

## RESEARCH ARTICLE

# A structured model and likelihood approach to estimate yeast prion propagon replication rates and their asymmetric transmission

Fabian Santiago<sup>1</sup>, Suzanne Sindi<sup>2\*</sup>

**1** Department of Mathematics, University of Arizona, Tucson, Arizona, United States of America,

**2** Department of Applied Mathematics, University of California Merced, Merced, California, United States of America

\* [ssindi@ucmerced.edu](mailto:ssindi@ucmerced.edu)



## OPEN ACCESS

**Citation:** Santiago F, Sindi S (2022) A structured model and likelihood approach to estimate yeast prion propagon replication rates and their asymmetric transmission. *PLoS Comput Biol* 18(7): e1010107. <https://doi.org/10.1371/journal.pcbi.1010107>

**Editor:** Attila Csikász-Nagy, Pázmány Péter Catholic University; Pazmany Peter Katolikus Egyetem, HUNGARY

**Received:** October 12, 2021

**Accepted:** April 14, 2022

**Published:** July 1, 2022

**Copyright:** © 2022 Santiago, Sindi. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All data and code necessary to replicate the results of this study are available at the following github: [https://github.com/FS-CodeBase/propagon\\_replication\\_and\\_transmission/](https://github.com/FS-CodeBase/propagon_replication_and_transmission/).

**Funding:** FS and SS received funding in part from the National Science Foundation (Award No. 1840265 and 1429783), and the Joint Initiative to Support Research at the Interface of the Biological and Mathematical Sciences between the National

## Abstract

Prion proteins cause a variety of fatal neurodegenerative diseases in mammals but are generally harmless to Baker's yeast (*Saccharomyces cerevisiae*). This makes yeast an ideal model organism for investigating the protein dynamics associated with these diseases. The rate of disease onset is related to both the replication and transmission kinetics of propagons, the transmissible agents of prion diseases. Determining the kinetic parameters of propagon replication in yeast is complicated because the number of propagons in an individual cell depends on the intracellular replication dynamics and the asymmetric division of yeast cells within a growing yeast cell colony. We present a structured population model describing the distribution and replication of prion propagons in an actively dividing population of yeast cells. We then develop a likelihood approach for estimating the propagon replication rate and their transmission bias during cell division. We first demonstrate our ability to correctly recover known kinetic parameters from simulated data, then we apply our likelihood approach to estimate the kinetic parameters for six yeast prion variants using propagon recovery data. We find that, under our modeling framework, all variants are best described by a model with an asymmetric transmission bias. This demonstrates the strength of our framework over previous formulations assuming equal partitioning of intracellular constituents during cell division.

## Author summary

In this work we investigate the transmissible [*PSI*<sup>+</sup>] phenotype in yeast. The agents responsible for this phenotype are propagons, misfolded protein aggregates of a naturally occurring protein. These propagons increase in number within a cell and are distributed between cells during division. We use mathematical modeling to infer the replication rate of propagons within cells and if propagons are transmitted equally or unequally during cell division. Prior models in this area assumed only symmetric transmission when fitting replication rates. We couple this model with a novel likelihood framework allowing us to exclude influential outliers from our datasets when inferring parameters. We find that for

Science Foundation Division of Mathematical Sciences (NSF-DMS) and National Institutes of Health National Institute of General Medical Sciences (NIH-NIGMS) (Award No. GM126548). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

all six protein variants we study, propagons are transmitted asymmetrically with different biases. Our results can be reproduced with the code and data available at [https://github.com/FS-CodeBase/propagon\\_replication\\_and\\_transmission/](https://github.com/FS-CodeBase/propagon_replication_and_transmission/).

## Introduction

Today in the United States, millions of individuals suffer from Alzheimer's disease and other similar dementias. The cause of these diseases is thought to be the accumulation of misfolded proteins in the brain [1]. Currently, Alzheimer's is the sixth-leading cause of death in the United States and the fifth-leading cause of death for individuals 65 and over [1]. Beyond Alzheimer's, there are many other disorders caused by protein misfolding. Neurodegenerative diseases such as Parkinson's and Huntington's disease to less well-known diseases like Kuru and Creutzfeldt-Jakob Disease [2, 3], have different pathology but are characterized by the accumulation of amyloid aggregates. In addition, protein misfolding disorders have also been observed in other mammals such as scrapie in sheep, chronic wasting disease in deer and Bovine Spongiform Encephalopathy in cattle [4, 5]. Collectively, these diseases are untreatable and invariably fatal. While the details differ, all share two key commonalities: (1) a misfolded form of a protein *appears*; (2) the misfolded form of the protein forms aggregates, which then spread their folded confirmation through contact with other normally folded proteins. Because of the growing body of literature suggesting potential regulatory effects of prions (see [6] for example), in our work below we will refer to prion protein as alternatively folded protein instead of misfolded protein.

A promising biological system allowing for insight into alternatively folded mammalian protein and prion disease is the yeast *Saccharomyces cerevisiae*. A number of harmless, heritable phenotypes in yeast are shown to be transmitted vertically to new daughter cells by heritable prion elements termed propagons [4, 7]. Because these propagons are harmless to the yeast cells, researchers are able to study the protein dynamics themselves in the absence of harming the host. Indeed, biologists have many experimental tools that have been developed for yeast which allow for a detailed interrogation of the protein aggregation system in ways that are not possible to do in mammalian systems *in vivo* [6]. However, a unique challenge to working with yeast as a model system is that during the experimental time course, the yeast cells themselves continue to divide [8]. Without cell division, we know that the number of propagons in a cell will increase until it reaches a steady-state concentration where the propagon number is in balance with the soluble protein level [8–11]. However, the number of propagons in a cell will decrease when the cell divides as any propagons (i.e., transmissible aggregates, see [7] for more details) will be separated between the resulting mother and daughter cells [4, 8]. This means that for meaningful quantitative comparisons between mathematical models and experimental systems, care must be taken to understand the population averages but also heterogeneity amongst cells in the same population.

In this work, we consider not only the protein dynamics within yeast cells, but the cellular populations themselves. More specifically, in this work we develop a structured population model of yeast cells where the transmission of prion propagons are tracked as cells divide and as propagons replicate. We use our model to infer both the amplification rate and transmission bias of propagons from six prion variants (some are variants of  $[PSI^+]$  and some are mutants, see [12] for full details). This work is not the first to use the structured population framework, nor to infer a quantitative rate for transmission bias. A structured population model was previously applied to propagon data in [13]. The transmission bias of propagons has been directly

observed using propagon counting assays (see [14] for example). In 2009, Byrne et al. [15] used propagon curing data to quantitatively link the propagon transmission bias to the relative volumes of mother and daughter cells at cytokinesis. In comparison to these prior studies, our work offers several novel contributions. First, the model we develop here is more general, allowing for propagon amplification and asymmetric transmission. Second, our inference framework is more general. In contrast to [13], rather than assuming the variance in the propagon counts is proportional to the system mean, we employ the full likelihood of the data. In contrast to [15], we consider both propagon replication and transmission. Finally, we study a larger set of prion variants.

With this more general framework, we find that all six variants are best described by an asymmetric transmission of propagons between actively dividing yeast cells. With our model selection framework we were also able to exclude influential outliers from our prion variant datasets for computing kinetic parameters of prion variants. Moreover, we find differences among the variants for both the propagon replication rate and the transmission bias. Encouragingly, prion variants with similar phenotypic properties are fit with similar kinetic parameters. As such, our framework offers the ability to infer meaningful properties about prion variants even with our simplified model of intracellular propagon dynamics.

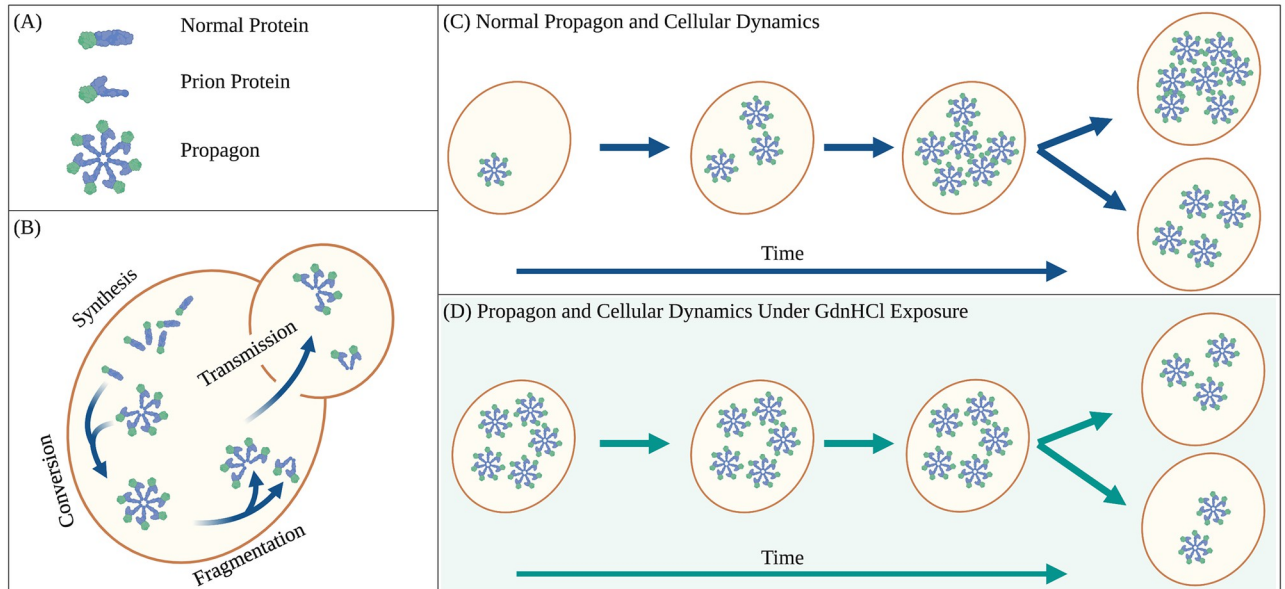
In the biological background section we develop the background of prion variants and the recovery assays we model. The methods section describes our propagon and generation structured population model and the likelihood approach we use for fitting kinetic parameters to experimental data. In the results section we first characterize the ability of our model and inference framework to recover the correct kinetic parameters and then apply our model and inference framework to recovery data from six distinct prion variants. In the discussion and conclusion section we discuss the implications of our study as well as factors to be considered in future studies on prion propagon dynamics.

## Biological background

As mentioned in the introduction, yeast prions were not discovered in the context of a disease but in one of mysterious heritable phenotypes [16, 17]. In addition, there is a considerably shorter history of knowledge about yeast prions than their mammalian counterparts [6, 9]. The  $[PSI^+]$  phenotype in yeast that we now know to be linked to a prion form of the protein Sup35, was discovered in 1965 by biologist Brian Cox. The phenotype corresponded to that of a white colored colony and the ability to grow a colony on media lacking adenine [16]. Remarkably, this phenotype appeared to be heritable, but the nature of transmission was by some unknown nonchromosomal cytoplasmic element [18]. For years researchers searched for a DNA or RNA basis for  $[PSI^+]$ . In 1994, Wickner, following insights from mammalian diseases [19, 20], hypothesized that  $[PSI^+]$  and  $[URE3]$  (another phenotype whose determinant was mysterious) were propagated by an alternatively folded (prion) form of their respective proteins [21]. In 1996, Paushkin et al. [22] and Patino et al. [23], separately demonstrated that the  $[PSI^+]$  phenotype was the result of an alternatively folded form of the protein Sup35 that was self-propagating. Today we know that many proteins in yeast are capable of forming prions and that a given prion protein may have multiple variants—distinct alternatively folded conformations—each of which is capable of propagating through this self-propagation process. Indeed, it is possible that prions may, at least in yeast, offer beneficial functions including serving as heritable bet-hedging devices diversifying microbial phenotypes [24]. In this study, as mentioned in the introduction, we consider propagon amplification data from six variants of the  $[PSI^+]$  prion in yeast [12].

For mammalian prion disease, the disease phenotype is observed at the level of a single organism (i.e., a cow, human, mouse, etc). However, in yeast prion biology the prion phenotypes are visible at the single-cell level with a different assay where the phenotype is observed with a yeast colony consisting of many cells founded by a single cell with the prion phenotype [25]. As such, it has been challenging to establish a precise link between the infective species, or propagon (that necessarily) reside in a single cell with the colony level phenotype [4]. While it is clear that the presence of a single propagon in the founding cell is necessary for the appearance of the prion phenotype at the colony level, it is not clear that it is sufficient [7]. In this work, following others [8, 13, 26] we will assume that the presence of a single propagon in a founding cell is both necessary and sufficient for the appearance of the prion phenotype at the colony level. As such, yeast prion dynamics are inherently a multi-scale process [27, 28].

Prion phenotypes occur when an alternatively folded form of a protein occurs, and rather than be cleared by cellular quality control machinery, the form persists and associates in aggregates (propagons), ordered structures of prion monomers as shown in Fig 1A. More specifically, four steps are essential to the maintenance of prion phenotypes (see Fig 1B). First, normal protein is continually produced by the cell. Second, prion propagons convert the normally folded protein to its alternatively folded confirmation through a templated conversion process when an aggregate incorporates the normally folded protein. This conversion process increases the size of the prion propagon. Third, the total number of propagons increases when the propagons are fragmented. This increases the total number of templating units and thus accelerates the conversion process. Fourth, propagons are transmitted between cells during division. Intriguingly, the cell division process creates an interesting phenomenon from the perspective of a single cell. Between cell divisions, the number of propagons increases, and it



**Fig 1. Multiscale yeast prion aggregate (propagon) dynamics.** (A) Within each cell in the colony is a mixture of normal protein and prion (alternatively folded) protein. Prion proteins are contained in propagons of multiple alternatively folded monomers. (B) Within each cell normal protein is produced (synthesis) and incorporated into existing propagons which leads to conversion of the normal protein to the prion form and increases the size of a propagon. Aggregates may increase in number by fragmentation and must be spread from mother to daughter cells during division (transmission). (C) Under normal growth conditions, the number of propagons increases during the lifetime of a cell and is split during cell division. (D) When cells are grown under GdnHCl fragmentation is assumed to stop and the number of propagons remains unchanged during the lifetime of a cell.

<https://doi.org/10.1371/journal.pcbi.1010107.g001>

will then decrease when cell division occurs Fig 1C. Due to the low frequency of spontaneous  $[PSI^+]$  appearance,  $\sim 10^{-8} - 10^{-7}$ /generation [29], in this work we assume there is no spontaneous appearance of propagons.

Yeast biologists take advantage of the fact that when cells are exposed to Guanadine Hydrochloride (GdnHCl) cell division is not impacted, but the propagon fragmentation process is assumed to halt. As such, the number of propagons within a cell, and indeed in the entire population is kept constant, while the number of cells in the colony continues to increase, see Fig 1D. This allows for experiments which probe the number of propagons in a single cell by regrowing colonies and assuming that any cell with at least one propagon will create a colony with a prion phenotype. A single yeast cell with propagons is introduced to a GdnHCl environment and the population of cells is allowed to grow normally. Because propagons will be split between mother and daughter cells during division, the expected number of propagons per cell will continue to decrease until the point where a cell in the population is extremely unlikely to have more than one propagon [26, 30, 31]. Then each cell in this population is allowed to form their own colony under normal growth conditions.

In propagon recovery experiments, yeast biologists use GdnHCl exposure in two phases to observe the amplification of propagons (see Fig 2). In the first phase, yeast cells are treated with GdnHCl until the propagons have sufficiently diluted. In this first phase, as propagons present at the beginning of GdnHCl exposure will exist for all time, the expected number of propagons per cell will continue to decrease as cell division continues as normal. In time, cells in the colony will contain a very low number of propagons per cell, ideally one. In the second phase, the resulting cells from phase one are transferred to a GdnHCl free environment and are allowed to form individual colonies. The number of white colonies is then assumed to correspond exactly to the number of cells with at least one propagon (see Fig 2). This is because the resulting colony whose founding cell had at least one propagon will have the  $[PSI^+]$  prion phenotype (white) while those founded by a cell with no propagons will have the  $[psi^-]$  non-prion phenotype (red).

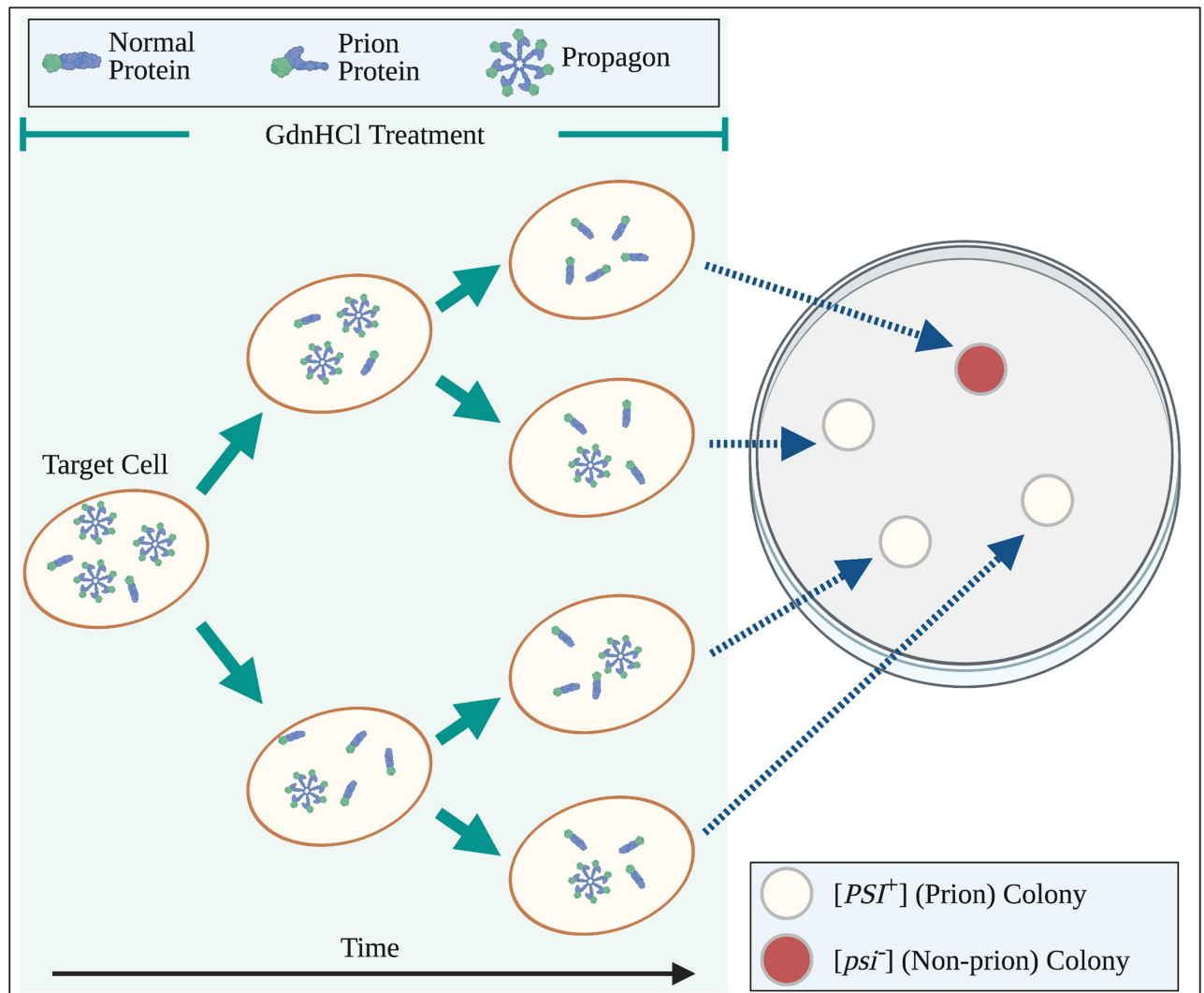
## Results

In this section we demonstrate that we can recover known parameters from simulated data using the adaptive Metropolis (AM) algorithm with our likelihood formulation, and study the effects of hourly sampling rate on these estimates. We then apply these parameter estimation methods to experimental data for six prion variants and perform model selection between symmetric division against asymmetric transmission of propagons during cell division.

### Parameter inference on simulated data

We first verify the capability of our likelihood formulation, Eq (16), and the AM algorithm 1, detailed in the methods to estimate known kinetic parameter values from simulated data before applying our methods to the experimental data. The simulated datasets are created using rejection sampling methods [32] on the asymmetric transmission of propagons (ATP) model presented in the methods. We take the cell division rates  $\alpha_i(t)$ , to be constant during each period of cell division, with a rate of  $\alpha_i(t) = 0.46 \text{ hr}^{-1}$ , and that death is negligible throughout the duration of the experiment by setting  $\beta_i(t) = 0$ . Furthermore, we assume the initial distribution of propagons to be a truncated normal distribution defined on the interval  $R = (0, \infty)$  with  $\mu = 10$  and  $\sigma = 1$ , that is  $Y(a) = \mathcal{N}^+(a; \mu = 10, \sigma = 1)$ , to simulate a colony with low number of propagons at the beginning of the experiment.





**Fig 2. Propagon amplification assay.** A two-step process is used to count the number of transmissible prion aggregates (propagons) in a single cell. *Left:* A single target cell is isolated, and propagon fragmentation is stopped through exposure to GdnHCl. Since propagons (pinwheels) cannot increase in number, they are diluted through cell division (green arrows). *Right:* After sufficient dilution, i.e. each yeast cell is likely to contain at most one propagon, the colony is replated onto solid media. In the absence of GdnHCl, each single cell serves as a founder of a distinct yeast colony. The number of propagons in the target cell corresponds to the number of white colonies in the plate.

<https://doi.org/10.1371/journal.pcbi.1010107.g002>

#### Algorithm 1 Adaptive Metropolis Algorithm

```

1:  $D \leftarrow$  load data
2:  $L \leftarrow$  load likelihood formulation
3:  $M \leftarrow$  load maximum number of iterations
4:  $k \leftarrow$  load non- adaptive period length
5:  $\theta^{(0)} \leftarrow$  initial value(s) for  $\theta$ 
6:  $\mathbf{v}^{(0:k)} \leftarrow$  initial covariance matrix values for non-adaptive period  $V_0$ 
7:  $sp \leftarrow$  set covariance scaling design parameter  $s_p$ 
8:  $ep \leftarrow$  set identity scaling  $\varepsilon$  for positive definite  $V$ 
9: for  $i = 1$  to  $k$  do ▷ Non-adaptive period
10:    $\theta_{new} \leftarrow N(\theta^{(i-1)}, \mathbf{v}^{(i-1)})$ 
11:    $u \sim U(0, 1)$ 
12:    $\alpha \leftarrow \min \{1, L(\theta_{new} | D) / L(\theta^{(i-1)} | D)\}$ 

```

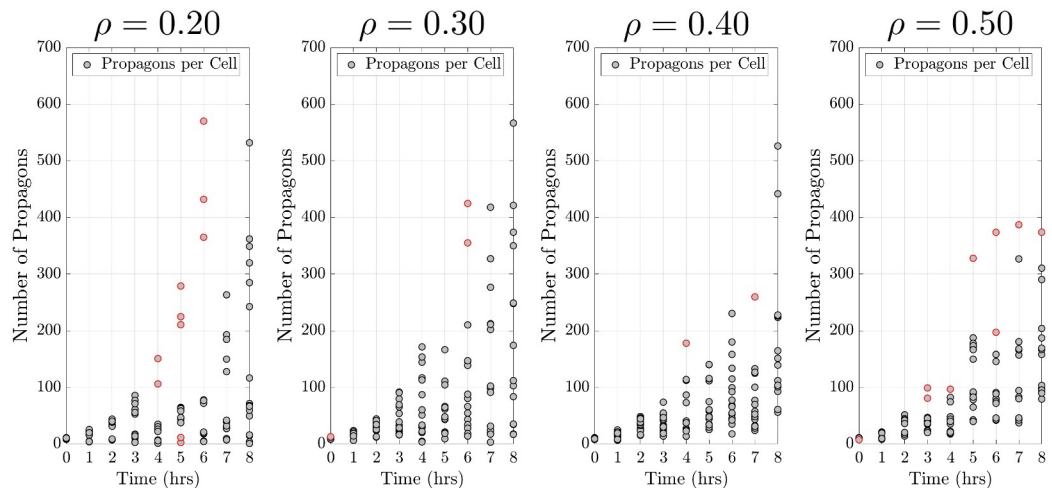
```

13: if  $u < \alpha$  then
14:    $\theta^{(i)} \leftarrow \theta_{new}$ 
15: else
16:    $\theta^{(i)} \leftarrow \theta^{(i-1)}$ 
17: for  $i = k + 1$  to  $M$  do ▷ Adaptive period
18:    $\theta_{new} \leftarrow N(\theta^{(i-1)}, \mathbf{v}^{(i-1)})$ 
19:    $u \sim U(0, 1)$ 
20:    $\alpha \leftarrow \min \{1, L(\theta_{new} | D) / L(\theta^{(i-1)} | D)\}$ 
21:   if  $u < \alpha$  then
22:      $\theta^{(i)} \leftarrow \theta_{new}$ 
23:   else
24:      $\theta^{(i)} \leftarrow \theta^{(i-1)}$ 
25:    $\mathbf{v}^{(i)} = sp * cov(\theta^{(0)}, \theta^{(1)}, \dots, \theta^{(i)}) + ep * I_p$ 
26: return  $\theta, \mathbf{V}$ 

```

In our investigation we considered the effects of replication rates, transmission biases, and sampling rates on our ability of recover known kinetic parameter values. Fig 3 shows four examples of simulated data produced by rejection sampling in the cases of four intracellular constituents biases  $\rho = 0.20, 0.30, 0.40, 0.50$ , a replication rate of  $\lambda = 0.70 \text{ hr}^{-1}$ , and a sampling rate of 16 samples per hour. The noise in the simulated data presented in Fig 3 is a consequence of simulating data with rejection sampling from the distribution of propagons over time in the ATP model.

Parameter estimations are made using the likelihood formulation and the AM algorithm 1 outlined in the methods section. We generate simulated data with three sampling rates: 8, 16, and 32 samples per hour, with replication rates  $\lambda = 0.5, 0.7, 0.9 \text{ (hr}^{-1}\text{)}$  and transmission bias  $\rho = 0.2, 0.3, 0.4, 0.5$ , for 36 total possible combinations. The sampling rates were chosen to reflect the experimental data which contain approximately 16 samples per hour per dataset. To robustly assess the effects of the sampling rates on our ability to recover known parameter choices, we tested our methods on 500 generated datasets for each of the 36 sampling rate and parameter combinations. The AM algorithm 1 was applied to each of the 500 simulated datasets and each resulting AM chain iteration is used to compute a mean parameter estimate. This results in 500 singleton parameter estimates for each of the two parameters. We use these



**Fig 3. Simulated data.** The simulated data was generated using a replication rate of  $\lambda = 0.70 \text{ hr}^{-1}$  and four different transmission biases ( $\rho$ ). Samples were generated per experimental hour at a rate of 16 samples per hour. Data points outlined in red were determined to be outliers by the IQR method.

<https://doi.org/10.1371/journal.pcbi.1010107.g003>

**Table 1. Credible intervals (95%) for parameter estimates on simulated data.** The table summarizes the parameter inference results for twelve  $(\lambda, \rho)$  parameter pairs and three sampling rates using data simulated from the ATP model.

| $\theta$  |        | 8 Samples/Hour |             | 16 Samples/Hour |             | 32 Samples/Hour |             |
|-----------|--------|----------------|-------------|-----------------|-------------|-----------------|-------------|
| $\lambda$ | $\rho$ | $\lambda$      | $\rho$      | $\lambda$       | $\rho$      | $\lambda$       | $\rho$      |
| 0.5       | 0.2    | (0.49,0.52)    | (0.19,0.21) | (0.50,0.52)     | (0.19,0.20) | (0.50,0.51)     | (0.20,0.20) |
|           | 0.3    | (0.49,0.53)    | (0.28,0.32) | (0.49,0.51)     | (0.29,0.31) | (0.50,0.51)     | (0.30,0.31) |
|           | 0.4    | (0.49,0.51)    | (0.38,0.42) | (0.50,0.51)     | (0.39,0.41) | (0.50,0.51)     | (0.39,0.41) |
|           | 0.5    | (0.49,0.50)    | (0.48,0.49) | (0.50,0.50)     | (0.49,0.50) | (0.50,0.50)     | (0.49,0.50) |
| 0.7       | 0.2    | (0.69,0.72)    | (0.19,0.21) | (0.70,0.72)     | (0.19,0.21) | (0.70,0.71)     | (0.20,0.20) |
|           | 0.3    | (0.69,0.73)    | (0.28,0.32) | (0.69,0.71)     | (0.29,0.31) | (0.69,0.71)     | (0.30,0.31) |
|           | 0.4    | (0.69,0.72)    | (0.38,0.42) | (0.70,0.71)     | (0.39,0.41) | (0.70,0.71)     | (0.39,0.41) |
|           | 0.5    | (0.69,0.70)    | (0.48,0.49) | (0.70,0.70)     | (0.48,0.50) | (0.70,0.70)     | (0.49,0.50) |
| 0.9       | 0.2    | (0.89,0.92)    | (0.19,0.21) | (0.90,0.92)     | (0.19,0.21) | (0.90,0.91)     | (0.20,0.20) |
|           | 0.3    | (0.89,0.93)    | (0.29,0.32) | (0.89,0.91)     | (0.29,0.31) | (0.90,0.91)     | (0.30,0.31) |
|           | 0.4    | (0.89,0.92)    | (0.39,0.42) | (0.90,0.91)     | (0.39,0.41) | (0.90,0.91)     | (0.39,0.41) |
|           | 0.5    | (0.89,0.91)    | (0.48,0.49) | (0.90,0.90)     | (0.48,0.50) | (0.90,0.90)     | (0.49,0.50) |

<https://doi.org/10.1371/journal.pcbi.1010107.t001>

500 singleton parameter estimates to form a 95% credible interval for each of the 36 parameter and sampling rate combinations. We opted not to store each of the 500 AM chain iterations and combine them to form the credible intervals due to the burden on storage capacity. We present these results in [Table 1](#).

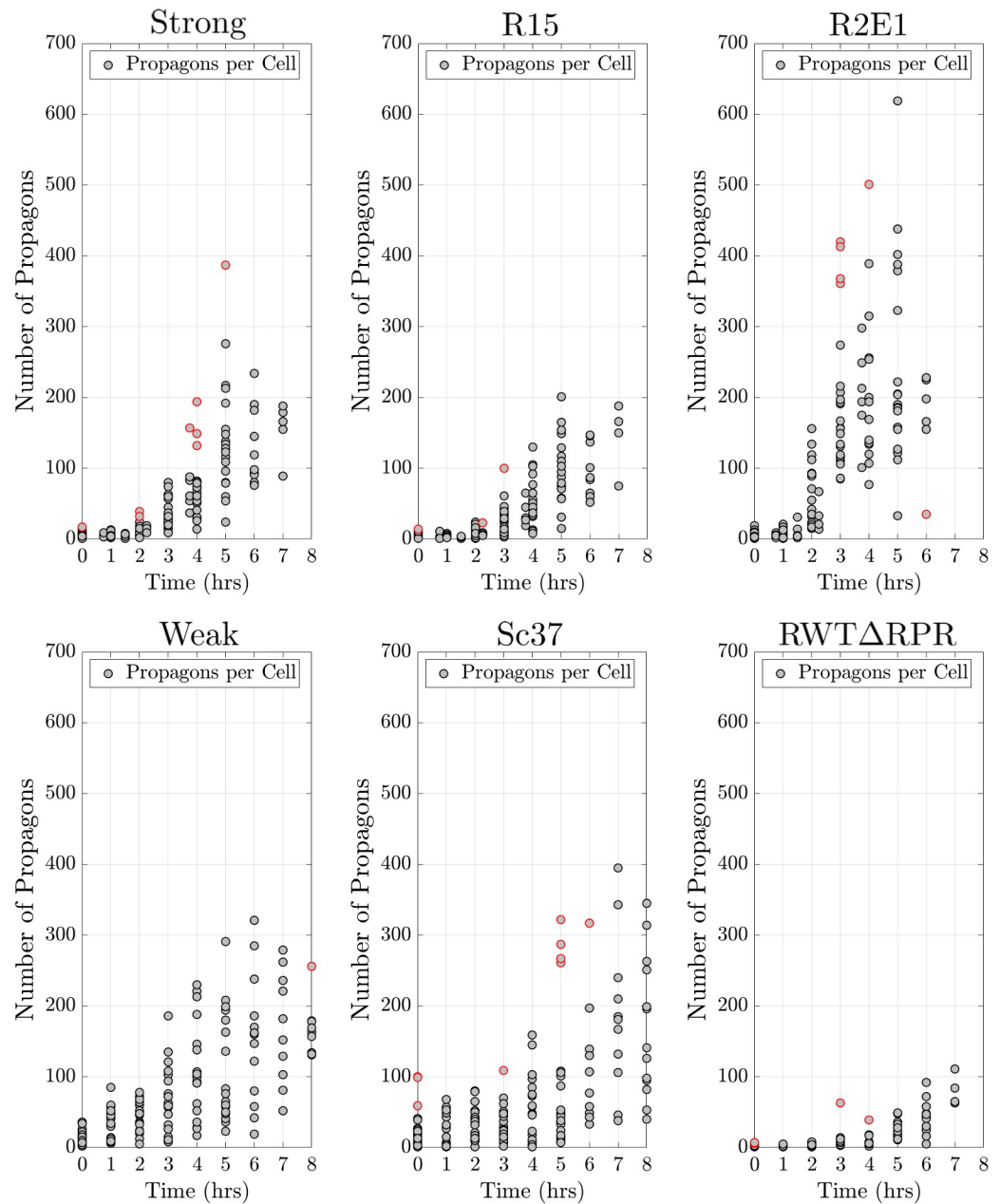
In all cases we found that we can successfully detect differences in replication rates and asymmetric transmission biases. As expected, we observed that increasing the sampling rate led to more precise parameter estimates. In the cases where we estimate a symmetric transmission bias ( $\rho = 0.5$ ), the credible intervals never capture this value. That is due to the fact that we estimate  $\rho$  in the interval  $[0, 0.5]$  because of the symmetry in our model about the point  $\rho = 0.5$  in the interval  $[0, 1]$ . However, in [Table 1](#) we see that as we increase the sampling rate, the estimates get closer to the true value of  $\rho = 0.5$ . In the experimental data we are interested in removing outliers using the interquartile range (IQR) method [33], so we studied the effects of removing possible “outliers” or extreme values using this method from our simulated data to study the effects on our ability to recover the true parameter values. We found that removing such outliers from our simulated data led to a slight improvement in our ability to recover the true parameter values (see [Table F](#) in [S1 Text](#)).

### Parameter inference on experimental data

After verifying that our methods allows us to recover known replication rate and transmission bias parameters from simulated datasets, we apply our methods to experimental data from six prion variants. We consider the propagon replication experiments for six Sup35 variants that have alternative folds, aggregate to form propagons, and transmit the  $[PSI^+]$  phenotype: Weak, Sc37, Strong [34], RWTΔRPR, R15, and R2E1 [12]. The experimental results from propagon recovery assays for the six prion variants are presented in [Fig 4](#). Note that all variants exhibit heteroscedasticity in the number of propagons in time.

We consider two variations of our model for the division of propagons among dividing yeast cells. First we consider the model  $Z_S := Z_S(t, a; \lambda, \rho = 0.5)$ , symmetric division of propagons between dividing cells where our model collapses to the model first proposed by [35] and we estimate the replication rate  $\lambda$  and fix the transmission bias at  $\rho = 0.5$ . Secondly we consider





**Fig 4. Experimental propagon counts for six prion variants.** This experimental data was obtained through propagon recovery experiments (see the biological background section for more details). Data points outlined in red were determined to be outliers by the IQR method.

<https://doi.org/10.1371/journal.pcbi.1010107.g004>

the model  $Z_A := Z_A(t, a; \lambda, \rho)$ , asymmetric transmission of propagons where we estimate both the propagon replication rate  $\lambda$  and the propagon transmission bias  $\rho$ .

The general formulation of our likelihood also allows us to consider the possibility that the start of the exponential phase of prion amplification begins at a time after the start of the experiment (Fig 4). For this, application of the AM algorithm on the experimental data  $\{a(t_i)\}_{i=1}^m$  is performed by using the data where  $t_i \geq T_\Delta$  for  $T_\Delta$  hours into the experiment (see Tables A-E in S1 Text). Then we apply model selection to determine  $T_\Delta$ , the point of time into

**Table 2. Parameter estimates and credible intervals (95%) for six prion variants.** The column labeled  $T_{\Delta}$  indicates the point of time into the experiment that best describes the beginning of the exponential phase of propagon replication. The column labeled %W presents the percent model weight for  $Z_A$ , the model for asymmetric division of propagons. The asterisk (\*) indicates the dataset not filtered for outliers (raw data) was selected using AIC<sup>c</sup>.

| Prion Variant    | Replication Rate ( $\lambda$ , hr <sup>-1</sup> ) | Transmission Bias ( $\rho$ , proportion) | $T_{\Delta}$ (hrs) | %W    | Model   |
|------------------|---|--|--------------------|-------|---------|
| Weak             | 0.88(0.87,0.88)                                   | 0.27(0.26,0.27)                          | 0*                 | 99.34 | $Z_A^*$ |
| Sc37             | 0.81(0.79,0.82)                                   | 0.23(0.22,0.23)                          | 0                  | 100.0 | $Z_A$   |
| RWT $\Delta$ RPR | 0.77(0.72,0.80)                                   | 0.31(0.29,0.33)                          | 2                  | 100.0 | $Z_A$   |
| Strong           | 1.13(1.11,1.16)                                   | 0.36(0.34,0.38)                          | 0.75               | 100.0 | $Z_A$   |
| R15              | 1.14(1.10,1.18)                                   | 0.32(0.29,0.35)                          | 0.75               | 100.0 | $Z_A$   |
| R2E1             | 1.25(1.22,1.29)                                   | 0.29(0.26,0.33)                          | 0                  | 100.0 | $Z_A$   |

<https://doi.org/10.1371/journal.pcbi.1010107.t002>

the experiment that best describes the beginning of the exponential phase of propagon replication.

In Table 2, we present the replication rate, transmission bias, the start of exponential growth phase, and the percent model weight for the model,  $Z_A$  or  $Z_S$ , that best explains each dataset. Through our analysis we found that a model which includes the cell replication rate and asymmetric transmission of propagons during cell division ( $Z_A$ ), best explains the experimental propagon recovery data for all prion variants (model weights > 99%, Table 2). In the case of the Weak prion variant, we found that the dataset where outliers were kept and an asymmetric division model,  $Z_A$ , best explained the propagon dynamics for the Weak variant with a model weight of 99.34%. For each variant we detected differences in the replication rates and transmission bias. Across the six prion variants we detected propagon transmission biases, the proportion of propagons transmitted to daughter cells, to be between 22% and 38%,  $\rho \in (0.22, 0.38)$ . For the three prion variants RWT $\Delta$ RPR, Strong, and R15, we found a delay in the start of the exponential growth phase of propagon replication. RWT $\Delta$ RPR was found to have the largest delay ( $T_{\Delta} = 2$  hrs) before the beginning of the exponential growth phase, which was identified from experimental data without the outliers identified through the IQR method. Without the work to detect  $T_{\Delta}$ , we underestimate the replication rate for this variant (see Tables C, G, and H in S1 Text).

The parameter estimates, corrected AIC, percent model weights, and the postprocessing of the Metropolis chain iterations for all the cases considered are summarized in the methods section.

## Discussion and conclusion

In this work we presented a structured population model that generalizes the work presented in [13, 35], to study propagon replication dynamics and the bias in their transmission among proliferating yeast cells. We developed an inverse problem formulation that consists of an interpretable likelihood formulation descriptive of the propagon recovery experiment. We first verified that we could recover known propagon replication parameters from simulated data, then used our inverse problem formulation to study propagon replication in six prion variants. Additionally, the likelihood formulation allowed us to consider the presence of influential outliers and if there was a delay in observing the exponential phase of propagon replication in the experimental data.

We were able to detect differences in propagon replication rates and bias in their transmission during cell division among the six prion variants that we studied. We found that the Weak prion variant had a lower replication rate and transmission bias,  $\lambda = 0.88$  hr<sup>-1</sup> and  $\rho = 0.27$  respectively, when compared to the Strong prion variant which had a higher replication

**Table 3. Relative volume of daughter cells by generation.** Where  $V_M$  is the mother cell volume and  $V_D$  is the daughter volume. This table was adapted from Table 2 in [37].

| Generation        | 1    | 2    | 3    | 4    |
|-------------------|------|------|------|------|
| $V_D/(V_D + V_M)$ | 0.40 | 0.32 | 0.29 | 0.25 |

<https://doi.org/10.1371/journal.pcbi.1010107.t003>

rate and transmission bias,  $\lambda = 1.13 \text{ hr}^{-1}$  and  $\rho = 0.36$  respectively. These findings are consistent with previous studies of propagon size [7, 36], and fragmentation rates [8, 25] of Weak and Strong propagons, where Weak propagons are larger and replicate at a lower rate than Strong propagons, thus it would be reasonable to observe a higher rate of propagon replication and a more significant transmission bias in Strong than in Weak.

In this work we detected differences in transmission biases for each of the six prion variants, however the reason for these differences cannot be found with our current model formulation. We would expect the transmission bias to be proportional to the volume of the daughter cell to that of the mother and daughter volumes or  $\rho_1 \approx V_D/(V_M + V_D) \approx 40\%$ , and  $\rho_2 \approx V_M/(V_M + V_D) \approx 60\%$ , as has been found in [15] for different prion variants not considered in this work. From Table 3 adapted from [37], these ratios are generation dependent. However, our estimates for the transmission biases remain close to these generational dependent values. Therefore, it is possible that the differences in transmission biases we are capturing come from these generation based volume differences. In future models, we will need to incorporate generation based volume differences to discern differences in transmission biases from generation based daughter cell volume.

To our knowledge our work is the first to use propagon amplification assays and a structured population model (ATP model) to recover both the propagon replication rate and the transmission bias during cell division. The work in [13] uses propagon amplification assays and a structured population model that assumes symmetric transmission of propagons between dividing cells to determine propagon replication rates for two yeast variants but did not directly address the asymmetric bias in propagon transmission during cell division. Separate work presented in [15] has used curing experiments, where the application of GdnHCl causes inhibition of propagon replication and leads to elimination of the  $[PSI^+]$  phenotype through dilution by cell division and a model that captures asymmetric cell division through unequal transmission of propagons through cell division [38], to recover the propagon transmission bias. They considered the propagon transmission bias as the probability of propagon transmission to a budding daughter cell in their work. The curing experiment eliminates  $[PSI^+]$  phenotype and as such, a propagon replication rate cannot be computed. We note that our model can also be used to calculate the asymmetric transmission bias using data from curing experiments by setting the replication rate equal to zero ( $\lambda = 0 \text{ hr}^{-1}$ ). The number of propagons will decrease through asymmetric transmission of propagons between dividing cells.

In this work we detail the importance of the propagon replication rate  $\lambda$ , the asymmetric propagon transmission bias  $\rho$  and detection of a delay  $T_\Delta$ , before observing the exponential growth phase of prion replication, to explain prion replication in a proliferating yeast cell colony. In future work we intend to build on this framework by incorporating cell maturation, a state during which cells can grow but cannot divide until they are fully mature. We have made an effort to be clear about the difference between *asymmetric cell division* and *asymmetric transmission of intracellular constituents*, but in future models we plan to incorporate cell volume into our models, so that we can distinguish between effects due to volume and those due

to transmission bias. Also, in this work our intracellular propagon replication model assumed exponential growth because a clear steady state, or carrying capacity, was not observed in the experimental data, but perhaps by running the propagon amplification assays over a longer period of time we would begin to observe such a steady state. This would require the application of a more complex intracellular model to capture such a carrying capacity and would require methods such as those presented in [39] to numerically solve such a propagon replication model. As previously mentioned, another possible application with our modeling framework that we have not yet considered is the possibility of its application to study curing experiments.

## Methods

In this section we begin by presenting our model, a system of partial differential equations (PDEs) for the intracellular process of propagon replication and their transmission through the cellular process of division. We present intermediate quantities that allow the decoupling of the PDE system and present the explicit solutions to the model that we consider in this work. Then we show that the intermediate quantities used to derive explicit solutions facilitate a likelihood formulation for parameter estimation and model selection. We describe how to generate simulated data with the model solutions and that we can recover the true kinetic parameters with our likelihood formulation and our implementation of the adaptive Metropolis (AM) algorithm. We conclude this section by detailing how we overcome numerical issues with the implementation of our likelihood formulation.

### Asymmetric transmission of propagons model

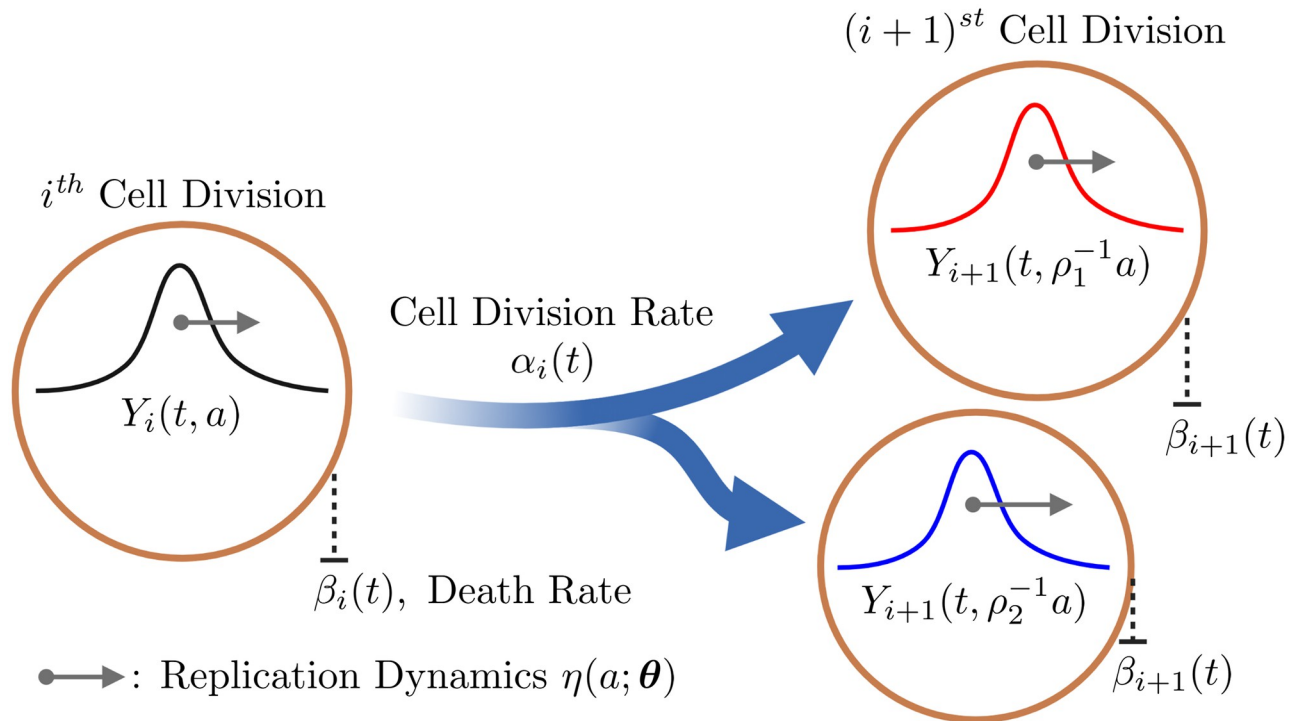
We seek to model the number of propagons or a single prion variant in a population of actively dividing cells. Let  $a(t)$  be the number of propagons a cell has  $t$  hours after dividing, then

$$\frac{da}{dt} = \eta(a; \theta) \quad (1)$$

where  $\eta$  is the intracellular propagon amplification model that depends on the current number of propagons  $a$  and  $\theta$ , the kinetic parameter(s) that govern the propagon replication and transmission dynamics (i.e. replication rate and transmission bias). We model the propagon distribution dynamics in the yeast cell population  $Y(t, a)$ , as evolving in time according to the transport equation:

$$\frac{\partial}{\partial t} Y(t, a) + \frac{\partial}{\partial a} (\eta(a; \theta) Y(t, a)) = 0. \quad (2)$$

However, we are interested in tracking  $Y_i(t, a)$ , the distribution of propagons in cells that have undergone  $i$  divisions,  $t$  hours since the start of the experiment (see Fig 5). These propagon dynamics in the population of dividing cells are captured by the following system of  $M + 1$



**Fig 5. Asymmetric transmission of propagons model schematic.** The model dynamics of intracellular propagon replication and cell division from generation  $i$  to generation  $i + 1$ . The black arrow ( $\rightarrow$ ) illustrates the intracellular increase in the number of propagons over time. The remaining parameters are detailed in the methods.

<https://doi.org/10.1371/journal.pcbi.1010107.g005>

coupled PDEs, which we refer to as the Asymmetric Transmission of Propagons (ATP) model:

$$\begin{aligned}
 \frac{\partial}{\partial t} Y_0(t, a) + \frac{\partial}{\partial a} (\eta(a; \theta) Y_0(t, a)) &= -(\alpha_0(t) + \beta_0(t)) Y_0(t, a), \\
 \frac{\partial}{\partial t} Y_1(t, a) + \frac{\partial}{\partial a} (\eta(a; \theta) Y_1(t, a)) &= -(\alpha_1(t) + \beta_1(t)) Y_1(t, a) + D_1(t, a), \\
 &\vdots \\
 \frac{\partial}{\partial t} Y_i(t, a) + \frac{\partial}{\partial a} (\eta(a; \theta) Y_i(t, a)) &= -(\alpha_i(t) + \beta_i(t)) Y_i(t, a) + D_i(t, a), \\
 &\vdots \\
 \frac{\partial}{\partial t} Y_M(t, a) + \frac{\partial}{\partial a} (\eta(a; \theta) Y_M(t, a)) &= -(\alpha_M(t) + \beta_M(t)) Y_M(t, a) + D_M(t, a).
 \end{aligned} \tag{3}$$

The right hand side of each PDE captures the rate of cells lost in that generation from cell division and cell death, and accounts for new cells from cell division in the previous generation. Here  $\alpha_i(t)$  is the rate of cell division, and  $\beta_i(t)$  the rate of cell death in the  $i^{\text{th}}$  generation. The rate of increase in the number of cells in the  $i^{\text{th}}$  generation is represented by the term

$$D_i(t, a) = \rho_1^{-1} \alpha_{i-1}(t) Y_{i-1}(t, \rho_1^{-1} a) + \rho_2^{-1} \alpha_{i-1}(t) Y_{i-1}(t, \rho_2^{-1} a), \tag{4}$$



where  $\rho_1$  and  $\rho_2$  control the transmission of propagons between dividing cells. To maintain the conservation of propagons during division we require  $\rho_1 + \rho_2 = 1$ , showing only one degree of freedom or that if  $\rho_1 = \rho \in (0, 1)$ , then  $\rho_2 = 1 - \rho$ . Note that letting  $\rho_1 = \rho_2 = 1/\gamma$  reduces our model to that proposed by [35]. Also, note that this formulation implies that for conservation and symmetric transmission of propagons during cell division  $\rho = 1/\gamma = 1/2$ . The work by [40] first generalized the work by [35] to study T cells traced with carboxyfluorescein diacetate succinimidyl ester to include their resting and cycling phases (delay in division) while considering an asymmetric division with a system of delay hyperbolic PDEs. However, in our modeling framework because of the duration of the experiment (8 hrs), we expect most cell divisions to be the product of mature cells actively producing daughter cells so we do not consider synchronous delays in division.

In this work, we are interested in determining the effect of  $\rho$ , the propagon transmission biased between dividing cells. To this end, we can bound the number of cell divisions ( $M$ ) because the duration of the experiment of interest is finite. We bound the maximum number of cell divisions up to  $M = 6$  generations because yeast cells divide every 1.5 hours and the longest duration of the propagon recovery experiments considered in this work is eight hours. Then to solve the ATP model, Eq (3), we must specify initial distributions for each generation

$$Y_0(0, a) = Y(a) \text{ and } Y_i(0, a) = 0 \text{ for all } i > 0. \quad (5)$$

Where  $Y(a)$ , is the initial intracellular distribution of propagons at the start of the experiment.

### Analytic solutions and model decomposition

The formulation of the ATP model, Eqs (3)–(5) presented in the methods, allows for computation of intermediate quantities that allow for the decoupling of population and intracellular propagon dynamics in Eq (3). First, the total number of cells resulting from the  $i^{\text{th}}$  division, since the beginning of the experiment, at time  $t$  is given by

$$n_i(t) = \int_R Y_i(t, a) da, \quad (6)$$

where  $R$  depends on the intracellular propagon amplification dynamics model in Eq (1), and is taken to be  $R = (0, \infty)$  in our work (see Corollary 1). The normalized density of propagons can then be defined by

$$y_i(t, a) = \frac{Y_i(t, a)}{n_i(t)}, \text{ for } n_i(t) > 0, \quad (7)$$

and  $y_i(t, a) = 0$  otherwise. This quantity represents the intracellular dynamics of the dividing cells in our population of interest. We define the initial number of cells at the start of the experiment by

$$N_0 = \int_R Y_0(0, a) da, \quad (8)$$

and the initial normalized propagon density

$$y_0(0, a) = \frac{Y_0(0, a)}{N_0}. \quad (9)$$

Using the quantities defined above and Eq (3), the following theorem holds for Eq (1) defined by an exponential growth model  $\eta(a; \theta) = \lambda a$ , where  $a \in R = (0, \infty)$ .

**Theorem 1.** *The solution of the system defined by Eqs (3) and (4), with initial conditions given by Eq (5) is given by:*

$$Y_i(t, a) = n_i(t)y_i(t, a), \text{ for } 0 \leq i \leq M, \tag{10}$$

in which:

1.  $n_i(t)$  is the solution of the system of ODEs:

$$\begin{aligned} i = 0 : \quad & \frac{dn_0}{dt} = -(\alpha_0(t) + \beta_0(t))n_0, \\ \text{for } 1 \leq i \leq M : \quad & \frac{dn_i}{dt} = -(\alpha_i(t) + \beta_i(t))n_i + 2\alpha_{i-1}(t)n_{i-1}, \end{aligned} \tag{11}$$

where  $n_0(0) = N_0$ , and  $n_i(0) = 0$  for  $i \geq 1$ .

2.  $y_i(t, a)$  is the solution to the PDE

$$\frac{\partial y_i(t, a)}{\partial t} + \frac{\partial(\eta(a; \theta)y_i(t, a))}{\partial a} = 0 \tag{12}$$

with initial conditions  $y_i(0, a) := (\frac{1}{2})^i \sum_{k=0}^i \binom{i}{k} \rho_1^{k-i} \rho_2^{-k} \Upsilon(\rho_1^{k-i} \rho_2^{-k} a)$ , and

3. the solution  $y_i(t, a)$  satisfies the recursive property

$$y_i(t, a) = \frac{1}{2} (\rho_1^{-1} y_{i-1}(t, \rho_1^{-1} a) + \rho_2^{-1} y_{i-1}(t, \rho_2^{-1} a)), \tag{13}$$

for all  $0 \leq i \leq M$ .

Unlike the theorems presented in [35] and [40], in Theorem 1 we highlight the fact that  $y_i(t, a)$  must satisfy Eqs (12) and (13), in order for Eq (10) to be a solution to the system defined by Eq (3). This is to highlight that for non-linear intracellular dynamics models (Eq (1)), the property given by Eq (13) may no longer be satisfied [39]. The proof of Theorem 1 follows the structure of the proof in [35], therefore we simply outline the proof of Theorem 1 here.

*Proof.* First, we substitute the  $Y_i(t, a)$  terms in Eq (3) with the decomposition given by Eq (10). Then we simplifying the terms of Eq (3), and using Eqs (11) and (12) leads to the recursive expression in Eq (13) completing the proof.

In this work we are modeling propagon replication where at the beginning of the experiment the initial condition is that of a distribution of low propagon counts in the yeast cell population as observed in the propagon recovery data (see Fig 4). This initial low number of propagons is followed by a period of exponential growth where a steady state in the number of propagons is not yet observed. Corollary 1 captures this phase of growth, where we assume that the rate of propagon replication is proportional to the current number of propagons present within a yeast cell. This type of propagon proliferation assumes that there is an unlimited amount of normally folded (soluble) protein that can be alternatively folded and lead to the continued formation of prion propagons.

**Corollary 1.** *The solution of the system defined by Eqs (3) and (4) with an intracellular propagon replication model  $\eta(t, a) = \lambda a$ , is*

$$Y_i(t, a) = n_i(t) \sum_{k=0}^i \binom{i}{k} \rho_1^{k-i} \rho_2^{-k} \exp(-\lambda t) \Upsilon(\rho_1^{k-i} \rho_2^{-k} a \exp(-\lambda t)) \text{ for } 0 \leq i \leq M. \tag{14}$$

Where  $n_i(t)$  is the solution to Eq (11).

*Proof.* Following [40], we solve Eq (12) via the method of characteristics. This yields

$$y_i(t, a) = \left(\frac{1}{2}\right)^i \sum_{k=0}^i \binom{i}{k} \rho_1^{k-i} \rho_2^{-k} \exp(-\lambda t) \Upsilon(\rho_1^{k-i} \rho_2^{-k} a \exp(-\lambda t)) \text{ for } 0 \leq i \leq M.$$

To show that  $y_i(t, a)$  satisfies the recursive property in Eq 13 we prove the equivalent expression  $y_{i+1}(t, a) = \frac{1}{2}(\rho_1^{-1}y_i(t, \rho_1^{-1}a) + \rho_2^{-1}y_i(t, \rho_2^{-1}a))$  holds, see S1 Text. Then, replacing  $y_i(t, a)$  in Eq (10), and inserting this result in Eq (3) proves Corollary 1.

The following corollary presented in [35] involves the solution of Eq (11) under a special case of constant cell division rate  $\alpha_i(t) = \alpha$  and constant cell death  $\beta_i(t) = \beta$ . In [35] the authors consider  $\beta_i(t) = \beta > 0$ , but the solution also holds for  $\beta = 0$ , which is the assumption in this work.

**Corollary 2.** Let  $\alpha_i(t) = \alpha \geq 0$  and  $\beta_i(t) = \beta \geq 0$ , for all  $0 \leq i \leq M$ , the solution to Eq (11) is

$$n_i(t) = \frac{(2\alpha t)^i}{i!} \exp(-(\alpha + \beta)t) N_0, \text{ for } 0 \leq i \leq M. \tag{15}$$

*Proof.* Straightforward by plugging Eq (15) into Eq (11).

### Likelihood problem formulation and model selection

In this work we are interested in estimating the kinetic parameters of our model, Eqs (3)–(5) with the propagon amplification models presented in the methods, using experimental data. We are interested in the probability of the observed data for a given value of the kinetic parameters  $\theta$  of the parameter space  $\Theta$ , denoted  $L(\theta|Data)$ . Let  $N(t) = \sum_{i=0}^M n_i(t)$ , the total number of cells at time  $t$ ,  $Z(t, a; \theta) = \sum_{i=0}^M Y_i(t, a)$  represent the total yeast cell population propagon density after  $M$  cell divisions, and  $\theta$  represent the kinetic parameters. Further, let the experimental observations  $a_k$  observed at time  $t_k$  consisting of propagon counts be  $\{(a_k, t_k)\}_{k=1}^m$  (see the Results section). Then, the likelihood of the kinetic parameters  $\theta$ , given the propagon data is defined as follows

$$\begin{aligned} L(\theta|Data) &= \prod_{k=1}^m P(\{t_k, a_k\}; \theta), \\ &= \prod_{k=1}^m \left\{ \sum_{i=0}^M \left[ \frac{\binom{n_i(t_k)}{a_k}}{\binom{N(t_k)}{a_k}} y_i(t_k, a_k; \theta) \right] \right\}, \\ &= \prod_{k=1}^m \sum_{i=0}^M \frac{Y_i(t_k, a_k; \theta)}{N(t_k)}, \\ &= \prod_{k=1}^m \frac{Z(t_k, a_k; \theta)}{N(t_k)}. \end{aligned} \tag{16}$$

where  $P(\{t_k, a_k\}; \theta)$  is the probability of observing  $a_k$  propagons at time  $t_k$  given model parameters  $\theta$ . This probability is the product of  $\frac{\binom{n_i(t_k)}{a_k}}{\binom{N(t_k)}{a_k}}$ , the probability that the observation  $a_k$  came from a cell in the  $i^{th}$  cell division and  $y_i(t_k, a_k; \theta)$ , the probability that the cell in the  $i^{th}$  cell division at time  $t_k$  contains  $a_k$  propagons. The product of both terms is summed over the number of divisions that can be observed during the experiment. This is because a cell at any point in the experiment results from a finite number of cell divisions since the beginning of the experiment.

In this work we are interested in considering two mathematical modeling scenarios. The first  $Z_S(t, a; \lambda, \rho = 0.5)$ , symmetric cell division where we estimate the posterior distribution of the replication rate ( $\lambda$ ) and fix the transmission bias at  $\rho = 0.5$ , and the second  $Z_A(t, a; \lambda, \rho)$ , asymmetric cell division where we estimate the posterior distributions of the replication rate  $\lambda$  and the posterior distribution of the propagon transmission bias  $\rho$ . With the likelihood formulation we compute the Akaike Information Criterion (AIC)

$$AIC = -2 \cdot \log(L(\hat{\theta}|Data)) + 2K,$$

where  $\hat{\theta}$  are the mean parameter estimates that best explain the data and  $K$  is the number of free parameters [41]. To account for bias due to different number of parameters and data sizes, we use the AIC bias correction

$$AIC^c = AIC + \frac{2K(K + 1)}{m - K - 1},$$

where  $m$  is the sample size [42]. These  $AIC^c$  values are used to compute the  $AIC^c$  difference

$$\Delta_i = AIC_i^c - \min(AIC^c),$$

for the  $i^{th}$  model and prion variant data. Finally, we can compute the relative model weights

$$W_i = \frac{\exp(-\frac{1}{2}\Delta_i)}{\sum_{m=1}^M \exp(-\frac{1}{2}\Delta_m)},$$

of each model [43] with each prion variant dataset. We interpret this quantity as the probability that a model is the best approximation to the replication and division of propagons during cellular proliferation given the experimental data.

### Adaptive metropolis algorithm

In this work we follow the procedure described in [44], a Metropolis algorithm with an adaptive Metropolis (AM) step to estimate the target distribution  $\pi(\theta)$ , with  $p$  kinetic parameter variables  $\theta$ . Using an initial value for each kinetic parameter  $\theta_0$ , as the starting condition, a random candidate  $\theta_{new}$  is drawn from a proposal distribution  $J$  of the parameters  $\theta$ . Thus  $\theta_{new} \sim J(\theta_{new}|\theta_{i-1})$ , is drawn in every iteration.

The target distribution  $\pi(\theta)$  is given using previously defined likelihood function  $L(\theta|Data)$ ,

$$\pi(\theta) = \frac{L(\theta|Data)\pi_0(\theta)}{\pi(Data)}, \tag{17}$$

where the random parameter variables  $\theta$  have a known and possibly uninformative prior density  $\pi_0(\theta)$ ,  $\pi(Data) = \int_{\mathbb{R}^p} L(\theta|Data)\pi_0(\theta)d\theta$ , and  $p$  is the dimension of the parameter set. In this work we choose the noninformative prior  $\pi_0(\theta) = U(0, 1)^p$  and  $J(\theta_{new}|\theta_{i-1}) = N(\theta_{i-1}; V)$  to be normally distributed with covariance matrix  $V$ . The acceptance probability then follows

$$\alpha(\theta_{new}|\theta_{i-1}) = \min\left\{1, \frac{\pi(\theta_{new})J(\theta_{i-1}|\theta_{new})}{\pi(\theta_{i-1})J(\theta_{new}|\theta_{i-1})}\right\} = \min\left\{1, \frac{L(\theta_{new}|Data)}{L(\theta_{i-1}|Data)}\right\}. \tag{18}$$

In Eq (18), the terms involving  $J$  cancel because by design  $J$  is a symmetric proposal distribution. Now, with probability  $\alpha(\theta_{new}|\theta_{i-1})$ , we accept  $\theta_{new}$ , and set  $\theta_i := \theta_{new}$ . Otherwise, with probability  $1 - \alpha(\theta_{new}|\theta_{i-1})$ , we set  $\theta_i := \theta_{i-1}$ .

Following [45], the adaptive step is preceded by a non-adaptive period of length  $k$  where  $\theta_1, \theta_2, \dots, \theta_k$  are computed using an initial covariant matrix  $V_0 = V$ . Following the non-adaptive step, the covariance matrix is computed using the previous chain values

$$V_i = s_p \text{cov}(\theta_0, \theta_1, \dots, \theta_i) + \varepsilon I_p, \text{ for } i \geq k. \tag{19}$$

Here  $s_p$  is a design parameter that depends on  $p$ . The term  $\varepsilon I_p$  consists of the  $p$ -dimensional identity matrix and  $\varepsilon \geq 0$  to ensure that  $V_i$  remains positive definite. This formulation can quickly incur large computational cost, but this cost can be drastically reduced by use of the recursive update of the covariance [45]

$$V_i = \frac{i-2}{i-1} V_{i-1} + \frac{s_p}{i-1} \left( (i-1) \hat{\theta}_{i-2} \hat{\theta}_{i-2}^T - i \hat{\theta}_{(i-1)} \hat{\theta}_{(i-1)}^T + \theta_{(i-1)} \theta_{(i-1)}^T \right). \tag{20}$$

In the application of the adaptive Metropolis algorithm to the simulated data and experimental data, we take  $s_p = 1.0, \varepsilon = 1 \times 10^{-6}$ , and

$$V_0 = \begin{bmatrix} \sigma_{\lambda\lambda}^2 & \sigma_{\lambda\rho}^2 \\ \sigma_{\rho\lambda}^2 & \sigma_{\rho\rho}^2 \end{bmatrix},$$

where  $\sigma_{\lambda\lambda}^2 = 0.1406, \sigma_{\lambda\rho}^2 = \sigma_{\rho\lambda}^2 = 0$ , and  $\sigma_{\rho\rho}^2 = 0.0156$ . A non-adaptive period length of  $k = 1500$  for the simulated data and experimental data were established using Geweke’s convergence diagnostic as presented in [46], and performed a total of  $2 \times 10^6$  iterations. We avoid numerical issues in the direct evaluation of the likelihood formulation presented in Eq (16) by working with natural-logarithm version of the acceptance probability in Eq (18) using the numerically stable procedure outlined in the methods section.

To reduce the correlation among the AM estimates ( $\theta$ ), estimates after the burn-in period and after thinning at regular intervals of 50 iterations are used in our work. We use the integrated autocorrelation time (iac) as a measure of autocorrelation across the autocorrelation function (ACF) [47]. The chain iteration and final iacs are included in Figs A and B in S1 Text.

### Numerical implementation of the likelihood formulation

In our initial implementation of the likelihood formulation, we encountered significant numerical underflow, so we opted to work with the natural-logarithm form of the likelihood or log-likelihood. Recalling that  $N(t) = \sum_{i=0}^M n_i(t)$  and that  $Z(t) = \sum_{i=0}^M Y_i(t, a)$ , taking the natural logarithm of Eq (16), we have

$$\begin{aligned} \ln L(\theta|Data) &= \sum_{k=1}^m \ln(Z(t, a; \theta)) - \ln(N(t)), \\ &= \sum_{k=1}^m \ln \left( \sum_{i=0}^M Y_i(t_k, a_k; \theta) \right) - \ln \left( \sum_{i=0}^M n_i(t_k) \right). \end{aligned} \tag{21}$$

Numerical evaluation of this log-likelihood is performed by generalizing the algebraic and numerically stable property  $\ln(a_0 + a_1) = \ln(a_0) + \ln(1 + \exp(\ln(a_1) - \ln(a_0)))$  where  $\ln(a_0) > \ln(a_1)$ . That is we evaluate  $\ln(\sum_{i=0}^M Y_i(t_k, a_k; \theta))$  by applying the natural log directly to each  $i^{\text{th}}$  analytic model solution  $Y_i(t, a; \theta)$ , then evaluating the resulting expression. The evaluation of  $\ln(\sum_{i=0}^M Y_i(t_k, a_k; \theta))$  is then found recursively using the  $\ln(Y_i(t_k, a_k; \theta))$  terms as follows: let  $\ln q_i = \ln Y_i(t, a; \theta)|_{t=t_k, a=a_k}$ , and  $\mathbf{q} = \{\ln q_j\}_{j=0}^M$  be the sorted values  $\ln q_i$  in descending order,



then  $\ln(\sum_{i=0}^M q_i)$  is given by the numerically stable recursive function  $\xi$  as follows

$$\xi(\mathbf{q}) = \begin{cases} \mathbf{q}(1) & |\mathbf{q}| = 1 \\ \mathbf{q}(1) + \ln(1 + \exp(\xi(\mathbf{q}(\sim 1)) - \mathbf{q}(1))) & |\mathbf{q}| > 1 \end{cases} \quad (22)$$

where  $|\cdot|$  is the cardinality or the number of elements in  $\mathbf{q}$  and the notation “ $\sim 1$ ” indicates all elements in  $\mathbf{q}$  except the first. Implementing this formulation removed numerical underflow when evaluating the likelihood formulation (Eq (16)).

## Supporting information

**S1 Text. Recursive property to corollary 1, additional notes on parameter estimation, model selection, and experimental data.**

(PDF)

## Acknowledgments

We would like to acknowledge Dr. Tricia Serio for providing us with the experimental data used in this investigation and for the insightful discussions regarding this work.

## Author Contributions

**Conceptualization:** Fabian Santiago, Suzanne Sindi.

**Data curation:** Fabian Santiago, Suzanne Sindi.

**Formal analysis:** Fabian Santiago, Suzanne Sindi.

**Funding acquisition:** Suzanne Sindi.

**Investigation:** Fabian Santiago, Suzanne Sindi.

**Methodology:** Fabian Santiago, Suzanne Sindi.

**Project administration:** Fabian Santiago, Suzanne Sindi.

**Resources:** Fabian Santiago, Suzanne Sindi.

**Software:** Fabian Santiago.

**Supervision:** Suzanne Sindi.

**Validation:** Fabian Santiago, Suzanne Sindi.

**Visualization:** Fabian Santiago, Suzanne Sindi.

**Writing – original draft:** Fabian Santiago, Suzanne Sindi.

**Writing – review & editing:** Fabian Santiago, Suzanne Sindi.

## References

1. Association A. 2019 Alzheimer’s disease facts and figures. *Alzheimer’s & dementia*. 2019; 15(3):321–387. <https://doi.org/10.1016/j.jalz.2019.01.010>
2. Soto C. Unfolding the role of protein misfolding in neurodegenerative diseases. *Nature Reviews Neuroscience*. 2003; 4(1):49–60. <https://doi.org/10.1038/nrn1007> PMID: 12511861
3. Head MW. *Human prion diseases*. Cambridge University Press; 2014.
4. Sindi SS, Serio TR. Prion dynamics and the quest for the genetic determinant in protein-only inheritance. *Current opinion in microbiology*. 2009; 12(6):623–630. <https://doi.org/10.1016/j.mib.2009.09.003> PMID: 19864176

5. Prusiner SB. Novel Proteinaceous Infectious Particles Cause Scrapie. *Science*. 1982; 216(4542):136–144. <https://doi.org/10.1126/science.6801762> PMID: 6801762
6. Tuite MF, Serio TR. The prion hypothesis: from biological anomaly to basic regulatory mechanism. *Nature Reviews Molecular Cell Biology*. 2010; 11(12):823–833. <https://doi.org/10.1038/nrm3007> PMID: 21081963
7. Villali J, Dark J, Brechtel TM, Pei F, Sindi SS, Serio TR. Nucleation seed size determines amyloid clearance and establishes a barrier to prion appearance in yeast. *Nature structural & molecular biology*. 2020; 27(6):540–549. <https://doi.org/10.1038/s41594-020-0416-6> PMID: 32367069
8. Derdowski A, Sindi SS, Klaips CL, DiSalvo S, Serio TR. A size threshold limits prion transmission and establishes phenotypic diversity. *Science*. 2010; 330(6004):680–683. <https://doi.org/10.1126/science.1197785> PMID: 21030659
9. Sindi SS. Mathematical Modeling of Prion Disease. In: *Prion-An Overview*. InTech; 2017. p. 207–227.
10. Greer ML, Pujo-Menjouet L, Webb GF. A mathematical analysis of the dynamics of prion proliferation. *Journal of Theoretical Biology*. 2006; 242(3):598–606. <https://doi.org/10.1016/j.jtbi.2006.04.010> PMID: 16753184
11. Pruss J, Pujo-Menjouet L, Webb GF, Zacher R. Analysis of a model for the dynamics of prions. *Discrete and Continuous Dynamical Systems—Series B*. 2006; 6(1):225–235. <https://doi.org/10.3934/dcdsb.2006.6.225>
12. Langlois CR, Pei F, Sindi SS, Serio TR. Distinct prion domain sequences ensure efficient amyloid propagation by promoting chaperone binding or processing in vivo. *PLoS genetics*. 2016; 12(11):e1006417. <https://doi.org/10.1371/journal.pgen.1006417> PMID: 27814358
13. Banks H, Flores KB, Langlois CR, Serio TR, Sindi SS. Estimating the rate of prion aggregate amplification in yeast with a generation and structured population model. *Inverse Problems in Science and Engineering*. 2018; 26(2):257–279. <https://doi.org/10.1080/17415977.2017.1316498>
14. Cox B, Ness F, Tuite M. Analysis of the generation and segregation of propagons: entities that propagate the [PSI<sup>+</sup>] prion in yeast. *Genetics*. 2003; 165(1):23–33. <https://doi.org/10.1093/genetics/165.1.23> PMID: 14504215
15. Byrne LJ, Cole DJ, Cox BS, Ridout MS, Morgan B, Tuite MF. The number and transmission of [PSI<sup>+</sup>] prion seeds (Propagons) in the yeast *Saccharomyces cerevisiae*. *PLoS One*. 2009; 4(3):e4670–e4670. <https://doi.org/10.1371/journal.pone.0004670> PMID: 19262693
16. Cox B. *psi*, a cytoplasmic suppressor of super-suppressor in yeast. *Heredity*. 1965; 20(4):505. <https://doi.org/10.1038/hdy.1965.65>
17. Tuite MF, Cox BS. The [PSI<sup>+</sup>] prion of yeast: a problem of inheritance. *Methods*. 2006; 39(1):9–22. <https://doi.org/10.1016/j.jymeth.2006.04.001> PMID: 16757178
18. Tuite MF, Staniforth GL, Cox BS. [PSI<sup>+</sup>] turns 50. *Prion*. 2015; 9(5):318–332. <https://doi.org/10.1080/19336896.2015.1111508> PMID: 26645632
19. Weissmann C. A 'unified theory' of prion propagation. *Nature*. 1991; 352(6337):679–683. <https://doi.org/10.1038/352679a0> PMID: 1876183
20. Prusiner SB. Molecular biology of prion diseases. *Science*. 1991; 252(5012):1515–1522. <https://doi.org/10.1126/science.1675487> PMID: 1675487
21. Wickner RB. [URE3] as an altered URE2 protein: evidence for a prion analog in *Saccharomyces cerevisiae*. *Science*. 1994; 264(5158):566–569. <https://doi.org/10.1126/science.7909170> PMID: 7909170
22. Patino MM, Liu JJ, Glover JR, Lindquist S. Support for the prion hypothesis for inheritance of a phenotypic trait in yeast. *Science*. 1996; 273(5275):622–626. <https://doi.org/10.1126/science.273.5275.622> PMID: 8662547
23. Paushkin S, Kushnirov V, Smirnov V, Ter-Avanesyan M. Propagation of the yeast prion-like [PSI<sup>+</sup>] determinant is mediated by oligomerization of the SUP35-encoded polypeptide chain release factor. *The EMBO journal*. 1996; 15(12):3127–3134. <https://doi.org/10.1002/j.1460-2075.1996.tb00675.x> PMID: 8670813
24. Newby GA, Lindquist S. Blessings in disguise: biological benefits of prion-like mechanisms. *Trends in cell biology*. 2013; 23(6):251–259. <https://doi.org/10.1016/j.tcb.2013.01.007> PMID: 23485338
25. Tanaka M, Collins SR, Toyama BH, Weissman JS. The physical basis of how prion conformations determine strain phenotypes. *Nature*. 2006; 442(7102):585–589. <https://doi.org/10.1038/nature04922> PMID: 16810177
26. Sindi SS, Olofsson P. A Discrete-Time Branching Process Model of Yeast Prion Curing Curves. *Mathematical Population Studies*. 2013; 20(1):1–13. <https://doi.org/10.1080/08898480.2013.748566> PMID: 30287982

27. Banwarth-Kuhn M, Collignon J, Sindi S. Quantifying the Biophysical Impact of Budding Cell Division on the Spatial Organization of Growing Yeast Colonies. *Applied Sciences*. 2020; 10(17):5780. <https://doi.org/10.3390/app10175780>
28. Lemarre P, Pujo-Menjouet L, Sindi SS. Generalizing a mathematical model of prion aggregation allows strain coexistence and co-stability by including a novel misfolded species. *Journal of mathematical biology*. 2019; 78(1):465–495. <https://doi.org/10.1007/s00285-018-1280-4> PMID: 30116882
29. Lancaster AK, Bardill JP, True HL, Masel J. The spontaneous appearance rate of the yeast prion [*PSI*<sup>+</sup>] and its implications for the evolution of the evolvability properties of the [*PSI*<sup>+</sup>] system. *Genetics*. 2010; 184(2):393–400. <https://doi.org/10.1534/genetics.109.110213> PMID: 19917766
30. Cole DJ, Ridout MS, Morgan BJT, Byrne LJ, Tuite MF. Approximations for Expected Generation Number. *Biometrics*. 2007; 63(4):1023–1030. <https://doi.org/10.1111/j.1541-0420.2007.00780.x> PMID: 17425634
31. Morgan BJT, Ridout MS, Ruddock LW. Models for Yeast Prions. *Biometrics*. 2003; 59(3):562–569. <https://doi.org/10.1111/1541-0420.00066> PMID: 14601757
32. Casella G, Robert CP, Wells MT, et al. Generalized accept-reject sampling schemes. In: *A festschrift for herman rubin*. Institute of Mathematical Statistics; 2004. p. 342–347.
33. Chihara LM, Hesterberg TC. *Mathematical statistics with resampling and R*. John Wiley & Sons; 2018.
34. Pei F, DiSalvo S, Sindi SS, Serio TR. A dominant-negative mutant inhibits multiple prion variants through a common mechanism. *PLoS genetics*. 2017; 13(10):e1007085. <https://doi.org/10.1371/journal.pgen.1007085> PMID: 29084237
35. Hasenauer J, Schittler D, Allgöwer F. Analysis and simulation of division-and label-structured population models. *Bulletin of mathematical biology*. 2012; 74(11):2692–2732. PMID: 23086287
36. Kryndushki DS, Alexandrov IM, Ter-Avanesyan MD, Kushnirov VV. Yeast [*PSI*<sup>+</sup>] Prion Aggregates Are Formed by Small Sup35 Polymers Fragmented by Hsp104. *Journal of Biological Chemistry*. 2003; 278(49):49636–43. <https://doi.org/10.1074/jbc.M307996200>
37. Woldringh C, Huls P, Vischer N. Volume growth of daughter and parent cells during the cell cycle of *Saccharomyces cerevisiae* a/alpha as determined by image cytometry. *Journal of bacteriology*. 1993; 175(10):3174–3181. <https://doi.org/10.1128/jb.175.10.3174-3181.1993> PMID: 8491731
38. Cole DJ, Morgan BJ, Ridout MS, Byrne LJ, Tuite MF. Estimating the number of prions in yeast cells. *Mathematical Medicine and Biology*. 2004; 21(4):369–395. <https://doi.org/10.1093/mammb/21.4.369> PMID: 15567890
39. Santiago F, Sindi S, Flores K. Numerical Approaches to Division and Label Structured Population Models. *Letters in Biomathematics*. 2020; 7(1):153–170.
40. Luzyanina T, Cupovic J, Ludwig B, Bocharov G. Mathematical models for CFSE labelled lymphocyte dynamics: asymmetry and time-lag in division. *Journal of mathematical biology*. 2014; 69(6):1547–1583. <https://doi.org/10.1007/s00285-013-0741-z> PMID: 24337680
41. Akaike H. A new look at the statistical model identification. *IEEE transactions on automatic control*. 1974; 19(6):716–723. <https://doi.org/10.1109/TAC.1974.1100705>
42. Hurvich CM, Tsai CL. Regression and time series model selection in small samples. *Biometrika*. 1989; 76(2):297–307. <https://doi.org/10.1093/biomet/76.2.297>
43. Posada D. *Bioinformatics for DNA sequence analysis*. Springer; 2009.
44. Smith RC. *Uncertainty quantification: theory, implementation, and applications*. Philadelphia: SIAM; 2014.
45. Haario H, Saksman E, Tamminen J, et al. An adaptive Metropolis algorithm. *Bernoulli*. 2001; 7(2):223–242. <https://doi.org/10.2307/3318737>
46. Cowles MK, Carlin BP. Markov chain Monte Carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association*. 1996; 91(434):883–904. <https://doi.org/10.1080/01621459.1996.10476956>
47. Roberts GO, Rosenthal JS. Optimal scaling for various Metropolis-Hastings algorithms. *Statistical Science*. 2001; 16(4):351–367. <https://doi.org/10.1214/ss/1015346320>