

Modelling prion dynamics in yeast

Ridout, Martin, Giagos, Vasileios and Morgan, Byron

University of Kent, School of Mathematics, Statistics and Actuarial Science

Canterbury, Kent, CT2 7NF, U.K.

E-mail: M.S.Ridout@kent.ac.uk, V.Giagos@kent.ac.uk, B.J.T.Morgan@kent.ac.uk

Naeimi, Wesley, von der Haar, Tobias and Tuite, Mick

University of Kent, School of Biosciences

Canterbury, Kent, CT2 7NJ, U.K.

E-mail: W.R.Naeimi@kent.ac.uk, T.von-der-Haar@kent.ac.uk, M.F.Tuite@kent.ac.uk

Introduction

The word prion was coined by Nobel laureate Stanley Prusiner to describe the infectious agent in a group of fatal neurodegenerative diseases that includes scrapie in sheep, bovine spongiform encephalopathy (BSE) in cattle and variant Creutzfeldt-Jakob disease (vCJD) in humans. The infectious agent is a misfolded form of the protein, PrP, denoted PrP^{Sc}, which forms polymers that grow by recruiting molecules of the normal cellular form of the protein, PrP^C. Once attached to the PrP^{Sc} complex, the PrP^C molecule refolds into the prion conformation.

A key property of prions is their ability to self-replicate, through a process known as *nucleated polymerization*. Nucleation refers to the fact that the minimal stable unit of PrP^{Sc} is a ‘nucleus’ or ‘seed’ consisting of several molecules. Very occasionally, such seeds can arise spontaneously; alternatively they are inherited or result from infection through eating meat from an infected animal. The seeds grow by polymerization and the polymers can subsequently fragment into two parts. Each fragment then acts as a separate seed, provided it is larger than the minimum stable size. This combination of growth and fragmentation provides the mechanism of replication.

Yeast prions are a group of proteins in the budding yeast *Saccharomyces cerevisiae* that, though functionally unrelated to PrP, also self-replicate through nucleated polymerization and can be inherited through cell division. However, yeast prions are not associated with disease in *S. cerevisiae*. Tuite and Serio (2010) provide a recent overview of the biology.

One widely-studied yeast prion is $[PSI^+]$, a prion form of the essential yeast-cell protein Sup35p. Sup35p is involved in ensuring that the translation of RNA transcripts into protein molecules terminates correctly. Normally it occurs as a soluble monomer (single molecule), but there is also a polymeric prion form that is insoluble. Yeast cells in which all the Sup35p is in the normal form are termed $[psi^-]$ whereas those that contain some Sup35p in the prion form are termed $[PSI^+]$. A colour assay is available to distinguish the two types of cell.

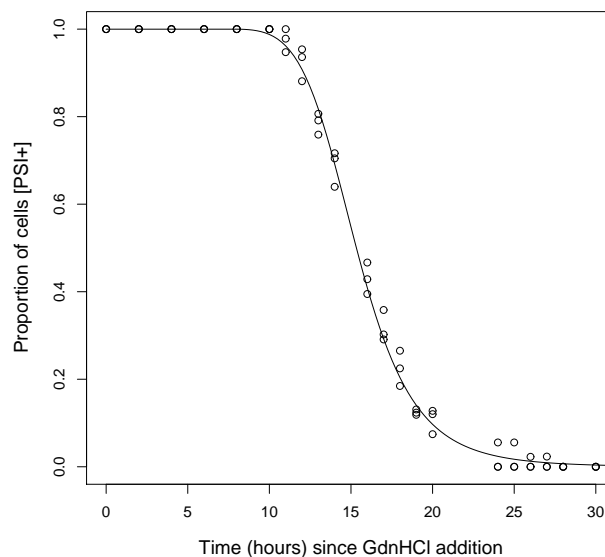
This paper gives an overview of a long-running collaboration between bioscientists and statisticians at the University of Kent, studying $[PSI^+]$. We adopt a chronological approach, showing how initial phenomenological models have been refined progressively both to incorporate greater biological realism and to probe deeper into the underlying mechanisms. We begin by describing $[PSI^+]$ curing by the chemical guanidine hydrochloride (GdnHCl), which has been an important experimental technique for understanding $[PSI^+]$, and outline the stochastic models that we have developed to describe curing data. We then look at mathematical models of nucleated polymerization and conclude by outlining current and future areas of work.

GdnHCl-induced curing of the $[PSI^+]$ prion

If GdnHCl is added at low concentrations to the growth medium of a culture of $[PSI^+]$ yeast

cells, the proportion of $[PSI^+]$ cells decreases gradually to zero over time, after an initial lag period where no effect is observable (Fig. 1). This is termed *curing* and occurs because GdnHCl inhibits replication of the prion polymers, which are then diluted through cell division until $[psi^-]$ cells with no polymers appear (Byrne *et al.*, 2007).

Figure 1. Curing data for the yeast strain YJW512. The fitted curve is based on the improved model of GdnHCl-induced prion curing described below.



Curing data are obtained by extracting samples of cells at various times after GdnHCl addition and determining the number of cells that are $[PSI^+]$. A reasonable starting point to model the data is to treat the number of $[PSI^+]$ cells at time t as binomial, with index equal to the number of cells sampled and with probability $p_+(t)$, say. In practice, replicate data sometimes show mild overdispersion relative to the binomial distribution.

Our initial work on yeast prions was concerned with deriving expressions for the function $p_+(t)$, which depends on the mean number of polymers per cell present initially, n_0 , and on the rate of cell division. The experimental data can then be used to estimate these parameters; the parameter of most interest biologically is n_0 . We outline two different models. The first contains the essential ideas, but assumes that cells divide symmetrically. The second incorporates a more realistic model of yeast cell division. To describe the models, we introduce two related terms. The *generation time* of a cell is the time from when the cell is generated, through a cell division event, until it too divides. The *generation number* of a cell at time t is the number of cell divisions that have occurred in the ancestry of this cell since time $t = 0$. Thus, cells present at the start of the experiment have generation number 0, these divide to give cells of generation number 1, and so on.

An initial model of GdnHCl-induced prion curing

Morgan *et al.* (2003) assume (i) that the number of polymers per cell at the outset is $\text{Poisson}(n_0)$, (ii) that this number is fixed when GdnHCl is added, (iii) that at cell division the polymers are divided independently and randomly between the two daughter cells and (iv) that the generation times of the cells are i.i.d. random variables. This latter assumption implies that the cell population grows as an *age-dependent branching process*.

Suppose that, at time t , we pick a cell at random from the population and that its generation time is $G(t)$. It follows from the assumptions that the number of prions in the cell is $\text{Poisson}\{n_0 2^{-G(t)}\}$,

and the probability that the cell is $[PSI^+]$ is therefore

$$(1) \quad p_+ \{t|G(t) = g\} = 1 - \exp \{-n_0 2^{-g}\}.$$

To obtain the unconditional probability $p_+(t)$, we require the distribution of $G(t)$. Suppose we pick a cell at random at the start of the experiment and follow a random path through its descendants until time t ; each time there is a cell division, we choose randomly between the two offspring cells, arriving finally at a particular cell at time t . Let $H(t)$ denote the generation number of this cell. This has a different distribution to $G(t)$ because the sampling protocol for arriving at the chosen cell is different. The distribution of $H(t)$ is available in terms of the generation-time distribution, using standard results in renewal theory. We then have the approximation

$$(2) \quad \Pr \{G(t) = g\} = \frac{2^g \Pr \{H(t) = g\}}{\sum_{h=0}^{\infty} 2^h \Pr \{H(t) = h\}},$$

which is the expected number of cells of generation g at time t divided by the expected total number of cells at time t . Morgan *et al.* (2003) fitted this model to several sets of curing data, assuming that cell division times follow a gamma distribution with known coefficient of variation. This provided estimates of n_0 and also of the mean cell generation time. Estimates of the latter parameter were shown to be in good agreement with estimates based on a different experimental technique.

Better models of yeast cell division

In reality, *S. cerevisiae* reproduces not by simple symmetric division, but by an asymmetric budding process. A small bud appears on a mother cell which grows until it is about two thirds of the size of the mother cell before separating from the mother. A mother cell can produce numerous buds in succession in this way until eventually the cell dies. The new daughter cell also produces buds in due course, but the first bud takes longer to produce than subsequent buds.

Green (1981) proposed a model for yeast cell division in which a mother cell takes time W_M to generate a bud to the point where it separates from the mother. The corresponding time for a newly generated daughter cell to produce its own offspring is $W_D + W_M$, where W_D represents the additional time that a daughter cell requires to mature to the point where it is similar to a mother cell. Different cells have independent realisations of the random variables W_D and W_M .

This is a generalisation of the age-dependent branching process. Green showed that, asymptotically, the population grows exponentially at rate ϕ , where ϕ , which is termed the *Malthusian parameter*, is the unique solution to the equation

$$[1 + E \{\exp(-\phi W_D)\}] E \{\exp(-\phi W_M)\} = 1.$$

Ridout *et al.* (2006) gave accurate approximations for ϕ based on just the first two moments of the distributions of W_M and W_D . Cole *et al.* (2007) provided similar approximations to the expected generation number of a randomly chosen cell. The parameter ϕ also determines the proportion of daughter cells in the population asymptotically, which is

$$P_D = E \{\exp(-\phi W_M)\}.$$

An improved model of GdnHCl-induced prion curing

Cole *et al.* (2004) extended the initial model of curing to allow for the asymmetrical budding process, by introducing an additional parameter π to represent the probability that a given prion polymer is passed from the mother cell to the daughter cell during budding; the earlier model assumed $\pi = 1/2$.

We can again consider a cell chosen at random from the population at time t , but now need to take into account not just the generation number of the cell, $G(t)$, but also the number of ancestors of the current cell that were daughter cells, which we denote by $D(t)$. The conditional probability that the cell is $[PSI^+]$ is then

$$p_+ \{t|G(t) = g, D(t) = d\} = 1 - \exp \left\{ -n_0 \pi^d (1 - \pi)^d \right\}.$$

To obtain the marginal probability $p_+(t)$, we need to sum over the joint distribution of $G(t)$ and $D(t)$. The model does not consider cell death, but this is reasonable because most cells in a dividing population are ‘young’; Cole et al. (2004) provided simulation evidence to support this conclusion.

The new model was fitted to the same data as in Morgan *et al.* (2003). This required assumptions about the distributions of W_M and W_D , which were taken to be $\text{Gamma}(\alpha_M, \beta)$ and $\text{Gamma}(\alpha_D, \beta)$, respectively. Choosing a common value of β means that $W_M + W_D$ also has a gamma distribution. The main parameters of interest are not those of the distributions of W_M and W_D but n_0 , the mean number of polymers at $t = 0$, as before, and π , the probability that a polymer is transmitted to the daughter cell. It is not practical to estimate all of the model parameters from curing data alone and supplementary cell growth data were used to estimate the parameters of the distributions of W_M and W_D .

Calculation of $p_+(t)$ requires the numerical evaluation of various integrals involving the densities of W_M and W_D . Ridout (2008) shows that $p_+(t)$ can be calculated efficiently using numerical Laplace transform inversion, for any distributions of W_M and W_D for which the Laplace transforms are available; the distributions need not be from the same family (Palmer *et al.*, 2008).

The analyses that we have described so far provide estimates of the mean number of polymers present in a cell, n_0 . An alternative experimental approach, proposed by Cox *et al.* (2003), is to isolate a single cell onto medium containing GdnHCl and let this cell reproduce. Because of the dilution of the prion polymers present in the parent cell through its subsequent descendants, eventually each of the descendants will either be $[psi^-]$ or will be $[PSI^+]$ but will contain only a single polymer. Therefore, at this point, the number of polymers present in the single initial cell can be determined simply by counting the number of $[PSI^+]$ cells amongst the descendants. Whilst this technique is laborious and has some experimental limitations, it does provide valuable additional data. Moreover, it is possible to separate mother-daughter pairs to determine the numbers of prion polymers that they contain, providing direct information about the transmission of polymers from mothers to daughters.

To gain further biological insight we need to consider what is occurring within an individual cell. We therefore turn our attention to mathematical models of the process of nucleated polymerization.

Models of nucleated polymerization

The prevailing model of prion aggregates, which is described by Masel *et al.* (1999), treats the aggregates as one-dimensional linear fibrils. This does not imply that the polymers are strictly one-dimensional at the microscopic level – for example, they may have a helical structure – but they are ‘macroscopically linear’. The model is deterministic and describes the system at time t by the number of monomers, $x(t)$, and the number of polymers of length i , $y_i(t)$; monomers are normal protein molecules whereas polymers are in the prion form. We also introduce two further variables, $y(t) = \sum y_i(t)$ and $z(t) = \sum iy_i(t)$. Thus $y(t)$ is the total number of polymers and $z(t)$ is the total number of protein molecules that are in polymers, or equivalently, the total length of the polymers.

Five types of reaction are considered:

1. **Production of a new monomer**, which occurs at rate $\alpha \mu\text{M min}^{-1}$.
2. **Binding of a monomer to a polymer of length i** . The monomer is converted to the prion form and therefore gives a prion polymer of length $i + 1$. For each polymer, this occurs at a rate $\beta x(t)$, where the units of β are $\mu\text{M min}^{-1}$; the overall rate is $\beta x(t)y(t)$. This is the rate that is expected if monomers and polymers mix homogeneously, for example by diffusing freely within the cell, and if the rate of binding to a polymer is independent to its length. There has been debate in the literature about whether monomers can bind at both ends of a polymer or only one; however, β simply reflects the overall rate of binding, however this occurs.
3. **Splitting of a polymer**. A linear polymer of length i has $i - 1$ bonds and the breakage rate per bond is assumed to be $\gamma \text{ min}^{-1}$. There appears to be a minimum polymer length, n , that is needed for stability and it is assumed that if polymer splitting leaves a fragment of length less than n , then this disintegrates instantaneously into (non-prion) monomers. Hence, $y_i(t) = 0$ for $i < n$. The value of n is not known precisely, but is usually taken to be in the range 2 – 6.
4. **Degradation ('death') of a monomer**, which occurs at rate $\lambda_{mon} \text{ min}^{-1}$ per monomer.
5. **Degradation ('death') of a polymer**, which is assumed to occur at rate $\lambda_{pol} \text{ min}^{-1}$ per polymer, independently of polymer length. Since polymers are assumed to be more stable than monomers, one expects $\lambda_{pol} \ll \lambda_{mon}$.

The evolution of the system over time is described by a system of differential equations:

$$(3) \quad \frac{dx}{dt} = \alpha - \lambda_{mon}x - \beta xy + 2\gamma \sum_{i=1}^{n-1} \sum_{j=i+1}^{\infty} iy_j$$

$$(4) \quad \frac{dy_i}{dt} = \beta x(y_{i-1} - y_i) - \lambda_{pol}y_i - \gamma(i-1)y_i + 2\gamma \sum_{j=i+1}^{\infty} y_j \quad (i \geq n)$$

Summing these equations leads to a closed system of equations for x, y and z ,

$$\begin{aligned} \frac{dx}{dt} &= \alpha - \lambda_{mon}x - \beta xy + n(n-1)\gamma y \\ \frac{dy}{dt} &= \gamma z - \lambda_{pol}y - (2n-1)\gamma y \\ \frac{dz}{dt} &= \beta xy - \lambda_{pol}z - n(n-1)\gamma y \end{aligned}$$

Setting the right hand sides of these equations to zero reveals that there are two steady-state solutions. When the system is started from any other state, it converges to one or other of these steady states as $t \rightarrow \infty$, depending on whether a particular function of the rate parameters is above or below a threshold value, similarly to many epidemic models (Greer *et al.*, 2006). One steady state is $\{x(t) = \alpha/\lambda, y(t) = 0, z(t) = 0\}$, which corresponds to the prion-free state; in the other steady state there are prion polymers as well as (non-prion) monomers.

In fact, in common with a number of other recent papers, Greer *et al.* (2006) consider a slightly different form of the model in which polymer length is treated as a continuous rather than a discrete variable. This simplifies the mathematical analysis of the model. In particular, the infinite set of differential equations (4) is replaced by a single partial integro-differential equation and Greer *et al.* (2006) give an explicit expression for the steady-state distribution of the number of polymers of different lengths when the steady state includes prions.

In their model, the mean polymer length in equilibrium is given by $\mu = \lambda_{pol}/\gamma + 2n_0$. If we let $U(> n_0)$ denote polymer length, then it can be shown from their results that the variable $V = (U - n_0)/(\mu - n_0) + 1$ has the (fixed) probability density function

$$f_V(v) = (v^2 - 1)e^{(v^2-1)/2} \quad v \geq 1.$$

Thus, the steady-state distribution of polymer length is determined entirely by its mean; the distribution has coefficient of variation 0.5580 and skewness 0.7589.

Similar models have been considered for yeast by Tanaka *et al.* (2006), Derdowski *et al.* (2010) and Palmer *et al.* (2011), all of whom also consider stochastic formulations of the model. Yeast models have to take account of the fact that the cells are growing, which affects the reaction rates. Palmer *et al.* (2011) consider a stochastic version of the model based on the five types of reaction listed above, with random variables X , Y and Z in place of x , y and z . The cell volume is modelled as $V(t) = V(0)\exp(Rt)$, where R is the relative cell growth rate. Following Wilkinson (2006, pp. 156–157), the rate parameters α and β are replaced by time-dependent rates $\alpha_g(t) = \alpha A_1 V(t)$ and $\beta_g(t) = \beta/[A_1 V(t)]$, respectively, where $A_1 = 6.0221 \times 10^{-17}$ is the number of molecules in $1\mu\text{M}$. The increasing rate $\alpha_g(t)$ ensures that the cell maintains the same total *concentration* of protein molecules as it enlarges. Similarly, the rate $\beta_g(t)$ decreases to reflect the fact that, for a given number of monomers and polymers, molecular collisions occur less frequently in a larger cell than a smaller cell. Ideally, the size-dependence of the rates would be verified experimentally, though this has not been possible to date.

Starting from the Kolmogorov forward equations, Palmer (2009, Section 3.2) derives a partial differential equation for the joint probability generating function of (X, Y_n, Y_{n+1}, \dots) . Although this is rather unwieldy, if one makes the standard simplifying assumption that $E(XY_i) = E(X)E(Y_i)$, a set of three differential equations for $E(X)$, $E(Y)$ and $E(Z)$ can be derived that are identical to the equations for x, y and z , except that the constants α and β are replaced by $\alpha_g(t)$ and $\beta_g(t)$, respectively. If degradation is ignored (i.e. $\lambda_{mon} = \lambda_{pol} = 0$), the threshold condition for the stable existence of the $[PSI^+]$ state is

$$(5) \quad \alpha\beta\gamma > R(R + n\gamma) [R + (n - 1)\gamma]$$

and the limiting mean polymer length as $t \rightarrow \infty$ is approximately $R/\gamma + 2n - 1$. The lack of dependence of mean polymer length on the monomer-to-polymer binding rate β is perhaps surprising at first sight, but is due to the fact that γ is the breakage rate per polymer *bond*, not per polymer. Thus, with larger β , monomers will bind to polymers more frequently, but polymers will also split more frequently.

Extension to asymmetrically dividing cells

The results discussed so far describe what occurs in a single cell that grows indefinitely. With minor modification, they also describe what occurs in a population of dividing cells provided that the cells divide symmetrically and that the population of monomers and polymers is shared equally between the two offspring cells. However, as already discussed, yeast cells divide asymmetrically, with daughter cells being smaller and taking longer to divide subsequently than mother cells. Additionally, there is experimental evidence (e.g. Sindi & Serio, 2009) that it is more difficult for long polymers to be transmitted to the daughter cell than short polymers.

Suppose that generation times for mother and daughter cells are fixed with values t_M and $t_D = t_M + t_d$, where t_d is the additional time needed for a daughter cell to reach maturity. Let $V_M(t)$ denote the volume of a mother cell measured at time t since the start of its cell cycle and $V_D(t)$ denote the corresponding volume of a daughter cell. Palmer *et al.* (2011) assume that growth of the mother cell is purely due to the growth of the bud, so that

$$V_M(t_M) = V_M(0) + V_D(0).$$

Also, at the time the bud separates from the mother, a proportion θ of the total volume accrues to the daughter, i.e. $V_D(0) = \theta V_M(t_M)$ and $V_M(0) = (1 - \theta)V_M(t_M)$.

The probability of transmission of a polymer from mother cell to daughter cell is modelled as a decreasing logistic function of the length of the polymer, w ,

$$(6) \quad \xi(w) = \frac{\theta}{1 + \exp[\rho_1(w - \rho_2)]},$$

where ρ_1 determines the rate of decrease and ρ_2 is the length at which the probability of transmission is reduced to $\theta/2$. The parameters are chosen so that there is little effect of polymer length for the range of polymer lengths that typically occurs under natural conditions, though this function becomes important when modelling experiments involving GdnHCl. Independently, Dewdowski *et al.* (2010) proposed a similar model, but assuming an abrupt transition from transmission to non-transmission, equivalent to taking the limit $\rho_1 \rightarrow \infty$.

Palmer *et al.* (2011) show that in specific examples, approximations to the mean number of monomers, the mean number of polymers and the mean polymer length still apply after being adjusted to take account of the values of θ , t_M and t_D . Simulation was used to incorporate variable generation times. This involved simulating the five basic reaction types, using the algorithm of Gibson and Bruck (2000), a variant of the standard Gillespie algorithm for simulating biochemical reactions that reduces the quantity of random numbers that need to be generated.

Application to modelling the results of experiments involving GdnHCl

The basic assumption of earlier models was that the addition of GdnHCl to the culture medium prevented replication of the prion. In terms of the nucleated polymerization model, this entails setting the parameter γ to zero. In fact, when other parameters are held fixed, the inequality that must be satisfied for existence of a stable $[PSI^+]$ population (equation 5) is quadratic in γ and a population that is initially $[PSI^+]$ will revert to $[psi^-]$ if γ is either below a lower threshold or above an upper threshold. We consider the upper threshold later, but the existence of a lower threshold that is larger than zero implies that it is not necessary for GdnHCl to block prion replication completely to cause curing, provided that γ is reduced to a value below the threshold.

A decrease in γ results in an increase in mean polymer length, as observed experimentally. This makes it increasingly difficult for polymers to be transmitted to the daughter cell, according to equation (6), also in qualitative agreement with experimental data (e.g. Cox *et al.*, 2003). In contrast to the model of Cole *et al.* (2004), where the expected proportion of polymers that is transmitted to the daughter cell, π , is assumed to be constant, the current model predicts that this proportion should decrease with time, as polymers become longer; this should be testable using the polymer counting technique of Cox *et al.* (2003) described earlier.

Transferring cells back to GdnHCl-free medium halts curing. This is consistent with the value of γ being restored to its original value. Similarly, reintroducing cells into medium containing GdnHCl reactivates curing at a rate that depends on the duration of the GdnHCl-free period (Cox *et al.*, 2003). Palmer *et al.* (2011) present results of simulations with changing values of γ that mimic these experimental findings.

The role of chaperones

Molecular chaperones are proteins that ‘look after’ other proteins in various ways, for example by assisting in protein folding. In yeast, the mechanism of $[PSI^+]$ polymer splitting involves a hexameric (6-molecule) ring of the chaperone Hsp104, though other chaperones, including Hsp40 and Hsp70, are also involved in ways that are not yet fully understood. Thus, in reality, the rate γ , which we have

regarded hitherto as a fixed parameter, is in reality dependent on the concentration of Hsp104 in the cell. It is known that GdnHCl binds to Hsp104, effectively inactivating the GdnHCl and leading to the reduction in γ . Derdowski *et al.* (2010) extend the standard nucleated polymerization model to include hexamers of Hsp104. In their model, the overall fragmentation rate, which is $\gamma(Z - Y)$ in the standard model, becomes $\nu H(Z - Y)/(H/2 + Z - Y)$, where H is the number of Hsp104 hexamers in the cell.

Interestingly, if yeast cells are modified genetically to over-express Hsp104, this also results in $[PSI^+]$ curing and it is tempting to speculate that this is because the upper threshold value of γ for stable existence of the $[PSI^+]$ phenotype has been breached, so that fragmentation occurs at such a high rate that a polymer population cannot be maintained. However, the upper threshold value turns out to be unrealistically large and the resulting curing curves do not match experimental data. Therefore, the explanation of curing by Hsp104 over-expression appears to lie elsewhere.

Current and future work

Refining the model

A key scientific challenge is to disentangle the roles played by the different chaperones (Hsp40, Hsp70, Hsp104) in determining the rate of polymer fragmentation which, as we have seen, is crucial to prion kinetics and to maintenance of the $[PSI^+]$ state. A combined approach involving experimentation and modelling seems a promising way forward. Genetic manipulations can give cells with reduced or elevated chaperone levels and we are developing experimental techniques that measure these levels. On the modelling side, one of us (VG) has developed a C++ template library, called `YeastSim++`, which provides simulation code for different kinetic schemes in a largely automated way, enabling different possible modes of interaction between chaperones to be explored quickly and easily.

Other possible developments include the following. Firstly, our current models distinguish between mother cells and new daughter cells, but Derdowski *et al.* (2010) have suggested that it may be important to take into account the *replicative age* of the mother, i.e. the number of daughter cells that it has produced previously. Secondly, the models could be refined to take account of the fact that protein production tends to occur in bursts rather than in the smooth fashion implied by differential equation models; Paulsson (2005) reviews models in this area. Finally, novel imaging techniques are enabling us to learn more about proteins within cells, for example to see whether they are localised in particular cellular compartments and to track their movements. Early results with Sup35p constructs suggest that movements may be more complex than simple random walks (Tsuji *et al.*, 2011).

Statistical inference

Little has been said in this paper about statistical inference, but of course it is important to be able to infer parameter values from experimental data. Morgan *et al.* (2003) and Cole *et al.* (2004) used likelihood based methods, with explicit computation of the likelihood. This approach is unlikely to be viable for more complex models based on intracellular kinetic and simulation-based methods, such as Approximate Bayesian Computation, seem more promising (e.g. Wilkinson, 2011). Palmer (2009, Chapter 4) considered a modification of the model of Cole *et al.* (2004) in which the parameter π was treated as a decreasing function of time, in keeping with the predictions of the intracellular model. However, the function was chosen empirically to introduce just one additional parameter into the model. Simulation-based inference was attempted, but the broad conclusion was that there is insufficient information in the curing data alone to allow the additional parameter to be estimated.

Thus, it appears that other sources of data will be needed, in conjunction with curing data, to estimate parameters reliably, with associated issues of how best to combine data from different

sources (Henderson *et al.*, 2010). Currently, we are trying to obtain experimental data on polymer-length distributions.

Faster simulation

Simulation of biochemical processes within cells is expensive computationally because large numbers of events occur in short time periods. We are exploring two approaches to faster simulation, one based in hardware and the other algorithmic. Although neither approach is novel, they need to be adapted to be effective in simulating polymer growth and fragmentation.

The first approach uses graphics processing units (GPUs) to provide a cheap source of highly parallel processing (Dematté & Prandi, 2010). A challenge here is to design algorithms that exploit the combined computing power of the processors effectively whilst controlling the overheads of moving data between processors. The second approach uses an approximate rather than an exact approach to simulation. Exact simulation typically requires all hazard rates to be updated after each event occurs. An alternative approach, known as *tau-leaping*, updates rates only after a time interval of length τ . The challenge of this approach is to choose τ large enough to obtain a worthwhile improvement in computational speed but not so large as to entail a substantial loss of quality in the simulated data. This has been a very active area of research in recent years (e.g. Hu & Li, 2009).

Conclusions

We hope to have shown how stochastic modelling in this area has evolved as understanding of yeast prion biology has increased. For statisticians, this type of project requires investment of time and effort in becoming familiar with some of the underlying biology and its associated terminology, but the payoff is a chance to work at the cutting edge of a fast-moving and topical area of bioscience that presents challenging problems of both mathematical and statistical interest.

Acknowledgements

We thank Lee Byrne, Diana Cole, Brian Cox and Karen Palmer for their contributions to the work described here. We thank the Biotechnology and Biological Sciences Research Council (BBSRC) for twice funding matching postdoctoral appointments in Biosciences and in Statistics, and the Engineering and Physical Sciences Research Council (EPSRC) for funding the PhD work of Karen Palmer.

REFERENCES (RÉFÉRENCES)

- Byrne L.J., Cox B.S., Cole D.J., Ridout M.S., Morgan B.J.T. & Tuite M.F. (2007) Cell division is essential for elimination of the yeast $[PSI^+]$ prion by guanidine hydrochloride. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 11688–11693.
- Cole D.J., Morgan B.J.T., Ridout M.S., Byrne L.J. & Tuite M.F. (2004) Estimating the number of prions in yeast cells. *Mathematical Medicine and Biology*, **21**, 369–395.
- Cole D.J., Ridout M.S., Morgan B.J.T., Byrne L.J. & Tuite M.F. (2007) Approximations for expected generation number. *Biometrics*, **63**, 1023–1030.
- Cox, B.S., Ness, F. & Tuite, M.F. (2003) Analysis of the generation and segregation of propagons: entities that propagate the $[PSI^+]$ prion in yeast. *Genetics* **165**, 23–33.
- Dematté, L. & Prandi, D. (2010) GPU computing for systems biology. *Briefings in Bioinformatics*, **11**, 323–333.
- Derdowski, A. Sindi, S.S., Klaips, C.L., DiSalvo, S. & Serio, T.R. (2010) A size threshold limits prion transmission and establishes phenotypic diversity. *Science*, **330**, 680–683.
- Gibson, M.A. & Bruck, J. (2000) Efficient exact stochastic simulation of chemical systems with many

species and many channels. *Journal of Physical Chemistry*, **104**, 1876–1889.

Green, P.J. (1981). Modelling yeast cell growth using stochastic branching processes. *Journal of Applied Probability*, **18**, 799–808.

Greer, M.L., Pujo-Menjouet, L. & Webb, G.F. (2006) A mathematical analysis of the dynamics of prion proliferation. *Journal of Theoretical Biology*, **242**, 598–606.

Henderson, D.A., Boys, R.J. & Wilkinson, D.J. (2010). Bayesian calibration of a stochastic kinetic computer model using multiple data sources. *Biometrics*, **66**, 249–256.

Hu, Y. & Li, T. (2009) Highly accurate tau-leaping methods with random corrections. *Journal of Chemical Physics*, **130**, 124109.

Masel, J., Jansen, V.A.A. & Nowak, M.A. (1999) Quantifying the kinetic parameters of prion replication. *Biophysical Chemistry*, **77**, 139–152.

Morgan B.J.T., Ridout M.S. & Ruddock L.W. (2003) Models for yeast prions. *Biometrics*, **59**, 562–569.

Palmer, K.J., Ridout, M.S. & Morgan, B.J.T. (2008) Modelling cell generation times using the tempered stable distribution. *Applied Statistics*, **57**, 379–397.

Palmer, K.J. (2009) *Stochastic Models for Cell Biology*. PhD Thesis, University of Kent.

Palmer, K.J., Ridout, M.S. & Morgan, B.J.T. (2011) Kinetic models of guanidine hydrochloride-induced curing of the yeast [PSI⁺]prion. *Journal of Theoretical Biology*, **274**, 1–11.

Paulson, J. (2005) Models of stochastic gene expression. *Physics of Life Reviews*, **2**, 157–175.

Ridout M.S., Cole D.J., Morgan B.J.T., Byrne L.J. & Tuite M.F. (2006) New approximations to the Malthusian parameter. *Biometrics*, **62**, 1216–1223.

Ridout, M.S. (2008) Computational methods for yeast prion curing curves. *Mathematical Biosciences*, **215**, 152–157.

Sindi, S.S. & Serio, T.R. (2009) Prion dynamics and the quest for the genetic determinant in protein-only inheritance. *Current Opinion in Microbiology*, **12**, 623–30.

Tanaka, M., Collins, S.R., Toyama, B.H., & Weissman, J.S. (2006) The physical basis of how prion conformations determine strain phenotypes. *Nature*, **442**, 585–589.

Tsuji T., Kawai-Noma S., Pack C.G., Terajima H., Yajima J., Nishizaka T., Kinjo M. & Taguchi H. (2011) Single-particle tracking of quantum dot-conjugated prion proteins inside yeast cells. *Biochemical & Biophysical Research Communications*, **405**, 638–643.

Tuite, M.F. & Serio, T.R. (2010) The prion hypothesis: from biological anomaly to basic regulatory mechanism. *Molecular Cell Biology*, **11**, 823–833.

Wilkinson, D.J. (2006) *Stochastic Modelling for Systems Biology*. Chapman & Hall/CRC, London, UK.

Wilkinson, D.J. (2011) Parameter inference for stochastic kinetic models of bacterial gene regulation: A Bayesian approach to systems biology (with discussion). In *Bayesian Statistics 9*, J.M. Bernardo *et al.* (eds). Oxford University Press, Oxford, UK, in press.