantly. With just a few interchanged amino acids a

Date: AUGUST 28, 2024.

protein will fold completely differently. Even if there was a mutation that would fix one component of the structure, if all the other components were wrong it would have almost no impact on the fitness of that particular protein within its host organism.

The status quo could change, however, if some amount of instability was introduced to the protein's development. Imagine that some foldings of the protein took longer to complete than others, and that if a long enough time threshold had passed without a finished structure, the development would reset to the beginning, so that every choice becomes available again.

How could this reset infrastructure have come about? In some biological processes, the time taken to assemble a structure is an attribute (just as the actual function of the structure is) that natural selection may act on. This is seen to be true for kinetic proofreading, where some trade-off of increased error rate is made to lower the time taken to complete the process [2]. The reason for the trade-off is quite intuitive: if the proofreading takes too much time, it slows the growth of the organism, which is disadvantageous for it evolutionarily. This same act of selection would have to occur for reset mechanisms to take hold in the folding of proteins: cutting off pathways that have too much time cost.

The presence (or possibility) of a reset mechanism is an assumption of all models analyzed here. The groundwork for this reset, or "dynamic instability," has been researched in [1]. This paper proposes the existence of a multi-component structure whose creation involves a time cost to its host organism, which could motivate the presence of dynamic instability. The physics behind this reset are complex, occurring via non-equilibrium dynamics, but it can be witnessed in micro-tubules: they self-assemble, but participate in occasional large-scale-disassembly. The authors of [1] noticed a correlation between slow-to-complete structures and disordered[1] structures, that correlation being the latter identity causing the former, and they conjectured that selecting for fast completion times would simultaneously select for ordered structures. The paper concluded that resets could be advantageous in eliminating time costs if two conditions were met: first that the time cost of the slow trajectories must be very large, and second that the probability of taking those trajectories must be quite small. These conditions are required such that, on average, resetting slow-built structures is more beneficial than allowing them to complete.

Our analysis continues the focus on a multi-component structure, but looks to the consequences of the reset mechanism on the evolutionary process, and how it amplifies the impact of otherwise insignificant mutations. Further, we make assumptions about the fitness of different foldings such that the conclusions of [1] are not necessarily applicable here.

2. One-Correct-Structure Model

In this section, we explore various generalizations of one concept: a structure with a single correct development, and many incorrect developments. We then investigate how changing the probability of achieving the correct development impacts the fitness and evolvability of that structure.

[1]Here "order" just means adherence to a specific design of the structure, i.e. with high order there is little variation in the assembly of the components.

2.1. **2-Step Toy Model.** Imagine a gene as if it guides an organism to choosing one of n possible options, only 1 of which could possibly lead to future survival and reproduction. If an organism has k different genes G_i, then any of n^k paths may be taken by an organism, but where $\prod_{i=1}^{k} P_i$ is the probability of the single correct path being taken. In this subsection we will work with $k = 2$.

Assumptions 2.1. This toy model of protein folding is both similar and distinct from the other models we will mention. Its simplicity and effectiveness at demonstrating the way reset infrastructure modifies the impacts of resets makes it very useful to explore. The simplicity comes from the toy-protein we are analyzing, which is constructed from only two decision points, each with n options. This model's physicality relies on a number of assumptions:

- Our protein folding process is determined by just 2 genes: G_1 and G_2. Each gene has n possible ways to carry out its function, only one of which is correct. This gives us n^2 possible outcomes, only one of which is correct. These two genes are each given some probability of doing their respective step of the process correctly, P_1 and P_2.
- The fitness of the protein folding is determined entirely by the probability that the correct folding is reached, which requires that both genes make the correct decision.
- A mutation to the ith gene changes the probability that the ith step of the folding process is done correctly: a mutation to G_i changes P_i.
- In absence of resets: the correct folding always finishes at time $\varepsilon_t \ll 1$, whereas incorrect foldings take some time according to the Gaussian distribution in Figure 1b to finish. This presumed correlation between "fast" and "correct" foldings is an exaggerated application of the results from [1].
- Some proportion S of the $n^2 - 1$ incorrect paths will be reset to the beginning, where both genes will then be able to make new decisions, independently of the decisions they made previously.
- The best mutation for a protein (and its host organism) is that which maximizes the change in fitness caused by the mutation, or $\Delta_{[mut]_i} F$.

We may now assume that "incorrect" paths generally take much more time to complete than that of the "correct" path, that being $\varepsilon_t \ll 1$, and the time taken to completion for the incorrect paths follows from some nearly-Gaussian distribution. If we then add a "reset" mechanism that resets at some threshold time, $T > \varepsilon_t$ and resets some proportion S of incorrect paths, then we can model the fitness F, for a system of 2 genes G_1 and G_2, as follows:

Definition 2.2. The fitness, F is defined as according to our current model:

$$\begin{aligned} F &= \sum_{m=0}^{\infty} \left[P_1 P_2 \ \times \underbrace{(1 - P_1 P_2)^m S^m}_{\text{proportion of paths to be reset on the } m\text{th reset}} \right] \\ &= \underbrace{\frac{P_1 P_2}{1 - S + S P_1 P_2}}_{\text{probability that folding is correct, given reset proportion } S} \end{aligned} \tag{2.3}$$

where $F|_{S=0} = P_1 P_2$ as one would expect, given that F is equal to the probability that the correct outcome is reached and that without any reset function, $P_1 P_2$ is that probability.

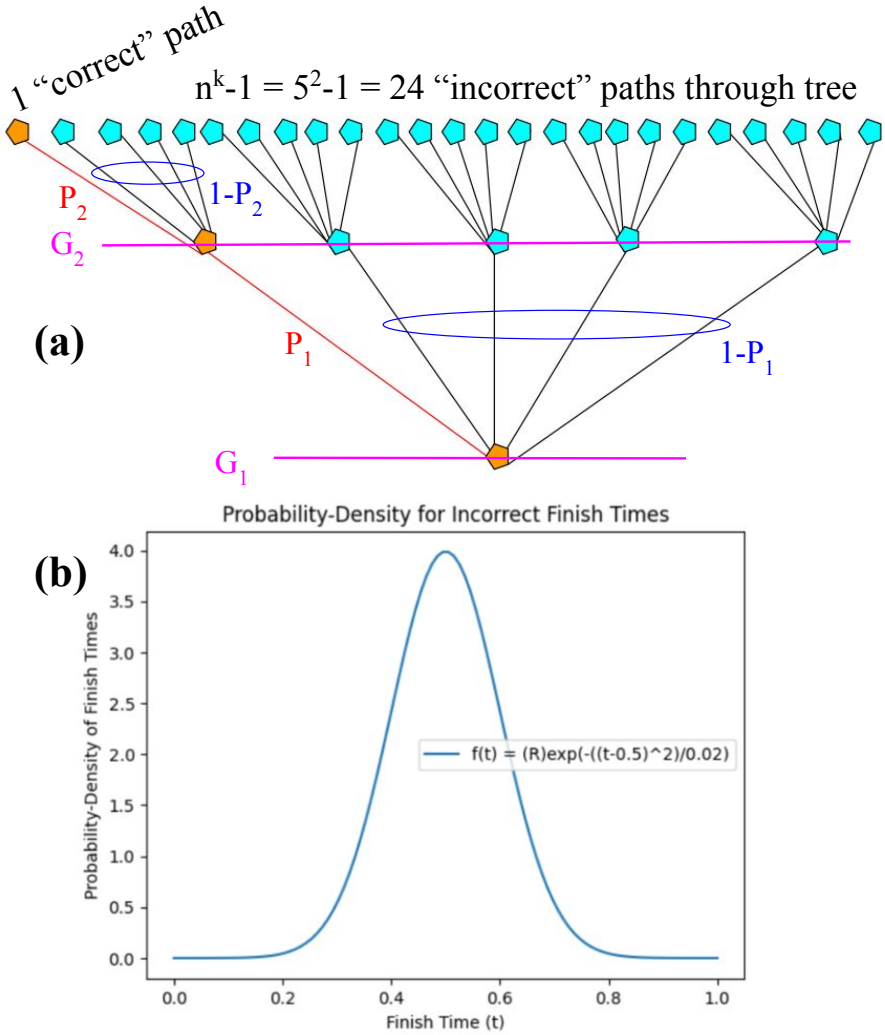

FIGURE 1. (a) A schematic of the model, where the fitness is equal to the probability of going down the correct path, which takes time $\varepsilon_t \ll 1$. The $n^2 - 1$ incorrect pathways that do not contribute to fitness, each take some time as according to a roughly Gaussian distribution. (b) The distribution of times to completion for incorrect paths, a nearly-Gaussian distribution with mean 0.5 and standard deviation 0.1, only defined on $[\varepsilon_t, \infty)$, and normalized so that the integral across the whole domain is unity. The horizontal axis is the finish time, and the vertical axis is the probability density function for incorrect pathway finish times (so that the integral is just the proportion). A normal distribution is chosen because of our assumption that incorrect finish times should roughly follow a uni-modal distribution, and the Gaussian is the most reasonable uni-modal non-singular distribution.

Definition 2.4. The proportion of incorrect protein foldings that get reset to the beginning, S, is defined using the distribution in Figure 1b:

$$S = S(T) = R\int_T^\infty \exp\left(-\frac{(t-0.5)^2}{0.02}\right) dt \quad \text{where } R = \frac{1}{\int_{\varepsilon_t}^\infty \exp\left(-\frac{(t-0.5)^2}{0.02}\right) dt} \tag{2.5}$$

and where T is used to parameterize S in terms of our distribution of incorrect finish times. Now instead of saying that some proportion S of incorrect structures are reset, we can say that given our distribution of finish times, all the finish times above the threshold time T are reset.

The times to completion for incorrect paths have been assigned a Gaussian distribution with mean 0.5 and standard deviation 0.1. A correction was then made to ensure that $S(\varepsilon_t) = 1$, keeping with our physical expectation that if the threshold for a pathway reset was just over $t = \varepsilon_t$, every incorrect pathway would be reset. The correct pathway will never be reset because its time taken ε_t is negligible in comparison to the threshold time, T. This model, with (2.3), yields an equation for fitness in terms of T:

$$F(T, P_1, P_2) = \frac{P_1P_2}{1 + (P_1P_2 - 1)R\int_T^\infty \exp\left(-\frac{(t-0.5)^2}{0.02}\right) dt}. \tag{2.6}$$

Note that we can take $\lim_{T\to\infty} F = P_1P_2$, which is the same as saying that when $S = 0$ (no reset), $F = \prod_{i=1}^k P_i$ where, for our case, $k = 2$.

We can now consider how an infinitesimal change to one of the genes, and consequently its associated probability, impacts the fitness:

$$\frac{\partial F}{\partial P_1} = \frac{P_2(1 - R\int_T^\infty \exp\left(-\frac{(t-0.5)^2}{0.02}\right) dt)}{(1 + (P_1P_2 - 1)R\int_T^\infty \exp\left(-\frac{(t-0.5)^2}{0.02}\right) dt)^2},$$
$$\frac{\partial F}{\partial P_2} = \frac{P_1(1 - R\int_T^\infty \exp\left(-\frac{(t-0.5)^2}{0.02}\right) dt)}{(1 + (P_1P_2 - 1)R\int_T^\infty \exp\left(-\frac{(t-0.5)^2}{0.02}\right) dt)^2}. \tag{2.7}$$

Definition 2.8. To describe the effect of a mutation that changes one of our genes G_i and its probability P_i by some $\Delta P_i = P_i^{mut} - P_i^{WT}$, we can use the notation of $\frac{\Delta F}{\Delta[mut]_i}$: where P_i^{mut} denotes the probability of the correct choice being selected for the ith gene after mutation, and P_i^{WT} denotes the wild-type (pre-mutation) probability for the same gene G_i.

$$\frac{\Delta F}{\Delta[mut]_i} = \Delta_{[mut]_i}F = F(P_i^{mut}) - F(P_i^{WT}) = \int_{P_i^{WT}}^{P_i^{mut}} \frac{\partial F}{\partial P_i} dP_i. \tag{2.9}$$

From (2.6), it is clear that $\Delta_{[mut]_i}F$ is symmetric in P_1, P_2.

If we want to know the change due to a mutation with no resets, we simply carry out $\lim_{T\to\infty} \Delta_{[mut]_i}F$. This can be seen in Figure 2, where each graph approaches some asymptote as $T \to \infty$, which can be compared to the maximum value of $\Delta_{[mut]_1}F$ attained on the graph. A point of interest is how one could use the 3

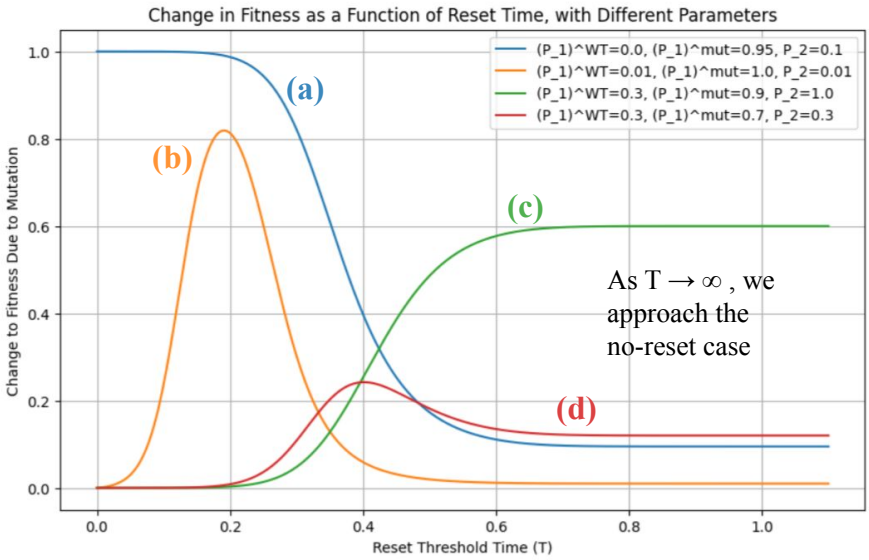

FIGURE 2. This is a graph of $\frac{\Delta F}{\Delta[mut]_1}$ over T, with different values given to P_1^{WT}, P_1^{mut}, and P_2 (which is held constant throughout the mutation to G_1). The values for each curve are as follows: (a) $P_1^{WT} = 0.00$, $P_1^{mut} = 0.95$, $P_2 = 0.10$. (b) $P_1^{WT} = 0.01$, $P_1^{mut} = 1.00$, $P_2 = 0.01$. (c) $P_1^{WT} = 0.30$, $P_1^{mut} = 0.90$, $P_2 = 1.00$. (d) $P_1^{WT} = 0.30$, $P_1^{mut} = 0.70$, $P_2 = 0.30$.

parameters (P_1^{WT}, P_1^{mut}, P_2) to maximize the ratio of the maximum value on the curve to the limit as $T \to \infty$ on the curve, which is done in Figure 3. This ratio physically corresponds to how much resets can amplify a mutation's impact on an organism, and so they demonstrate how important resets may be to rapid evolution.

The results of Figure 3 are in line with an intuitive analysis of the model. The biggest difference between Figure 3a and Figure 3b is the value approached by $\Delta_{[mut]_1}F$ as $T \to \varepsilon_t$, that being 1 and 0 respectively. In the case of Figure 3a, the physical meaning of this "initial" value is that the change in fitness, $\Delta_{[mut]_1}F$, due to the given mutation in a structure where every single incorrect structure is reset, is 1, traversing the entire domain of F, which is $[0, 1]$. Before the mutation, $P_1 = 0$, meaning that the correct pathway is impossible to complete within the structure, regardless of how common structure resets are, and so $F = 0$. After the mutation, $P_1 = 1$, $P_2 > 0$ (no matter how small), and because $T \to \varepsilon_t$, $S \to 1$, so that all incorrect paths will be reset until they inevitably turn into the correct pathway, making $F = 1$. Therefore the mutation to G_1 causes a change in F of 1, or equivalently, $\Delta_{[mut]_1}F = 1$. In Figure 3b, the change in fitness after the mutation, with every incorrect structure being reset, is 0. This is because $P_1, P_2 > 0$, no matter how small, so that the structure will reset until eventually the correct pathway is chosen, meaning $F = 1$ both before and after the mutation when $P_1 = 1 > 0$ still. This means that as $T \to \varepsilon_t$, $\Delta_{[mut]_1}F = 0$. The rest of the curves of both Figure 3a and Figure 3b are nearly identical because the mutations are very similar, and without significant reset the fitness will be very low

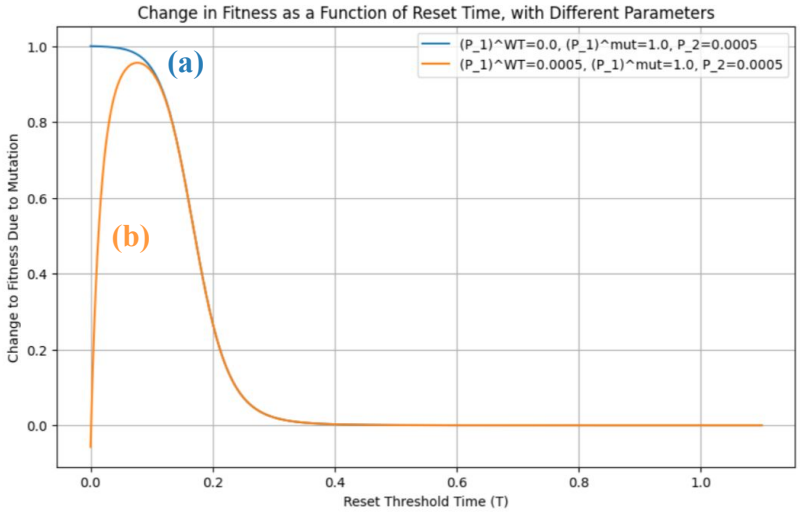

FIGURE 3. This is a follow-up graph to Figure 2, with the same axes and parameters, but with extreme parameter values. These curves, just like those in Figure 2, only describe physical reality on $T \in (\varepsilon_t, \infty)$, because the time threshold for reset cannot be non-positive. The initial values of the two curves are different due to very slightly varied parameters, but the overall curves and $\lim_{T\to\infty}$ are nearly identical. (a) $P_1^{WT} = 0.00$, $P_1^{mut} = 1.00$, $P_2 = \epsilon_p \ll 1$ (arbitrarily close to 0). As $T \to \varepsilon_t$, $\Delta_{[mut]_1} F \to 1$, and as $T \to \infty$, $\Delta_{[mut]_1} F \to \epsilon_p$. (b) $P_1^{WT} = \epsilon_p$, $P_1^{mut} = 1.00$, $P_2 = \epsilon_p$. As $T \to \varepsilon_t$, $\Delta_{[mut]_1} F \to 0$, and as $T \to \infty$, $\Delta_{[mut]_1} F \to (1 - \epsilon_p)\epsilon_p$. As P_2 becomes closer and closer to 0, the maximum of this curve approaches 1 as it simultaneously approaches $T \to \varepsilon_t$ (the curve of (b) approaches the curve of (a)).

both before and after the mutation, because P_2 is very low in both. This extreme example demonstrates the importance of resets to the fitness of a structure, which in the mutated states of both Figure 3a and Figure 3b makes the difference between an F of nearly 1 or of nearly 0.

2.2. **M-Steps.**

Assumptions 2.10. This M-Step model of protein folding is an extension of the last toy model with some notable changes. The 2-gene-protein has been changed to some arbitrary M-gene-protein, while we have also changed the incorrect pathway finish time distribution from a thin Gaussian to a pure delta function. This model's physicality relies on a number of assumptions, all of which will be noted here:

- Our protein folding process is determined by an array of M genes: $(G_1, G_2, \ldots, G_M)$. These M genes are each given some probability of doing their respective step of the process correctly, $(P_1, P_2, \ldots, P_M)$. Generally all $P_i = \epsilon_p \ll 1$ unless noted otherwise. Note that $\epsilon_p \neq \varepsilon_t$, though they are both small numbers, one is used to talk about probabilities, and the

other for finish times. There are n^M possible foldings, only one of which is correct.

- The (time-adjusted!) fitness of the protein folding receives a positive contribution from the probability that the "correct" folding is reached (requires all M genes to make the correct decision), and a negative contribution from how much time those correct foldings take to finish. This is a new concept to be motivated and defined later.
- A mutation to the ith gene changes the probability that the ith step of the folding process is done correctly. Generally this means changing P_i from some small ϵ_p to some larger x.
- In absence of resets: the correct folding finishes at time $M\varepsilon_t$, whereas all incorrect foldings finish at time M. This presumed correlation between "fast" and "correct" foldings is an exaggerated application of the results from [1].
- Some proportion S of the $n^M - 1$ incorrect paths will be reset to the beginning, where all genes will then be able to make new decisions, independently of the decisions they made previously. Because our incorrect finish time distribution is a delta function, S will either be 0 or 1.
- The best mutation for a protein (and its host organism) is that which maximizes the amplification to time-adjusted fitness caused by the mutation, also called the evolvability $\psi^{b\to g}_{[mut]_i}(\mathcal{F})$ (which we will define later).

Now consider a new model with M genes, G_i, each with some probability of picking the correct segment of the path from n choices, P_i, keeping with the logic from the last model. There are now n^M total pathways, with 1 correct and $n^M - 1$ incorrect. Each gene may be in 1 of 2 states, b or g (for bad and good), where $P_i^b = \epsilon_p$ (some value arbitrarily close to 0), and where $P_i^g = x \in (\epsilon_p\,, 1]$ (for us to vary). Similarly, we will now be (arbitrarily) defining a new distribution for both the correct and incorrect pathway completion times, those being delta functions.

Definition 2.11. The distribution of correct times is defined $\mathcal{T}_{Cor}(t) = \delta(t - M\varepsilon_t)$ where $M\varepsilon_t \ll 1$ (arbitrarily close to 0); the distribution of incorrect times, $\mathcal{T}_{Inc}(t) = \delta(t - M)$. Using delta functions is a way to state our deterministic finish times for correct and incorrect structures with "distributions," even though only one value is actually possible within the distribution. These distributions would be followed for a structure with no reset infrastructure present, i.e. $S = 0$.

With this more general model, the fitness F is adjusted accordingly (in line with (2.3):

$$\begin{aligned} F &= \sum_{j=0}^{\infty}\left[\left(\prod_{P\in X_M} P\right)\left(1 - \prod_{P\in X_M} P\right)^j S^j\right] \\ &= \frac{\epsilon_p^{M-\beta}x^\beta}{1 - S + S(\epsilon_p^{M-\beta}x^\beta)} \\ &= \frac{\epsilon_p^{M-\beta}x^\beta}{1 + (\epsilon_p^{M-\beta}x^\beta - 1)\int_T^\infty \delta(t-M)dt} \end{aligned} \tag{2.12}$$

where $X_M = (P_1, P_2, P_3, \ldots, P_M)$ are the probabilities associated with our array of genes, and β is the number of genes (out of M) in the g state. We can consider the

effect of a mutation to some G_i such that it transitions from the b state to the g state:

$$\Delta^{b\to g}_{[mut]_i} F = F(T, x, \epsilon_p, M, \beta + 1) - F(T, x, \epsilon_p, M, \beta). \tag{2.13}$$

There are only two cases for this, depending on T:

$$\Delta^{b\to g}_{[mut]_i} F = \begin{cases} 0 & M\varepsilon_t < T < M \\ (x/\epsilon_p - 1)\epsilon_p^{M-\beta} x^\beta & T \geq M \end{cases} \tag{2.14}$$

and as we can see, $\Delta^{b\to g}_{[mut]_i} F = 0$ when reset infrastructure is present. This follows from the findings of Figure 3b for when $S = 1$. Since every incorrect folding is reset, every folding will end up correct, regardless of a mutation.

We can now lay the groundwork for the concept of time-adjusted fitness, $\mathcal{F}$. This provides a more complete picture (than F) of what is selected for in the organism, by considering the time taken for the folding process. After all, this consideration of time is what would have manifested the reset function in the first place.

Definition 2.15. Let $\widetilde{\xi}$ be the median of the distribution $\xi(t)$, where the distribution describes a probability density function for correct finish times with a reset mechanism present. A unit of time for $\xi(t)$ is equal to that taken to complete a single component of an ultimately incorrect pathway (i.e. the time between gene "decision points"). This median is useful for evaluating how slow or fast protein folding takes, on average, given some X_M and T.

Definition 2.16. The time-adjusted fitness is defined by the following:

$$\mathcal{F}(T, X_M) \equiv kF - \ln\left(\widetilde{\xi}\right) \tag{2.17}$$

where k is a proportionality constant (to be assigned later). A logarithmic relation to $\widetilde{\xi}$ is chosen as opposed to any other because it is the most common choice from literature [1], and generally produces physical results.

If the reset time T is less than the time it would take an incorrect path to complete, M, then the structure will always finish the correct path given sufficient resets. Before the correct folding finishes, it must go through the correct pathway uninterrupted, which will complete many orders of magnitude faster than any incorrect pathways, so its time taken is negligible in the calculation of $\widetilde{\xi}$ if the structure reset even a single time. Let

$$\xi(t) = \underbrace{\epsilon_p^{M-\beta} x^\beta}_{\text{last path, correct}} \left[\underbrace{\delta(t - M\varepsilon_t)}_{\text{before reset}} + \left(\underbrace{\int_T^\infty \delta(t' - M) dt'}_{\text{Does reset occur?}} \right) \sum_{k=1}^{\infty} \underbrace{(1 - \epsilon_p^{M-\beta} x^\beta)^k}_{\text{incorrect, reset}} \underbrace{\delta(t - kT)}_{k^{\text{th}}\text{ reset}} \right]. \tag{2.18}$$

This function for correct-structure-finish-times, when reset occurs at intervals of T, is composed of delta functions that also occur at intervals of T. For $T \geq M$, no reset ever occurs, and we have $\widetilde{\xi} = M\varepsilon_t$, which is not very interesting. To go

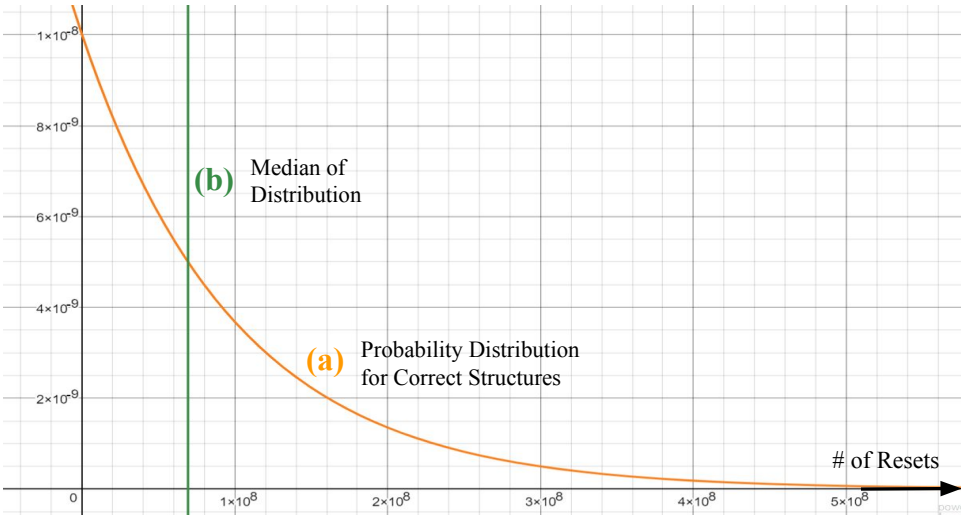

FIGURE 4. Here we assume $T < M$. (a) This is a graph of $\epsilon_p^{M-\beta}x^\beta(1-\epsilon_p^{M-\beta}x^\beta)^t$ (over t) for $t \in \mathbb{R}_0^+$, with parameters $T = 1$, and $\epsilon_p^{M-\beta}x^\beta = 10^{-8}$. This distribution works as an approximation for $\xi(t)$ on large scales, as shown in (2.19) and (2.23). (b) The median of the distribution is in green, $t_o = \widetilde{\xi} \approx 6.93147 \cdot 10^7$.

further we will assume $M\varepsilon_t \leq T < M$, such that

$$\begin{aligned} \int_0^{\alpha T} \xi(t)dt &= \epsilon_p^{M-\beta}x^\beta \sum_{t=0}^{\lfloor\alpha\rfloor}(1-\epsilon_p^{M-\beta}x^\beta)^t \text{ (for } \alpha > 0) \\ &= \epsilon_p^{M-\beta}x^\beta \int_0^{\alpha}(1-\epsilon_p^{M-\beta}x^\beta)^t dt + E(\alpha,\, \epsilon_p^{M-\beta}x^\beta) \end{aligned} \tag{2.19}$$

$$\int_0^{\infty} \xi(t)dt = \epsilon_p^{M-\beta}x^\beta \int_0^{\infty}(1-\epsilon_p^{M-\beta}x^\beta)^t dt + E(\epsilon_p^{M-\beta}x^\beta) = 1. \tag{2.20}$$

Where $E()$ is the error incurred by replacing the summation with an integral. We can write it explicitly as

$$\begin{aligned} E(\alpha,\, \epsilon_p^{M-\beta}x^\beta) &= \sum_{t=0}^{\lfloor\alpha\rfloor} f(t) - \int_0^{\alpha} f(t)dt \\ &= f(0) + \int_0^{\alpha}(t-\lfloor t\rfloor)f'(t)dt - (\alpha - \lfloor\alpha\rfloor)f(\alpha) \end{aligned} \tag{2.21}$$

where we have substituted $f(t) \equiv \epsilon_p^{M-\beta}x^\beta(1-\epsilon_p^{M-\beta}x^\beta)^t$ and $f'(t) \equiv \frac{df}{dt}$ for simplicity. Therefore we can solve

$$\epsilon_p^{M-\beta}x^\beta \int_0^{\widetilde{\xi}/T}(1-\epsilon_p^{M-\beta}x^\beta)^t dt + E(T,\, \epsilon_p^{M-\beta}x^\beta) = \frac{1}{2} \tag{2.22}$$

for $\widetilde{\xi}$:

$$\widetilde{\xi} = \underbrace{\frac{T\ln\left[1+\frac{\ln\left(1-\epsilon_p^{M-\beta}x^\beta\right)}{2\epsilon_p^{M-\beta}x^\beta}\right]}{\ln\left[1-\epsilon_p^{M-\beta}x^\beta\right]}}_{\equiv\widetilde{\xi}^*} - E_T \; ; \; M\varepsilon_t \leq T < M \tag{2.23}$$

where $E_T \leq \frac{3}{2}T$, and E_T does not depend on $\epsilon_p^{M-\beta}x^\beta$ at all. The only variation of E_T comes from the difference $\widetilde{\xi}^* - \lfloor\widetilde{\xi}^*\rfloor$, which occurs such that $\widetilde{\xi}$ is an integer multiple of T (otherwise $\widetilde{\xi}$ would not be consistent with the definition of $\xi(t)$ from (2.18)). However, because $\widetilde{\xi}$ grows very large for small $\epsilon_p^{M-\beta}x^\beta$, while E_T remains approximately constant, we see that $E_T \ll \widetilde{\xi}$ for $\epsilon_p^{M-\beta}x^\beta \ll 1$. Because of this, we ignore E_T, and assume $\epsilon_p^{M-\beta}x^\beta \ll 1$ and therefore $\widetilde{\xi} \approx \widetilde{\xi}^*$ in all further calculations for this model.

Knowing that $F \in [0,1]$ and $\widetilde{\xi} \in (0,\infty)$ ($\widetilde{\xi} \gg 1$ as long as $\epsilon_p^{M-\beta}x^\beta \ll 1$), we must use a proportionality constant in $\mathcal{F}$ to ensure that the relative importance of these two variables is taken into account (and also scaled for their differing co-domains). This constant could be anything, but we will choose the constant $k = 10 + \ln(6.93147 \cdot 10^7)$, such that the $\mathcal{F}$ of the structure in Figure 4 is approximately equal to 10. This is purely a choice of convenience, though k should be neither too large ($k \gg 1$) nor too small ($k < 1$) as to heavily impact the physicality of $\mathcal{F}$. Substituting (2.23) into (2.17), we get

$$\mathcal{F} = \begin{cases} k - \ln\left(\frac{T\ln\left[1+\frac{\ln\left(1-\epsilon_p^{M-\beta}x^\beta\right)}{2\epsilon_p^{M-\beta}x^\beta}\right]}{\ln\left[1-\epsilon_p^{M-\beta}x^\beta\right]}\right) & M\varepsilon_t \leq T < M,\ (\epsilon_p^{M-\beta}x^\beta) \ll 1 \\ k(\epsilon_p^{M-\beta}x^\beta) - \ln(M\varepsilon_t) & T \geq M. \end{cases} \tag{2.24}$$

Similarly to $\Delta^{b\to g}_{[mut]_i}F$, we may now use use $\mathcal{F}$ to analyze mutation impact.

Definition 2.25. We can define the amplification of $\mathcal{F}$ due to a mutation of one G_i from the b state to the g state, $\psi^{b\to g}_{[mut]_i}(\mathcal{F})$, called the evolvability of the structure:

$$\begin{aligned} \psi^{b\to g}_{[mut]_i}(\mathcal{F}) &\equiv \frac{\mathcal{F}(T,x,\epsilon_p,M,\beta+1)}{\mathcal{F}(T,x,\epsilon_p,M,\beta)} \\ &= \begin{cases} 1 + \frac{\ln(x/\epsilon_p)}{k+\ln(\frac{\epsilon_p^{M-\beta}x^\beta}{T\ln(2)})} & M\varepsilon_t \leq T < M,\ \epsilon_p^{M-\beta}x^\beta \ll 1 \\ 1 & T > M,\ \epsilon_p^{M-\beta}x^\beta \ll 1. \end{cases} \end{aligned} \tag{2.26}$$

We use a ratio for evolvability rather than a difference because when numbers are plugged in, $\mathcal{F}$ values often blow up to become very large. This means that any mutation may artificially create a large change, when the physical reality does not match. In (2.26), natural logarithms have been Taylor expanded when appropriate, though the conditions for expansion are no different than the conditions for which $\widetilde{\xi} \approx \widetilde{\xi}^*$, as has already been taken for granted.

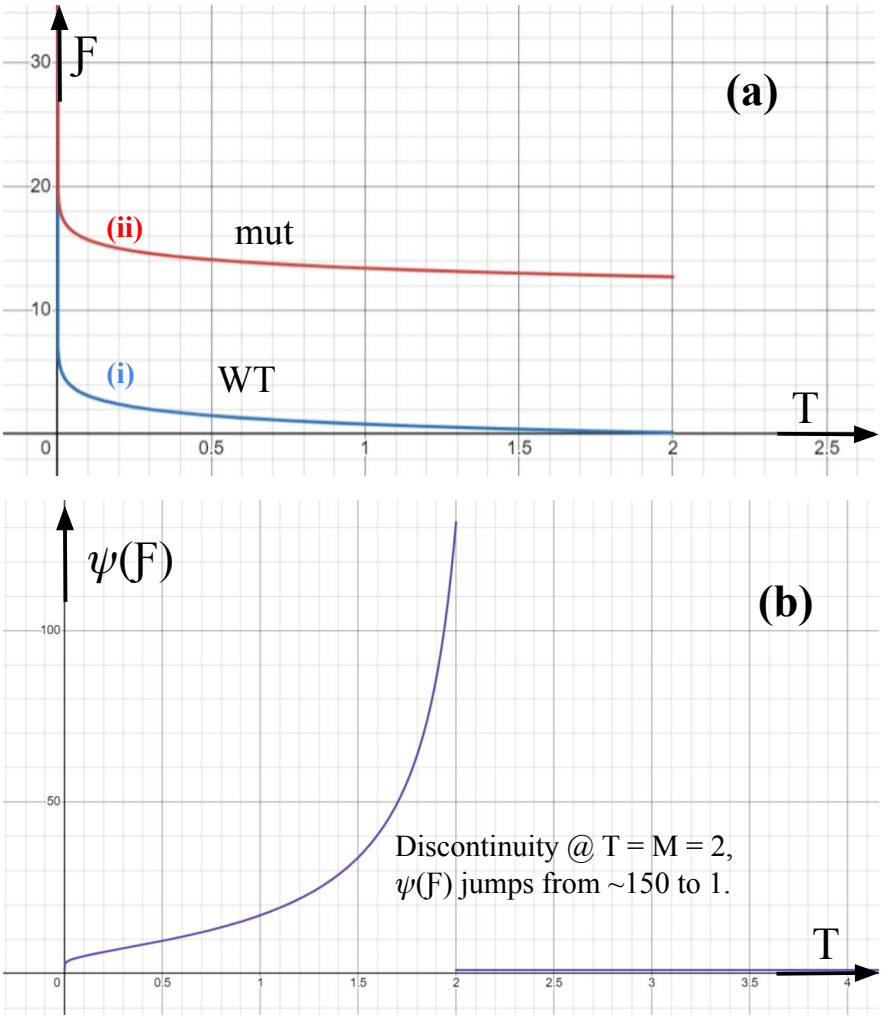

FIGURE 5. (a) and (b) involve the same mutation, with the same parameters: the genome changes $\beta : 0 \to 1$. (a) This is a graph of a pre-mutation, wild-type $\mathcal{F}(T, 0.3, 0.000001, 2, 0)$ (i) and the mutated $\mathcal{F}(T, 0.3, 0.000001, 2, 1)$ (ii). (b) This is a graph of the amplification from WT $\to$ mut in (a), called the evolvability $\psi^{b\to g}_{[mut]_i}(\mathcal{F})$. $M = 2$ here, and after $T > M$, $\psi^{b\to g}_{[mut]_i}(\mathcal{F}) \approx 1$.

Remark 2.27. Under this model, $\psi^{b\to g}_{[mut]_i}(\mathcal{F})$ is independent of i, i.e. the impact of the mutation on evolvability does not depend on which gene is mutated, just that a gene is mutated.

In Figure 5, a particular parameter regime of $\mathcal{F}$ and $\psi^{b\to g}_{[mut]_i}(\mathcal{F})$ are graphed over T, and the latter is then graphed over an even larger domain of values. The physical meaning of the amplification value changes depending on the value of the $\mathcal{F}$ that was mutated. Figure 5 has very straightforward graphs, in regards

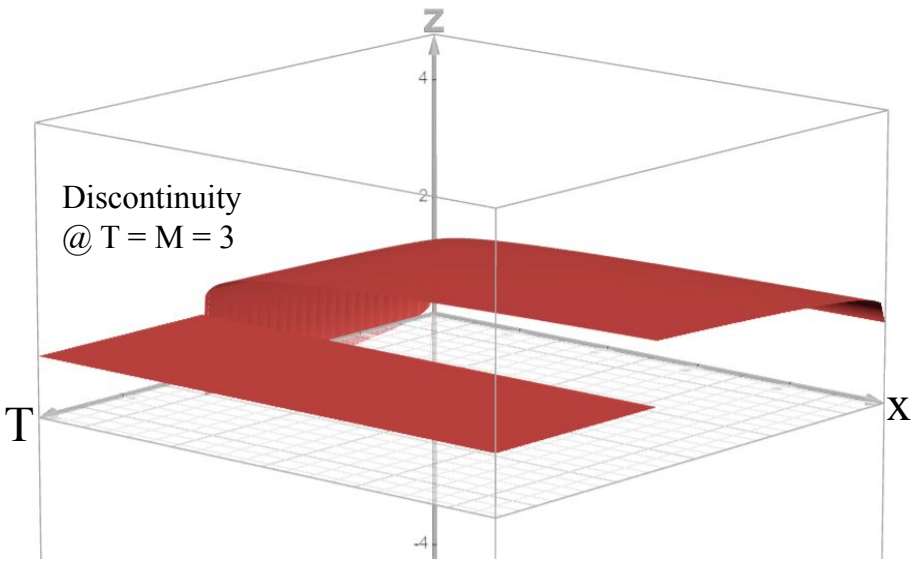

FIGURE 6. This figure consists of a graph of the amplification due to mutation, $\psi^{b\to g}_{[mut]_i}(\mathcal{F})$, with the mutation taking $\beta : 0 \to 1$, where $\epsilon_p = 0.001$, and M=3, giving $\psi^{b\to g}_{[mut]_i}[\mathcal{F}(T, x, 0.001, 3, 0)]$. In Figure 2, Figure 3, and Figure 5, T has been varied on one axis, as it is here, but in this figure we have added the probability associated with genes in the g state, that being $x \in (0, 1]$, which is on a second axis. This means the x-axis is constrained to only $(\epsilon_p, 1]$, and so those are the only x-values depicted on the graph. $\mathcal{F} > 0$ on the entire rectangle input pictured.

to both $\mathcal{F}$ and $\psi^{b\to g}_{[mut]_i}(\mathcal{F})$, which remain positive throughout $T < M$. Here the largest amplification occurs as $T \to M^-$. This is a consequence of our considering of median time: resets become more costly to time-adjusted fitness the closer the reset time T gets to M, meaning that there is now an even bigger benefit to needing fewer resets in order to achieve the median correct completion than before. This decrease in resets needed is achieved by the mutation that increases the probability of picking the correct pathway after each reset. This is not a trivial discovery, and one can see it in the analytical expression in (2.26), by tracking the dependence on T in the $M\varepsilon_t \leq T < M$ case.

In Figure 5b, the curve of $\psi^{b\to g}_{[mut]_i}(\mathcal{F})$ is approximately equal to 1 when $T > M$, which is actually independent of parameter regime. No matter the values plugged into $\psi^{b\to g}_{[mut]_i}(\mathcal{F})$, it would return 1 for $T > M$. This is because the median completion times of the no reset case, $\widetilde{\xi}$, are so low that its natural log dominates over F, regardless of the specific parameters or mutations. As long as the wild-type $F|_{T>M} = (\epsilon_p^{M-\beta} x^{\beta}) \ll 1$, and the mutated $F|_{T>M} = (\epsilon_p^{M-\beta-1} x^{\beta+1}) \ll 1$, the mutation is not significant enough to cause any large change in $\mathcal{F}$, meaning $\psi^{b\to g}_{[mut]_i}(\mathcal{F}) \approx 1$. This shows the importance of resets to the impact of mutations on

an organism's fitness. Under our model, a lack of reset means that any individual mutation is too insignificant for the environment to select for it.

In Figure 6, both T and x are varied on the same graph, and their impact on $\psi^{b\to g}_{[mut]_i}[\mathcal{F}(T, x, 0.001, 3, 0)]$ is observed. Varying x has a somewhat consistent impact on $\psi^{b\to g}_{[mut]_i}(\mathcal{F})$ values when $M\varepsilon_t \leq T < M$, i.e. when reset occurs in the structure. In this regime, increasing x increases $\psi^{b\to g}_{[mut]_i}(\mathcal{F})$. This makes physical sense because increasing x while holding ϵ_p constant (as we are) means that the mutation of a gene that changes its probability $\epsilon_p \to x$ is more significant of a change, and so will cause a larger amplification, $\psi^{b\to g}_{[mut]_i}(\mathcal{F})$. Here as well, the pattern of [larger T] $\to$ [larger $\psi^{b\to g}_{[mut]_i}(\mathcal{F})$] holds for $M\varepsilon_t \leq T < M$.

2.3. **Path-Dependent Time.**

Assumptions 2.28. This M-Step model of protein folding with path dependence, is an extension of the the standard M-Step model with some notable changes. We have completely reconstructed finish times, making them dependent on the path taken along the tree in Figure 7, resulting in an incorrect finish time distribution composed of delta functions at every integer value from 1 to M. This model's physicality relies on a number of assumptions, all of which will be noted here:

- Our protein folding process is still determined by an array of M genes: (G_1, G_2, ... , G_M). These M genes are each given some probability of doing their respective step of the process correctly, $(P_1, P_2, \ldots, P_M)$. Generally all $P_i = \epsilon_p \ll 1$ unless noted otherwise. This gives us n^M possible foldings, only one of which is correct. This is identical to that of the last model.
- The time-adjusted fitness of the protein folding receives a positive contribution from the probability that the "correct" folding is reached (requires all M genes to make the correct decision), and a negative contribution from how much time those correct foldings take to finish. This is identical to that of the last model.
- A mutation to the ith gene changes the probability that the ith step of the folding process is done correctly. Generally this means changing P_i from some small ϵ_p to some larger x. This is identical to that of the last model, though the placement of i now has importance (see next bullet-point) where it did not in any previous model.
- In absence of resets: the correct folding finishes at time $M\varepsilon_t$, whereas the incorrect foldings finish at times proportional to the number of incorrect steps done in the folding process. A distribution for incorrect finish times would include delta functions at every integer value from 1 to M, as elaborated on in Figure 7. This time-dependence is the only change made from the last model, though it has wide-ranging impacts. The presumed correlation between "fast" and "correct" foldings is an exaggerated application of the results from [1].
- Some proportion S of the $n^M - 1$ incorrect paths will be reset to the beginning, where all genes will then be able to make new decisions, independent of the decisions they made previous. This S depends on our incorrect finish time distribution, and so $S(T)$ will resemble a step function with M steps (whereas the previous model had only 1).

- The best mutation for a protein (and its host organism) is that which maximizes the amplification to time-adjusted fitness caused by the mutation, also called the evolvability $\psi^{b\to g}_{[mut]_i}(\mathcal{F})$. This is identical to that of the last model.

Up until now, we have only considered $\mathcal{T}_{Inc}$ distributions with a single peak, i.e. all incorrect paths were assigned a finish time from the same distribution (either a delta function or Gaussian) no matter the path taken through the tree.

Now, if the model was changed such that it accounted for the path taken through the genes, rather than just the destination, it would add an additional level of complexity to the finish times of incorrect paths. This model is initiated in Figure 7. Now, the order along the tree (i.e. when each gene makes its choice, either correct or incorrect) is important when it was not before. If all but one gene makes the correct choice, the structure could finish at $t = 1$, $t = M$ or at integer values between, depending on where in the order the incorrect choice was picked by a gene. Without reset, the distribution of correct times, $\mathcal{T}_{Cor}(t)$ will look identical to that of the last model, but the new $\mathcal{T}_{Inc}(t)$ will contain delta functions (of varying "area") at intervals of 1.

$$\mathcal{T}_{Cor}(t) = \delta(t - M\varepsilon_t), \tag{2.29}$$

$$\mathcal{T}_{Inc}(t) = \underbrace{\frac{1}{1-\prod_{P\in X_M}(P)}}_{\text{normalizes } \int_{M\varepsilon_t}^{\infty}\mathcal{T}_{Inc}\,dt=1} \sum_{j=1}^{M}\left[\underbrace{(1-P_j)}_{G_j \text{ incorrect}}\left(\underbrace{\prod_{i=0}^{j-1}P_i}_{\text{all } G_i \text{ correct}}\right)\underbrace{\delta(t-M-1+j)}_{M+1-j \text{ incorrect steps}}\right] \tag{2.30}$$

with $P_0 \equiv 1$. The $\frac{1}{1-\prod_{P\in X_M}(P)}$ factor out front of (2.30) is to scale it such that $\int_0^\infty \mathcal{T}_{Cor}(t)dt = \int_0^\infty \mathcal{T}_{Inc}(t)dt = 1$, communicating that both represent all possible correct and incorrect finish times, respectively. If we now account for the possibility of reset infrastructure with some threshold time T, then our distribution of correct finish times is given by

$$\xi(t) = \left[\underbrace{\delta(t-M\varepsilon_t)}_{\text{before reset}} + \underbrace{\sum_{h=1}^{\infty}\left(1-\prod_{P\in X_M}P\right)^h\left(\int_T^\infty \mathcal{T}_{Inc}(t')dt'\right)^h\underbrace{\delta(t-hT)}_{h^{\text{th}}\text{ reset}}}_{\text{incorrect, gets reset } h \text{ times}}\right]\underbrace{\prod_{P\in X_M}P.}_{\text{last path correct}} \tag{2.31}$$

The "path dependence" of (2.31) is hidden in $\int_T^\infty \mathcal{T}_{Inc}(t')dt'$, where $\mathcal{T}_{Inc}$'s dependence on the order of probabilities, $P_i \in X_M$, is related in (2.30). We can again approximate this distribution with one easier to work with:

$$\int_0^{\alpha T}\xi(t)dt = \left(\prod_{P\in X_M}P\right)\int_0^\alpha\left(1-\prod_{P\in X_M}P\right)^t\left(\int_T^\infty \mathcal{T}_{Inc}(\eta)d\eta\right)^t dt + E(X_M). \tag{2.32}$$

Where $E()$ is once again the error of the integral on the right side of (2.32) (its value is analogous to that in (2.21)), and will be used to find the error in our approximation of $\widetilde{\xi}$. The integral $\int_0^\infty \xi(t)dt = F$, as does the right hand side of

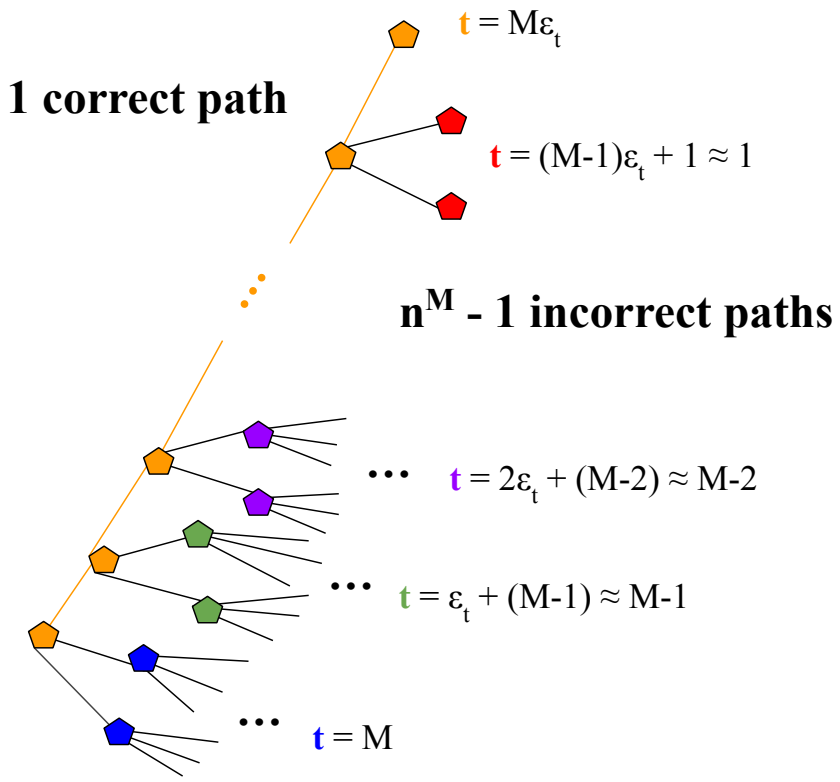

FIGURE 7. This figure is a schematic of a model with M genes (for a total of n^M possible structures, with the schematic depicting $n = 3$), but varying finish times for incorrect structures. The finish times are now dependent on the path taken to finish, rather than just the binary correct/incorrect identity associated with the finish. If the incorrect structure is on the correct path for any number of legs out of the total M, then the finish time will be faster because all legs of the correct path take time $\varepsilon_t \ll 1$, whereas all legs off the correct path take time 1.

(2.32) for $\alpha \to \infty$. Knowing that, we can now write:

(2.33)
$$\left(\prod_{P\in X_M} P\right)\int_0^{\widetilde{\xi}/T}\left(1-\prod_{P\in X_M} P\right)^t\left(\int_T^\infty \mathcal{T}_{Inc}(\eta)\,d\eta\right)^t dt + E(T\,,X_M) = \frac{1}{2}F.$$

From here we can solve for this model's $\widetilde{\xi}$, as was done in (2.23):

$$\widetilde{\xi} = \underbrace{\frac{T\ln\left[1+F\frac{\ln\left(\left[1-\prod_{P\in X_M}P\right]\left[\int_T^\infty \mathcal{T}_{Inc}(\eta)\,d\eta\right]\right)}{2\left(\prod_{P\in X_M}P\right)}\right]}{\ln\left[\left(1-\prod_{P\in X_M}P\right)\left(\int_T^\infty \mathcal{T}_{Inc}(\eta)\,d\eta\right)\right]}}_{=\widetilde{\xi}^*} - E_T \;;\; M\varepsilon_t \leq T < M, \tag{2.34}$$

where again, $E_T \leq \frac{3}{2}T$, and its exact value depends on $\widetilde{\xi}^* - \lfloor\widetilde{\xi}^*\rfloor$. E_T becomes insignificant compared to $\widetilde{\xi}$ as $\left[\left(1-\prod_{P\in X_M}P\right)\left(\int_T^\infty \mathcal{T}_{Inc}(\eta)\,d\eta\right)\right] \to 1$; not because E_T shrinks, but because $\widetilde{\xi}$ grows to be $\gg E_T$. Further, $\left(\prod_{P\in X_M}P\right) = (\epsilon_p^{M-\beta}x^\beta)$, though $\mathcal{T}_{Inc}(t)$ is dependent on the order of items within the list of probabilities, X_M, so the more explicit notation of $\left(\prod_{P\in X_M}P\right)$ has been opted for everywhere.

The standard fitness, F, can be found using the same logic as in (2.3) and (2.12):

$$F = \frac{\prod_{P\in X_M}P}{1-\left(1-\prod_{P\in X_M}P\right)\int_T^\infty \mathcal{T}_{Inc}(t)dt}. \tag{2.35}$$

But what about $\mathcal{F}$ and $\psi^{b\to g}_{[mut]_i}(\mathcal{F})$? How are those different under this new model? We compute:

(2.36)

$$\begin{aligned}\mathcal{F} &= \frac{k\prod_{P\in X_M}P}{1-\left(1-\prod_{P\in X_M}P\right)\int_T^\infty \mathcal{T}_{Inc}(t)dt} \\ &\qquad - \ln\left(\frac{T\ln\left[1+F\frac{\ln\left(\left[1-\prod_{P\in X_M}P\right]\left[\int_T^\infty \mathcal{T}_{Inc}(\eta)d\eta\right]\right)}{2\left(\prod_{P\in X_M}P\right)}\right]}{\ln\left[\left(1-\prod_{P\in X_M}P\right)\left(\int_T^\infty \mathcal{T}_{Inc}(\eta)d\eta\right)\right]}\right) \\ &= \frac{k\prod_{P\in X_M}P}{1-\left(1-\prod_{P\in X_M}P\right)\int_T^\infty \mathcal{T}_{Inc}(t)dt} \\ &\qquad - \ln\left(\frac{T\ln\left[1+\frac{\ln\left(\left[1-\prod_{P\in X_M}P\right]\left[\int_T^\infty \mathcal{T}_{Inc}(\eta)d\eta\right]\right)}{2-2\left(1-\prod_{P\in X_M}P\right)\int_T^\infty \mathcal{T}_{Inc}(t)dt}\right]}{\ln\left[\left(1-\prod_{P\in X_M}P\right)\left(\int_T^\infty \mathcal{T}_{Inc}(\eta)d\eta\right)\right]}\right).\end{aligned}$$

If $X_M = (\epsilon_p, \ldots, \epsilon_p)$, then $X_M^i = (\epsilon_p, \ldots, x, \ldots, \epsilon_p)$ is the list of probabilities with a mutation on G_i. Using this we can rewrite (2.26) for our current model:

$$\psi^{b\to g}_{[mut]_i}(\mathcal{F}) = \frac{\mathcal{F}(T, X_M^i)}{\mathcal{F}(T, X_M)}. \tag{2.37}$$

Remark 2.38. In Remark 2.27 we found that the standard M-Step model resulted in a $\psi^{b\to g}_{[mut]_i}(\mathcal{F})$ that was independent of i. Under this model, we find that is no longer the case: $\psi^{b\to g}_{[mut]_i}(\mathcal{F})$ can be dependent on i, i.e. the impact of the mutation on evolvability may depend on which gene is mutated.

Because of the complexity of computing $\psi^{b\to g}_{[mut]_i}(\mathcal{F})$, it is more interesting to analyze the relationships between all the variables than to actually find an analytical expression of the evolvability $\psi^{b\to g}_{[mut]_i}(\mathcal{F})$ as a function of T.

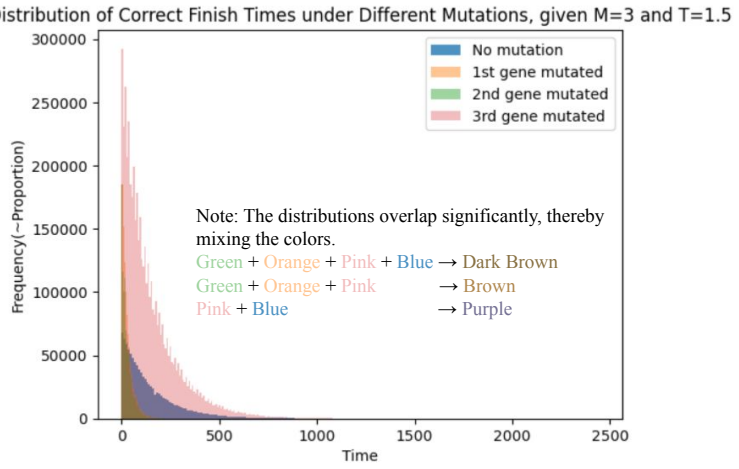

FIGURE 8. This figure demonstrates how, under some reset cases, certain mutations are more beneficial to the evolvability of a structure than others. The histograms show the differences in distributions between the unmutated $X_3 = (\epsilon_p, \epsilon_p, \epsilon_p)$, and the mutated X_3^1, X_3^2, and X_3^3. In each mutation, $\epsilon_p = 0.1$ and $x = 0.5$ for $x/\epsilon_p = 5$. The histograms model the finish times of the correct structures for their respective X_M^i and T values. If a histogram is shifted left such that it is thin and tall, it will have a lower median time than a histogram of comparative total area that is more widely distributed; if it is instead distributed the same but has a larger area, it will have a higher proportion of correct finishes (F). We show the specific case of $T = 1.5$, as it depicts these differences of mutation impact most cleanly. In this reset case there are two different impacts a reset can have. For both X_3^1 and X_3^2, the mutation just changes the median time of the distribution by a factor of $1/5$, but not the actual amount of correct finishes (the total area of the histogram). This means that these two distributions overlap almost entirely (and thereby mixing their histogram colors). For X_3^3, the mutation does not change the timing of the distribution, but does increase the proportion of correct finishes by a factor of 5.

Figure 8 reveals the behavior unique to this model in particular: different responses from the model depending on which gene is mutated, and what the reset condition is. In cases where some incorrect structures are reset and some are not, the ratio of the mutated probability to the unmutated probability, x/ϵ_p, can show up differently depending on which gene is mutated. The ratio represents either that with which the proportion of correct finishes increases, or the inverse of that with which the median time of the distribution decreases. If the mutation in question acts on a gene which, if it was the first to make the wrong "decision," would result in a reset structure that would start again at the base of the decision tree, the

mutation would help to minimize resets by decreasing this likelihood, and therefore reducing the median time to completion for correct structures. If, however, the mutation acts on a gene which, if it was the first to make the wrong decision, would result in an un-reset, incorrect structure, then the mutation would not be minimizing resets, but rather minimizing incorrect structures. The mutation would do this by increasing the chance that if a structure made it to that decision point, wholly correct in its past decisions, it would continue being correct, when the alternative is finishing incorrectly. This results in an increase in the probability of correct structures proportional to the increase in probability from the mutation.

To recap: depending on whether an incorrect decision by a gene could result in a reset, or a finished, incorrect structure, a mutation to that gene that changes its probability from $\epsilon_p \to x$ either changes the median time by a factor of ϵ_p/x or changes the proportion/chance of correct structures by a factor of x/ϵ_p, respectively. Now the question becomes: under our current model, which mutation results in a bigger evolvability? If we return to the definition of evolvability and that of time-adjusted fitness from (2.26) and (2.17) respectively, one can see evolvability has a linear relationship with the post-mutation F, but a logarithmic relationship with the post-mutation $\widetilde{\xi}$. Therefore a change in F will almost always have a larger impact than a proportional change in $\widetilde{\xi}$. The evolvability of X_3^3 is greater than that of both X_3^1 and X_3^2 in Figure 8. This stratification of different mutations only occurs under this model, where finish times for incorrect structures can vary depending on the "path" taken through the decision tree, and not the previous model, where the times did not vary.

3. Another Approach: Fitness Distribution

Assumptions 3.1. This M-Step model of protein folding is an entirely different approach to the question compared to previous models. We have now scrapped the "correct" path concept. Every pathway exists on a range of fitness (density) values from 0 to 1. Paths are no longer strictly correct or incorrect, but somewhere in the middle. Our goal is to find out if reset affects the impact of mutations the same way as in the other models, even though we removed the mode by which the impact has occurred. This model's physicality relies on a number of assumptions, all of which will be noted here:

- Our protein folding process is determined by an array of M genes: $(G_1, G_2, \ldots, G_M)$. These M genes are each given some probability of doing their respective step of the process in a way that keeps the folding on the fastest path, $(P_1, P_2, \ldots, P_M)$. Generally all $P_i = \epsilon_p = 1/n \ll 1$ unless noted otherwise (ϵ_p was just an arbitrarily small number in previous models, but here we have given it physical meaning).
- The fitness of the folding has returned to the original definition, no longer depending on time. It is again defined by the expected fitness value, but now considering all of our possible paths. Each path is assigned some arbitrary value $f_i \in [0, 1]$, and each is weighted by the probability that the path is taken. We add up all of these contributions, and receive our fitness. We choose not to use time-dependent fitness for simplicity, as to not muddle our results.
- A mutation to the ith gene changes the probability that the ith step continues along the fastest path, which leads to the fitness value, f_{n^M}. Only

looking at mutations that improve the probability of the fastest path (as opposed to any other path) is an arbitrary choice, but one that proves useful.

- In absence of resets: the path corresponding to the arbitrary fitness value f_{n^M} is the fastest, with all other $n^M - 1$ paths taking some longer time to finish, according to some unknown but arbitrary distribution. This distribution could be a near-Gaussian like in our 2-Step model, or a delta function, or anything else: we choose not to parameterize S in terms of T for simplicity here (in contrast to our previous models).
- In this model we are still paying attention to the fastest path, like in previous models, yet we do not assume it has the highest fitness value (nor do we label it "correct"). This is an attempt to generalize the results of previous models.
- Some proportion S of the $n^M - 1$ other paths will be reset to the beginning, where all genes will then be able to make new decisions, independently of the decisions they made previously. This is identical to that of previous models.
- The best mutation for a protein (and its host organism) is that which maximizes the change in fitness caused by the mutation, or $\Delta_{[mut]_i} F$. This is identical to our 2-Step model.

One aspect of these past models holds back the physicality of our results: the "correct" and "incorrect" structure developments. We can get rid of this, and instead randomize how advantageous every path is, according to some distribution. However, to best analyze the consequences of this, we will go back to only considering F and $\Delta^{b\to g}_{[mut]_i} F$ rather than $\mathcal{F}$ and $\psi^{b\to g}_{[mut]_i}(\mathcal{F})$. But first we must be more precise about the model: there are still n^M different possible structures, but now each one will randomly receive a fitness score from an unknown probability distribution on $[0, 1]$. We will still keep track of only 1 of these n^M endpoints, but we will not know the fitness score assigned to it, which will be denoted f_{n^M}. The fitness scores of all other endpoints are denoted f_1, f_2, ..., f_{n^M-1}. The average of this array of fitness scores will be denoted f_{avg}.

If every leg of every path had an equal $\epsilon_p \equiv 1/n$ probability of being chosen at every decision point, then every one of the n^M endpoints would have the same $\epsilon_p^M = 1/n^M$ probability of organically being chosen. We can then tally the total fitness of our structure (before any mutations occur):

$$\begin{aligned} F &= \epsilon_p^M f_1 + \epsilon_p^M f_2 + \cdots + \epsilon_p^M f_{n^M-1} + \epsilon_p^M f_{n^M} \\ &= (1 - \epsilon_p^M) f_{avg} + \epsilon_p^M f_{n^M}. \end{aligned} \tag{3.2}$$

If we then mutated one of the genes G_i along the path that we are tracking, such that its probability of continuing (if possible) along the f_{n^M} path, P_i, changes from $\epsilon_p \to x$, then we would also have to account for the decreased probabilities for the other $n - 1$ options given to that G_i:[2]

$$\Delta^{b\to g}_{[mut]_i} F = (x - \epsilon_p)\epsilon_p^{M-1}(f_{n^M} - f_{avg}). \tag{3.3}$$

What happens now if we were to re-introduce some reset infrastructure, not assuming anything about the distribution of finish times except that our specific

[2]See Appendix for more information.

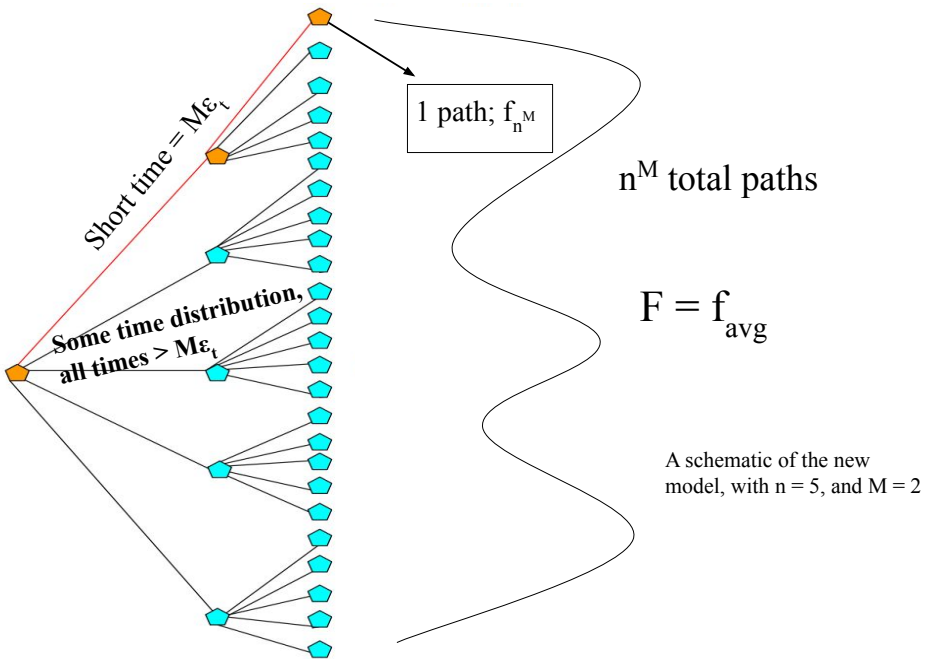

FIGURE 9. This is a schematic of the model with randomized fitness values assigned to every endpoint. The fitness of the structure is given by the average fitness of all of the endpoints, because the outcome is probabilistic and not deterministic. One of these paths is the fastest to complete comparative to the other $n^M - 1$ paths, and the fitness assigned to its completion is labeled f_{n^M}. Without mutations, we say every decision, and therefore every endpoint, is equally probable.

pathway, the one with f_{n^M}, finishes the fastest? If we use some $S \in [0,1]$ to represent the proportion of structures (not including the fastest one) that are reset, then the wild-type fitness would look like this:

$$F = \left(\underbrace{\frac{\epsilon_p^M}{1 - S + S\epsilon_p^M}}_{\text{proportion with } f_{n^M}} \right) f_{n^M} + f_{avg} \left(\underbrace{1 - \frac{\epsilon_p^M}{1 - S + S\epsilon_p^M}}_{\text{all other paths}} \right) + E_S \tag{3.4}$$

where the error E_S only becomes nonzero as S grows arbitrarily close to 1 (though $E_S = 0$ at $S = 1$). This is because, as $S \to 1$, the group of non-resetting paths may be too small to approximate their average with f_{avg}. In other words: our sample size becomes too low to assume the sample mean is the same as the population

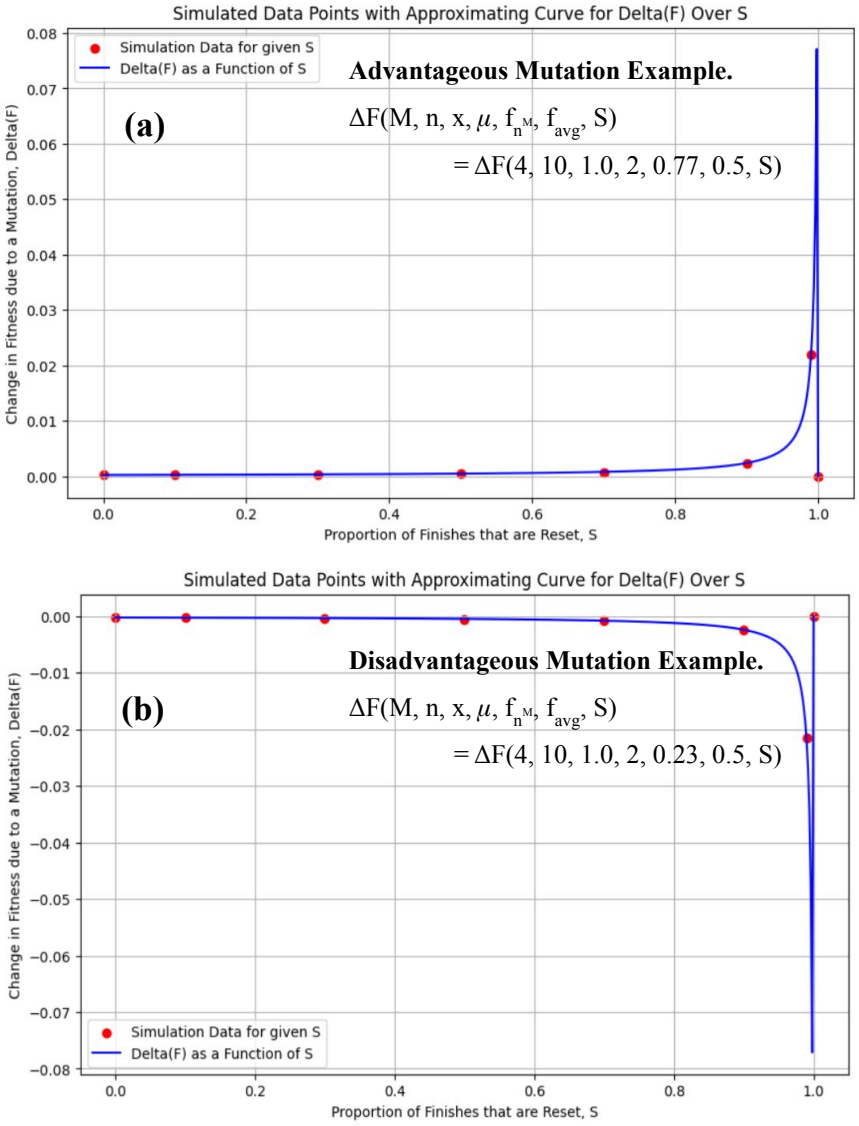

FIGURE 10. This figure contains 2 plots with the analytic function from (3.5) (ignoring the error term) matched up against 7 data points from simulations of the model, at $S = \{0.0, 0.1, 0.3, 0.5, 0.7, 0.9, 0.99, 1.0\}$. (a) $f_{n^M} = 0.77$. The mutation is beneficial because $f_{n^M} > f_{avg}$, and therefore $\Delta^{b\to g}_{[mut]_i} F > 0$ everywhere. This value increases as S increases. (b) $f_{n^M} = 0.23$. This mutation is not beneficial so $\Delta^{b\to g}_{[mut]_i} F < 0$ everywhere, though it further decreases as S increases. This means that a larger reset presence increases the impact of our mutation, regardless of its advantageousness or lack thereof.

mean, f_{avg}. We can now find $\Delta^{b\to g}_{[mut]_i}F$ when considering resets:

$$\Delta^{b\to g}_{[mut]_i}F = \left(\frac{x\epsilon_p^{M-1}}{1-S+Sx\epsilon_p^{M-1}} - \frac{\epsilon_p^M}{1-S+S\epsilon_p^M}\right)(f_{n^M} - f_{avg}) + E_S \tag{3.5}$$

$$= \frac{(x-\epsilon_p)\epsilon_p^{M-1}(f_{n^M} - f_{avg})}{1-S} + E_S + E_{x\epsilon_p^{M-1}} \tag{3.6}$$

where the E_S term has the same conditions as that in (3.5), and the $E_{x\epsilon_p^{M-1}}$ term goes to zero for very small $x\epsilon_p^{M-1}$. Collectively, to make a condition for both errors simultaneously, we can state that for $1-S \gg x\epsilon_p^{M-1}$; $E_S + E_{x\epsilon_p^{M-1}} \to 0$. Then (3.6) can be generalized to any list of probabilities X_M, even those that have $\beta > 0$, as long as our error condition is met:

$$\Delta^{b\to g}_{[mut]_i}F \approx \frac{(x-\epsilon_p)\epsilon_p^{M-\beta-1}x^\beta(f_{n^M} - f_{avg})}{1-S}; (1-S) \gg \epsilon_p^{M-\beta}x^\beta. \tag{3.7}$$

In Figure 10, one can see how the presence of reset makes a mutation more impactful to the evolvability of an organism and its genome, if the mutation increases the likelihood of a faster pathway being taken. The reset works to amplify the impact of the mutation, regardless of whether that impact is positive or negative. This is more general than our previous findings, because we don't assume our fast path to be "better" than the others, and we still achieve the same result regarding resetting's relationship with mutations. We now see how a bad mutation can critically lower an organism's evolvability to the extent that their genome may be selected against, which is the fundamental thesis of evolutionary biology.

Appendix

This is an elaboration on a statement made regarding calculating the fitness in the fitness distribution model, whose description is laid out in Assumptions 3.1.

In the case of a mutation to some G_i that changes its corresponding P_i from $\epsilon_p \to x$, that increased probability must come at the expense of the probabilities of the other paths available to G_i. Because of this, each endpoint descending from the mutated point on the tree now has an altered probability of being landed on.

We can write the mutated fitness, F_{mut}:

(4.1)

$$F_{mut} = \underbrace{\left(\epsilon_p - \frac{x-\epsilon_p}{n-1}\right)\epsilon_p^{M-1} f_1 + \cdots + \left(\epsilon_p - \frac{x-\epsilon_p}{n-1}\right)\epsilon_p^{M-1} f_{n^{M+1-i}-1}}_{\text{mutation decreases probability of all other endpoints downstream}}$$

$$+ \underbrace{\epsilon_p^M f_{n^{M+1-i}} + \cdots + \epsilon_p^M f_{n^M-1}}_{\text{paths unaffected by mutation (not downstream)}} + \underbrace{x\epsilon_p^{M-1} f_{n^M}}_{\text{our path}}$$

$$= \left(\underbrace{\frac{n^{M+1-i}-1}{n^M}}_{=(n^{M+1-i}-1)\epsilon_p^M} - \underbrace{(x-\epsilon_p)\epsilon_p^{M-1}}_{\text{change in probability}} \right) \times \underbrace{f_{\text{avg}}}_{\text{assuming } \frac{f_1+\cdots+f_{n^{M+1-i}-1}}{n^{M+1-i}-1}=f_{\text{avg}}}$$

$$+ \left(\underbrace{\frac{n^M - n^{M+1-i}}{n^M}}_{=1-\epsilon_p^M n^{M+1-i}} \right) \times \underbrace{f_{\text{avg}}}_{\text{assuming } \frac{f_{n^{M+1-i}}+\cdots+f_{n^M}}{n^M-n^{M+1-i}}=f_{\text{avg}}} + x\epsilon_p^{M-1} \times f_{n^M}$$

$$= \left(1 - x\epsilon_p^{M-1}\right) f_{\text{avg}} + x\epsilon_p^{M-1} f_{n^M}.$$

This result would be expected if one just used the logic of (3.2) without careful consideration of how a mutation may affect tree paths (and their endpoints) differently depending on their placement.

The result (3.3) follows immediately from subtracting (3.2) from this (4.1), as $\Delta^{b\to g}_{[mut]_i} F$ was defined in (2.9).

Acknowledgments

It is a pleasure to thank my mentor, Phillip Lo, for his effective guidance in my irregular project. I would further like to thank Dr. Arvind Murugan and Dr. Riccardo Ravasio for giving me interesting ideas to use in new models of dynamic instability.

References

[1] Riccardo Ravasio. Kabir Husain. A minimal scenario for the origin of non-equilibrium order. https://arxiv.org/abs/2405.10911. 2024.

[2] Arvind Murugan. Speed, dissipation, and error in kinetic proofreading. Proceedings of the National Academy of Sciences. 2012.