Bayesian inference elucidates activation of competing mechanisms that control DNA double strand break repair.

M. Woods[1] and C.P. Barnes[1,*]

[1] Department of Cell and Developmental Biology, University College London, England

*To whom correspondence should be addressed. Email: christopher.barnes@ucl.ac.uk

Abstract

Double strand breaks (DBSs) promote different repair pathways involving DNA end joining or homologous recombination, yet their relative contributions, interplay and regulatory interactions remain to be fully elucidated. These mechanisms give rise to different mutational processes and their propensity for activation directly affects genomic instability with implications across health and evolution. Here we present a new method to model the activation of at least three alternatives: non-homologous end joining (fast), homologous recombination (slow) and alternative end joining (intermediate) repair. We obtain predictions by employing Bayesian statistics to fit existing biological data to our model and gain insights into the dynamical processes underlying these repair pathways. Our results suggest that data on the repair of breaks using pulse field gel electrophoresis in wild type and mutants confirm at least three disjoint modes of repair. A density weighted integral is proposed as a tool to sum the predicted number of breaks processed by each mechanism from which we quantify the proportions of DSBs repaired by each. Further analysis suggests that the ratio between slow and intermediate repair depends on the presence or absence of DNAPKcs and Ku70. We outline how all these predictions can be directly tested using imaging and sequencing techniques. Most importantly of all, our approach is the first step towards providing a unifying theoretical framework for the dynamics of DNA repair processes.

1 Introduction

Double strand breaks and genetic mutations Double strand breaks (DSBs) are cytotoxic lesions in DNA that occur naturally by oxidative stress, DNA replication and exogenous sources [1,2]. When left unprocessed or during erroneous repair, DSBs cause changes to DNA structure creating mutations and potential genomic instability [3–8]. This is apparent with an increase in chromosomal aberrations observed in cells compromised of DSB end-joining by loss of Ku80, suggesting a caretaker gene role for regulating components [3]. Mutations that have been associated with DSBs include chromosome translocations [4, 5], small deletions or insertions [6, 7] and recombination leading to loss of heterozygosity [8]. Their variant classification depends on the repair mechanism, of which there have been multiple alternatives proposed in the literature including non homologous end joining (NHEJ) [7, 9–17], homologous recombination [18] including single strand annealing (SSA) [19, 20], microhomology mediated end joining (MMEJ) [21, 22] and alternative or back-up end joining (A-EJ) [23, 24]. The type of DNA lesion defined by the complexity of the break point affects the type of repair mechanism that is initiated, where simple breaks caused by restriction enzymes can be distinguished by those caused by ionising radiation (IR), for reviews see [25,26]. This leads to the suggestion that the type of break affects the probability of error prone repair, because particular mutations have been linked to specific repair mechanisms. For example, in mouse, chromosome translocations are promoted when two simple DSBs positioned on different chromosomes are repaired by SSA [4] and in *Saccharomyces cerevisiae*, NHEJ of simple DSBs is associated with small deletions or insertions [7]. Recently, mutations specific to alternative mechanisms have been identified, where next generation sequencing has revealed sequence specific chromosome translocations following A-EJ at dysfunctional telomeres [5]. In addition to the causal relationship between repair mechanism and mutation, *in vivo* studies of DSBs have shown that repair mechanism activation is cell cycle dependent and therefore dynamic. This has been proposed in human cells, where imaging of live single cells has suggested that the decision to activate a particular repair mechanism is not fixed at the time of damage and cells exhibit a pulse like repair [27]. This dynamic activation is supported by a molecular basis for cell cycle dependence in NHEJ, mediated by Xlf1 phosphorylation [28].

It is known that DSBs are associated with mutations in the genome. The type of mutation depends on the mechanism of repair which is in turn affected by cell cycle stage and complexity of the DSB. Some cancers are deficient in at least one repair mechanism and in these cases, alternative mechanisms of repair have been observed to compensate. One mechanism that has been shown to play a role is A-EJ, where Polθ has been shown to be a necessary regulator for cell survival in homologous recombination deficient cancer [5, 29]. Here, we suggest that mathematical modelling can be used to understand the interplay between multiple repair mechanisms. A better understanding of the interplay between DSB repair mechanisms could be applied to design potential lethal synthetic therapeutics in cancer [30]. Our model has been developed with the aim to provide predictions that can be used to test the model and further our understanding of the system as a whole.

DSB repair mechanisms We model three mechanisms that repair DSBs caused by ionising radiation (IR): Fast, slow and intermediate repair. We propose that these could describe NHEJ, SSA and A-EJ respectively, however other mechanisms could be included. A description of Rad51 dependent homologous recombination should not be included in our model predictions because mutants defective in Rad51 are not known to contribute to the repair of IR-induced DSBs [31]. NHEJ requires little or no homology and is a mechanism of DNA end joining in both unicellular and multicellular organisms [7]. In vertebrates, NHEJ initiates the recruitment and binding of several proteins (see Figure 1a)). These have been shown to include Ku70, Ku80, DNAPKcs, Artemis and Ligase IV in a cell free system [9]. Ku70 and Ku80 are subunits of the protein DNA-PK. Biochemical and genetic data suggests they bind to DNA ends and stimulate the assembly of NHEJ proteins by DNAPKcs [10, 12]. The local availability of DNA-PKcs leads to a fast rejoining process [15] and repair proceeds by Artemis facilitated overhang processing and end ligation via DNA Ligase IV [13, 14]. In mammalian cells, NHEJ has been suggested to repair the majority of DSBs caused by IR [16]. Ku deficient cells do not produce NHEJ products due to excessive degradation or inhibition of end joining [11]. However, the binding

of DNA-PKcs is thought to determine repair by NHEJ, as it precludes DNA end resection, a process that initiates homologous recombination [32]. Although well studied, new regulating components of NHEJ are still being discovered, for example the protein PAXX [17].

SSA is slower than NHEJ, is defined for both adjacent and intermolecular homologous sequences and was first described in mouse fibroblast cells by Lin *et al.* [19, 20]. Lesions are removed by aligning two complementary sequences on ssDNA with 3′ ends which are exposed through a 5′ to 3′ exonuclease end resection. Remaining overhangs are cut by an endonuclease and the DNA is reconstructed by DNA polymerase using the homologous sequences as a template. Some of the components that contribute to SSA have been identified in eukaryotes e.g. the complex MRN consisting of Mre11, Rad50 and Nibrin which facilitates DNA end resection [33]. Following resection, replication protein A (RPA) binds to the DNA and when phosphorylated, forms a complex with Rad52 [34], where Rad52 stimulates DNA annealing between the two complementary sequences [35]. Similarly to NHEJ, following gap repair, SSA is terminated with end ligation by Ligase III [36] (Figure 1a)). In yeast, it has been suggested that SSA constitutes a major role in the repair of DSBs accounting for three to four times more repairs than gene conversion during M phase [37]. Interestingly, inhibition of DNAPKcs but not Ku leads to elevated levels of resection and more HR [32]. Data of repair kinetics for mutants defective in Rad52 show limited slow repair in comparison to wild type repair curves in gamma irradiated cells in chicken B line cells [38], suggesting that SSA may be active in the repair of DSBs caused by IR.

One interesting finding in genetic studies is that when NHEJ is compromised, DSBs are removed by an alternative mechanism that we refer to as A-EJ [23, 39], for reviews see [40, 41]. It is still unclear how A-EJ is regulated or interacts with other processes and the mechanism has adopted various names in the literature, such as MMEJ in yeast [42] and back-up NHEJ (B-NHEJ) in higher eukaryotes [41]. Thought to act on break points with ends that are not complementary, biochemical studies have confirmed rejoining of breaks in the absence of NHEJ factors [39]. This mechanism is error prone, giving rise to chromosome translocations, of which there are more when NHEJ is inactive,

suggesting it's role as a back up mechanism in eukaryotes [43]. Complementary studies in yeast have suggested that A-EJ is repressed by RPA which promotes error-free homologous recombination by preventing spontaneous annealing between micro homologies which can lead to MMEJ [21]. An assortment of proteins have been proposed to regulate A-EJ, namely PARP-1, 53BP1, Lig3 and 1, Mre11, CtIP and Polθ (see Figure 1a)). PARP-1 is required and competes with Ku for binding to DNA ends through the PARP-1 DNA binding domain [24]. Inhibition of DNA-PKcs does not activate repair by PARP-1 mediated A-EJ, suggesting that the presence of Ku alone is sufficient to down regulate activation of backup NHEJ [24]. Similar roles have been suggested for the protein 53BP1, where the activation of 53BP1 in MMEJ is dependent on Ku70 and independent of DNAPKcs [22]. Additionally, CtIP has been associated with A-EJ through the use of microhomology [44] and the proteins required for end joining have been identified as Lig3 and Lig1 in the absence of XRCC1 [43, 45, 46]. This pathway has never been observed in single cells and it is unclear how A-EJ is related to other mechanisms. However, targeted RNAi screening for A-EJ has uncovered shared DNA damage response factors with homologous recombination [47].

Mathematical models of DSB repair Previous mathematical models of DSB repair have applied biphasic [48], biochemical kinetic [49–52], multi-scale [53, 54], and stochastic methods [55]. Biochemical kinetic methods have modelled NHEJ and SSA to predict repair time series and transient levels of recruitment proteins. In a study by Cucinotta *et al.* [49], a set of coupled nonlinear ordinary differential equations were developed. Based on the law of mass action, the model assumed a stepwise irreversible binding of repair proteins to describe NHEJ rejoining kinetics and the phosphorylation of H2AX by DNA-PKcs. A similar model was later applied to predict repair kinetics during SSA [50, 52] and in NHEJ to predict repair kinetics and DNA-PKcs recruitment for cells irradiated between 20-80Gy [51]. This approach was developed to model NHEJ, HR, SSA and two alternative pathways, providing a mechanistic description of all major pathways under a wide range of LET values and heavy ions [56]. Stochastic discrete simulation of reaction systems has been used to model fragment

rejoining during NHEJ. In a study by Li *et al.* [55], it was suggested that there are significant differences between the rejoining times for DSB fragments of different lengths and that NHEJ is dependent on the space required for the recruitment repair proteins to bind. Additional studies have used multi-scale modelling to describe protein recruitment and the spatial aspects of a DSB distribution for different types of radiation, using the software PARTRAC [53, 54]. Much of the mathematical modelling has focused on biochemical kinetic models that can be used to reproduce the experimental data observed. In these modelling approaches, large numbers of parameters are required to describe sequential steps in the repair process. This can cause difficulty in identifying parameter values because multiple parameter value combinations may be able to describe the data well. As such, predictions that are made are not unique, which could be detrimental in the design of a biological experiment. Consequently, designing a model which provides a unique prediction of repair dynamics is a challenge in current biology. One model that has addressed this in part has been applied to single strand break repair, where a model of random recruitment and reassembly of repair complexes during nucleotide excision repair has suggested that the dynamics are defined by slow first order kinetics. This slow repair is explained by the assembly of a repair complex from rapidly exchanging components [57].

In this study we have developed a new stochastic model to predict the interplay between multiple mechanisms using existing biological data. We simultaneously fit published datasets of DNA repair in wild type and mutants defective in proteins required for SSA, NHEJ and A-EJ using approximate Bayesian computation sequential Monte Carlo (ABC SMC) [58–61], which is one method that can be used to fit a model to multiple datasets when the likelihood is not available, and has previously been applied to estimate parameter values in a model of DNA methyllation [62]. Our approach strikes a balance between a detailed mechanistic description of the biochemical components with a traditional statistical model. This enables insights into the dynamical process underling repair pathways combined with novel and testable predictions.

2 Materials and Methods

A model of fast, slow and intermediate repair We assume repair can be described by at least three different mechanisms. A fast mechanism corresponding to NHEJ, a slow mechanism corresponding to homology based repair of SSA and an intermediate mechanism A-EJ see Figure 1 b). A-EJ is taken to be ten fold less active than NHEJ but we impose no restriction on it's dynamic behaviour by allowing the activation to change between the datasets. DNA repair is modelled with a stochastic reaction system consisting of six reactions on a population of DSBs $\bar{x}$. Each DSB in the population can be in one of four states, x_i, $i \in \{1,2,3,4\}$. All DSBs are initially in state x_1 and we define the set $R := (x_q : q \in \{2,3,4\})$ to represent DSBs that are being processed by fast repair (x_2), slow repair (x_3) and intermediate repair (x_4), Figure 1 b). The reactions assume irreversible binding of repair proteins:

$$\begin{aligned} x_1 + \mathrm{E}_{q-1} &\overset{K_{q-1}}{\rightarrow} x_q, q \in \{2,3,4\} \\ x_q &\overset{K_{q-1}}{\rightarrow} \emptyset. \end{aligned} \tag{1}$$

E_{q-1}, $q \in \{2,3,4\}$ represents repair proteins Ku, MRN and PARP-1 respectively. K_{q-1} represents the rate of binding for the initial protein recruitment and end ligation, see Figure 1 b). Following the approach of Cucinotta *et al.*, [49], we assume that the total amount of protein is conserved for each repair mechanism:

$$C = [E_{q-1}] + x_q. \tag{2}$$

This models the assumption that the sum of all bound and free protein does not change over time and in the deterministic system results in a nonlinear coupled ordinary differential equation (see supplementary material). We are interested in modelling live single cell DNA repair and because of intrinsic variation between cells we assume a stochastic model. To incorporate random recruitment and intrinsic stochasticity, we adopt a molecular approach to kinetics. In this method, binding is not deterministic and reactions depend on the probability that a DSB and a repair protein will be within a reacting distance. This is implemented in our code by formulating Kolmogorov's forward equation

for the stochastic petri network and simulating with the Gillespie algorithm [63]. The full set of reactions, prior distributions on parameters and initial conditions are presented in the supplementary appendix. At any time t the total observed DSBs $x(t)$ are given by the sum of all states for all DSBs in the population.

$$x(t) = \sum_{j=1}^{4} \sum_{i=1}^{N_j} x_j^i(t). \tag{3}$$

where N_j is the number of DSBs in state j. Proportions of DSBs repaired by each mechanism are estimated by calculating the cumulative number of DSBs $G_j(t)$ $\forall j \in R$ that enter each individual pathway, (see supplementary material).

Experimental data To infer parameter values we apply ABC SMC. With this method we attempt to construct the system that is most likely to give rise to the experimental data that we observe. The experimental data used in this study are published repair curves that are generated from methods of pulse field gel electrophoresis, a technique that distributes the DNA according to the length of the fragment. We model the dose equivalent number of DSBs that are obtained from the fraction of DNA released into the gel [64]. Table 1 lists the experimental data that are used for inference. Cells were γ-irradiated in a Cs^{137} chamber [65] or exposed to X-rays [15] and the number of DSBs within the population recorded over time.

Parameter estimation and approximate Bayesian computation We aim to build a model that can be used to obtain unique predictions, so it is advantageous to minimise the number of parameters that describe the system. To do this, we develop a hierarchical model [66], where the parameter values K_j, $j \in \{1,2,3\}$ are log normally distributed and share a common mean μ_j across all the datasets in which they are included, see Figure 2 a). By drawing parameters from one common hyper parameter μ_j across the datasets, the total number of parameters that are required to describe the data is reduced. For datasets in which a repair protein is repressed downstream of the initial protein that binds, we impose an additional hyper parameter μ_4. We include this additional hyper parameter because it is not clear if a repair mechanism remains active when individual regulating proteins are

repressed. Altogether, our model contains five hyper parameters to model eight independent datasets and each of the model parameters K_i are drawn from these four parameters accordingly (see table 2 and supplementary material). The fifth hyper parameter is the variance σ, which is shared amongst all parameters and the data. To assign values to our parameters, we perform ABC SMC to calculate the target posterior density $\pi(\bar{\mu}|\bar{D})$. This is the most probable set of parameters that can describe our data $\bar{D} =$ D1-D8. For further details on the hierarchical model and ABC SMC see supplementary material.

Data set	**Dose (Gy)**	**Phase**	**Cell line, Mutant**	**Repair Mechanisms**
D1	20	Asynchronous	WT MEFs* [31]	NHEJ, SSA, A-EJ
D2	20	G1	DNA-PKcs$^{-/-}$ MEFs* [31]	SSA, A-EJ
D3	20	G2	DNA-PKcs$^{-/-}$ MEFs* [31]	SSA, A-EJ
D4	80	Asynchronous	Rad52$^{-/-}$ DT40* [67]	NHEJ, A-EJ
D5	80	Asynchronous	Ku70$^{-/-}$/Rad54$^{-/-}$ DT40* [38]	SSA, A-EJ
D6	54	Asynchronous	Ku70$^{-/-}$ DT40* [38]	SSA, A-EJ
D7	52	Asynchronous	Ku70$^{-/-}$ + DPQ MEFs [24]	SSA
D8	32	Asynchronous	WT + 3'-AB MEFs [24]	NHEJ, SSA

Table 1: Table of data sets used for model fitting. The data contains DSB repair kinetics for cells that are irradiated at different doses or split into different phases of the cell cycle, G1 and G2. Data was traced from current literature, or where indicated was provided by G. Iliakis (*). References to the data and cell lines are provided. We chose a combination of mouse embryonic fibroblasts (MEFs) and DT40 cells because DT40 cells remove DSBs from their genome similarly to mammalian cells [67].

3 Results

3.1 DSBs require fast, slow and alternative mechanisms

ABC SMC was performed on the experimental data with the model parameters presented in table 2, (for prior distributions see supplementary material). The posterior distributions of the hyper parameters are shown in Figure 2 b). Inspection of the interquartile range of the hyper parameters confirms that a combination of fast, slow and intermediate repair is sufficient to describe the wild type and mutant data (Figure 2c), Furthermore a two sided Kolmogorov Smirnov test between the posterior

distributions for the hyper parameters confirmed that the four distributions were significantly different to one another (μ_1, μ_2 $D = 0.994$, μ_1, μ_3 $D = 0.7$, μ_2, μ_3 $D = 0.706$, μ_2, μ_4 $D = 1$, μ_4, μ_3 $D = 1$, all tests $p < 2.2e^{-16}$). For each data set (D1-D8) the posterior ranges of the parameters K_1, K_2 and K_3 were recorded (supplementary Figure S1). The posterior for the wild type data is shown in Figure 2 d-f). Analysis of the marginal distribution shows that the parameter distributions of K_1, K_2 and K_3 deviate from the hyper parameter distributions, suggesting that although the mechanisms are defined as fast, slow and intermediate, there is variation in activation of the mechanisms among different mutants (Figure 2 d)). There is some overlap in parameter values K_1, K_2 and K_3 (Figure 2 e)) but the interquartile ranges of the parameters K_1, K_2 and K_3 are disjoint (Figure 1f)). This is also observed in all eight datasets (supplementary Figure S1). For all posterior distributions of the parameters, see supplementary material Figure S2. The fit of the simulation to the data for all eight data sets is shown in Figure 2 g). In summary, we predict that the biological data can be explained by one fast, one slow and at least one intermediate mechanism.

Data	**Model Parameters**			**Hyper Parameters**		
	Fast	**|Slow**	**|A-EJ**	**Fast**	**|Slow**	**|A-EJ**
Wild type	K_{1d1}	K_{2d1}	K_{3d1}	μ_1, σ	μ_2, σ	μ_3, σ
DNA-PKcs$^{-/-}, G1$	K_{1d2}	K_{2d2}	K_{3d2}	μ_4, σ	μ_2, σ	μ_3, σ
DNA-PKcs$^{-/-}, G2$	K_{1d3}	K_{2d3}	K_{3d3}	μ_4, σ	μ_2, σ	μ_3, σ
Rad52$^{-/-}$	K_{1d4}	K_{2d4}	K_{3d4}	μ_1, σ	μ_4, σ	μ_3, σ
Ku70$^{-/-}$/Rad54$^{-/-}$	-	K_{2d5}	K_{3d5}	-	μ_2, σ	μ_3, σ
Ku70$^{-/-}$	-	K_{2d6}	K_{3d6}	-	μ_2, σ	μ_3, σ
Ku70$^{-/-}$ + DPQ	-	K_{2d7}	-	-	μ_2, σ	-
WT + 3'-AB	K_{1d8}	K_{2d8}	-	μ_1, σ	μ_2, σ	-

Table 2: Model parameters with their corresponding hyper parameters used in our hierarchical model. Their values are predicted following ABC SMC. For prior distributions on the hyper parameters, see supplementary material.

3.2 Number of DSBs repaired by each mechanism depends on regulating recruitment proteins

Repair and the cumulative repair were plotted for each data set (see Figure 3 a,d)), data sets in which NHEJ is active exhibited a faster repair with the cumulative number of DSBs reaching within 95% of the total within 2 hours. Next, we plotted the number of DSBs entering each repair mechanism as a time series (Figure 3 b,e)). The simulated data predicts that fast repair consistently processes most of the DSBs within two hours after radiation (red curves in Figure 3). Similarly, there were no clear differences amongst the data in the DSB processing by slow repair. Intriguingly, intermediate repair was slower in cells compromised of Ku70 than those without DNAPKcs (green curves, Figure 3 b,e)) To calculate the predicted number of DSBs repaired by fast, slow and alternative mechanisms, we computed the density weighted integral $G_j(t)$. The results are shown in Figure 3 c,f). The model predicts that the fast mechanism repairs most DSBs in the presence or absence of slow and intermediate mechanisms. Datasets for which cells were deficient in regulating components of NHEJ confirmed variation in the numbers of DSBs repaired by intermediate mechanisms. In agreement with the results obtained from the time series plots (Figure 3 b,e) there was a difference in the ratio of slow and intermediate mechanisms between data sets D2,D3 and D5,D6. We also observed a difference in the number of DSBs repaired by A-EJ and slow repair between G1 and G2, corroborating with experimental results in the literature. Both mechanisms increased and decreased significantly respectively (two sample t-test, $p<0.01$). The time taken for half the DSBs to be repaired by intermediate repair is shown in Figure 4 a). The majority of repair is fast, occurring within two hours, however, for cells deficient in Ku70, A-EJ adopts a slower repair with half maximum achieved at eight hours. Finally, we looked at the activation of A-EJ across the datasets by comparing the posterior distributions. The posterior distributions for A-EJ are shown in Figure 4 b). Activation corresponding to the role of DNA binding and end ligation is lowest in the wild type data, suggesting that intermediate mechanisms may compensate when either slow or fast repair is inhibited. The rate is highest in G2 when DNAPKcs is inhibited. We suggest that A-EJ could adopt a slow or fast repair and that the speed of repair depends on the presence or absence of DNAPKcs and Ku70 because inhibition of Rad52 had little

effect on the time until half-maximum for A-EJ, although the rate was increased from the wild type data. There are a number of ways in which this difference between Ku70 and DNAPKcs mutants can be interpreted. The first is that when Ku70 is inhibited, then two alternative mechanisms are activated, one that is fast and one that is slow. The other alternative is that A-EJ is one repair mechanism that repairs at a slower rate when Ku70 is inhibited.

3.3 Activation of competing repair mechanisms

Inspection of the time series data shows that at time $t = 0.5$hours, the majority of DSBs that are being processed are being repaired by the fast mechanism. Figure 4 c) shows a typical distribution of DSBs being processed by each mechanism over time. At time $t = 0$, the cells are exposed to a single dose of ionising radiation. Quickly, for example at time $t < 1$hour a large proportion of DSBs are processed by fast repair and possibly faster alternative repair mechanisms such as A-EJ. Later, after all DSBs processed by the faster mechanisms have been repaired, the remaining DSBs are still being processed by slower mechanisms. This change in the activity of repair mechanisms could potentially be investigated by recording changes in the level of recruitment proteins or gene expression as time series. To quantify this change in our simulated data, we plotted the percentage DSBs out of the total DSBs that remain in active repair mechanisms over time for the wild type data (see Figure 4 d). At a time of $t = 0.5$ and $t = 8$hours, the interquartile range shows an overlap in the percentage. To confirm if the dynamics presented in Figure 4 c) are representative of the whole data set, we considered all time series for all parameters, a total of 9000 simulations. For each parameter at every time point, we assigned a value of 1 if the corresponding mechanism for the parameter contained over 30% of the total DSBs being processed at that time point and a value of 0 if it contained less than 30%. The results are shown in Figure 4 e). There is a clear trend showing that the percentage of total activation decreases in time with an increase in repair rate K. This predicts that if a cell experiences a sudden creation of DSBs, then gene expression for slower repair mechanisms will be maintained for longer than those required for faster repair mechanisms such as NHEJ, a result that has been shown for NHEJ and HR (Figure 3 in [27]).

4 Discussion

In this study, we present a new hierarchical model of DSB repair and apply ABC SMC to make predictions on the activation of at least three repair mechanisms. Our Bayesian approach suggests that fast, slow and intermediate repair are sufficient to describe the data observed. Because the model assumptions are simple and exclude the full mechanistic details of the biological processes, we have created an identifiable framework that has generated unique insights. To obtain these insights, we have analysed time series for fast, slow and intermediate repair by assuming that repair attributed to different mechanisms is implicit in the experimental PFGE data. Because our simulated data are constrained to the biological data through Bayesian computation, the statistical analysis performed on the simulations drawn from the posterior distribution provides an additional method to quantify biological datasets. In contrast to previous studies, we have designed our model so that our predictions can be directly reproduced by experimental techniques to further aid our understanding of the system. In this study we have identified four major insights, some of which may already be hypothesised but each of which can be tested experimentally. The first insight is that the data can be explained by three independent mechanisms, for example a mechanism faster than Rad52 dependent HR is required to fit the experimental data to the model in datasets D2 and D3 (knock of DNA-PKcs). Another interesting insight is that intermediate repair is increased in G2 phase of the cell cycle. If we assume that intermediate repair corresponds to back-up end joining, then this is in agreement with experimental results in the literature, supporting the existing biological evidence of the role of A-EJ in DSB repair [31]. By analysing simulated data generated from our model, we observe differences in the half time of repair in A-EJ, this leads us to a second prediction that the speed of repair of A-EJ depends on the presence of regulating components in NHEJ and SSA. This prediction could be verified by recording the repair of DSBs in single cells with and without inhibitors for the regulating components and recording protein recruitment using time-lapse microscopy. There are existing experimental systems that would enable this type of experiment, for example the fluorescently tagged 53BP1 [27], a protein that colocalises with alternative DSB markers and fluorescently tagged PARP1, a candidate protein for A-EJ [68]. The third insight is obtained by applying a density weighted integral to compute the total

DSBs repaired by each mechanism. With this information, we can estimate the proportion of mutations that are expected following DSB repair in wild type and mutant cells. Because many cancer cells are deficient in certain repair mechanisms [29], this information will be important in predicting the numbers and types of mutations that we expect to observe. The fourth insight is that the expression profile of different DSB repair mechanisms changes over time, with slower repair mechanisms still remaining active many hours after the initial dose of radiation. Pulse like behaviour has been recorded in the repair of DSBs in human cells [27] and we suggest that this prediction could be further investigated using microarrays or RNA sequencing, although currently the genes involved in the different pathways - and how much they are shared - remains to be fully elucidated.

With additional data it will be possible to extend the model and include additional terms such as explicit repressive cross-talk interactions. However, from our simple assumptions we have generated *in silico* data and used it to produce a number of unique insights that can be tested experimentally. Mathematical modelling not only facilitates the analysis of disparate datasets but also enforces the explicit formalisation of the underlying assumptions of our models. Our framework is a significant step towards a theoretical understanding of the dynamics DNA repair pathways. As the collection of larger and more complex datasets increases, we anticipate these approaches will be absolutely essential for the reverse engineering of these complex biological processes.

5 Funding

This work was supported by a Wellcome Trust Research Career Development Fellowship [WT097319/Z/11/Z].

6 Acknowledgements

We gratefully acknowledge the advice and guidance of Geraint Thomas, David Hall, Miriam Leon, Lourdes Sri-Raja, Tanel Ozdemir, Alex Fedorec, David Gonzales and all members of the Barnes lab. M. Berger at NVIDIA Corporation and Fabrice Ducluzeau at University College London for hardware support. The authors would like to acknowledge that the work presented here made use of the Emer-

ald High Performance Computing facility made available by the Centre for Innovation. The Centre is formed by the universities of Oxford, Southampton, Bristol, and University College London in partnership with the STFC Rutherford-Appleton Laboratory.

References

[1] M H O'Dea K Mizuuchi, L M Fisher and M Gellert. DNA gyrase action involves the introduction of transient double-strand breaks into DNA. *Proc Natl Acad Sci U S A*, 77(4):1847–1851, 1980.

[2] David Freifelder. Lethal changes in bacteriophage dna produced by x-rays. *Radiation Research Supplement*, 6:pp. 80–96, 1966.

[3] M J Difilippantonio, J Zhu, H T Chen, E Meffre, M C Nussenzweig, E E Max, T Ried, and A Nussenzweig. DNA repair protein Ku80 suppresses chromosomal aberrations and malignant transformation. *Nature*, 404(6777):510–514, mar 2000.

[4] Maria Jasin and Christine Richardson. Frequent chromosomal translocations induced by DNA double-strand breaks. *Nature*, 405(6787):697–700, June 2000.

[5] P. A. Mateos-Gomez, F. Gong, N. Nair, K. M. Miller, Eros Lazzerini-Denchi, and Agnel Sfeir. Mammalian polymerase θ promotes alternative nhej and suppresses recombination. *Nature*, 518(7538):254–257, 02 2015.

[6] N Sugawara and J E Haber. Characterization of double-strand break-induced recombination: homology requirements and single-stranded dna formation. *Molecular and Cellular Biology*, 12(2):563–575, 1992.

[7] J K Moore and J E Haber. Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in saccharomyces cerevisiae. *Molecular and Cellular Biology*, 16(5):2164–73, 1996.

[8] Mary Ellen Moynahan and Maria Jasin. Loss of heterozygosity induced by a chromosomal double strand break. *Proceedings of the National Academy of Sciences*, 94(17):8988–8993, 1997.

[9] Joe Budman and Gilbert Chu. Processing of dna for nonhomologous end-joining by cell-free extract. *The EMBO Journal*, 24(4):849–860, 2005.

[10] W K Rathmell and G Chu. Involvement of the ku autoantigen in the cellular response to dna double-strand breaks. *Proceedings of the National Academy of Sciences*, 91(16):7623–7627, 1994.

[11] F Liang, P J Romanienko, D T Weaver, P A Jeggo, and M Jasin. Chromosomal double-strand break repair in ku80-deficient cells. *Proceedings of the National Academy of Sciences*, 93(17):8929–8933, 1996.

[12] Ola Hammarsten and Gilbert Chu. DNA-dependent protein kinase: DNA binding and activation in the absence of Ku. *Proceedings of the National Academy of Sciences*, 95(2):525–530, 1998.

[13] Yunmei Ma, Ulrich Pannicke, Klaus Schwarz, and Michael R. Lieber. Hairpin Opening and Overhang Processing by an Artemis/DNA-Dependent Protein Kinase Complex in Nonhomologous End Joining and V(D)J Recombination . *Cell*, 108(6):781 – 794, 2002.

[14] Ulf Grawunder, Matthias Wilm, Xiantuo Wu, Peter Kulesza, Thomas E Wilson, Matthias Mann, and Michael R Lieber. Activity of DNA ligase IV stimulated by complex formation with XRCC4 protein in mammalian cells. *Nature*, 388(6641):492–495.

[15] Steven J. DiBiase, Zhao-Chong Zeng, Richard Chen, Terry Hyslop, Walter J. Curran, and George Iliakis. DNA-dependent Protein Kinase Stimulates an Independently Active, Nonhomologous, End-Joining Apparatus. *Cancer Research*, 60(5):1245–1253, 2000.

[16] Andrea Beucher, Julie Birraux, Leopoldine Tchouandong, Olivia Barton, Atsushi Shibata, Sandro Conrad, Aaron A Goodarzi, Andrea Krempler, Penny A Jeggo, and Markus Löbrich. ATM and Artemis promote homologous recombination of radiation-induced DNA double-strand breaks in G2. *The EMBO Journal*, 28(21):3413–3427, 2009.

[17] Takashi Ochi, Andrew N. Blackford, Julia Coates, Satpal Jhujh, Shahid Mehmood, Naoka Tamura, Jon Travers, Qian Wu, Viji M. Draviam, Carol V. Robinson, Tom L. Blundell, and Stephen P. Jackson. PAXX, a paralog of XRCC4 and XLF, interacts with Ku to promote DNA double-strand break repair. *Science*, 347(6218):185–188, 2015.

[18] Frédéric Pâques and James E. Haber. Multiple Pathways of Recombination Induced by Double-Strand Breaks in Saccharomyces cerevisiae . *Microbiology and Molecular Biology Reviews*, 63(2):349–404, 1999.

[19] F L Lin, K Sperle, and N Sternberg. Model for homologous recombination during transfer of DNA into mouse L cells: role for DNA ends in the recombination process. *Molecular and Cellular Biology*, 4(6):1020–1034, 1984.

[20] F L Lin, K Sperle, and N Sternberg. Recombination in mouse L cells between DNA introduced into cells and homologous chromosomal sequences. *Proc Natl Acad Sci*, 82(5):1391–5, 1985.

[21] Sarah K Deng, Bryan Gibb, Mariana Justino de Almeida, Eric C Greene, and Lorraine S Symington. RPA antagonizes microhomology-mediated repair of DNA double-strand breaks. *Nat Struct Mol Biol*, 21(4):405–412, 04 2014.

[22] Xiahui Xiong, Zhanwen Du, Ying Wang, Zhihui Feng, Pan Fan, Chunhong Yan, Henning Willers, and Junran Zhang. 53BP1 promotes microhomology-mediated end-joining in G1-phase cells. *Nucleic Acids Research*, 43(3):1659–1670, 2015.

[23] Wael Y. Mansour, Tim Rhein, and Jochen Dahm-Daphi. The alternative end-joining pathway for repair of DNA double-strand breaks requires PARP1 but is not dependent upon microhomologies. *Nucleic Acids Research*, 38(18):6065–6077, 2010.

[24] Minli Wang, Weizhong Wu, Wenqi Wu, Bustanur Rosidi, Lihua Zhang, Huichen Wang, and George Iliakis. PARP-1 and Ku compete for repair of DNA double strand breaks by distinct NHEJ pathways. *Nucleic Acids Research*, 34(21):6170–6182, 2006.

[25] Agnes Schipler and George Iliakis. Dna double-strand-break complexity levels and their possible contributions to the probability for error-prone processing and repair pathway choice. *Nucleic Acids Research*, 41(16):7589–7605, 2013.

[26] Michael M. Vilenchik and Alfred G. Knudson. Endogenous dna double-strand breaks: Production, fidelity of repair, and induction of cancer. *Proceedings of the National Academy of Sciences*, 100(22):12871–12876, 2003.

[27] Ketki Karanam, Ran Kafri, Alexander Loewer, and Galit Lahav. Quantitative Live Cell Imaging Reveals a Gradual Shift between DNA Repair Mechanisms and a Maximal Use of HR in Mid S Phase. *Molecular Cell*, 47(2):320 – 329, 2012.

[28] Pierre Hentges, Helen Waller, Clara C. Reis, Miguel Godinho Ferreira, and Aidan J. Doherty. Cdk1 Restrains NHEJ through Phosphorylation of XRCC4-like Factor Xlf1. *Cell Reports*, 9(6):2011–2017, 2015/03/23.

[29] Raphael Ceccaldi, Jessica C. Liu, Ravindra Amunugama, Ildiko Hajdu, Benjamin Primack, Mark I. R. Petalcorin, Kevin W. O'Connor, Panagiotis A. Konstantinopoulos, Stephen J. Elledge, Simon J. Boulton, Timur Yusufzai, and Alan D. D'Andrea. Homologous-recombination-deficient tumours are dependent on Polθ-mediated repair. *Nature*, 518(7538):258–262, 02 2015.

[30] Nam Woo Cho and Roger A. Greenberg. Dna repair: Familiar ends with alternative endings. *Nature*, 518(7538):174–176, 02 2015.

[31] Wenqi Wu, Minli Wang, Weizhong Wu, Satyendra K Singh, Tamara Mussfeldt, and George Iliakis. Repair of radiation induced DNA double strand breaks by backup NHEJ is enhanced in G2. *DNA Repair*, 7(2):329–338, February 2008.

[32] Atsushi Shibata, Sandro Conrad, Julie Birraux, Verena Geuting, Olivia Barton, Amani Ismail, Andreas Kakarougkas, Katheryn Meek, Gisela Taucher-Scholz, Markus Löbrich, and Penny A Jeggo. Factors determining DNA double-strand break repair pathway choice in G2 phase. *The EMBO Journal*, 30(6):1079–1092, 2011.

[33] Kelly M. Trujillo, Shyng-Shiou F. Yuan, Eva Y.-H. P. Lee, and Patrick Sung. Nuclease Activities in a Complex of Human Recombination and DNA Repair Factors Rad50, Mre11, and p95. *Journal of Biological Chemistry*, 273(34):21447–21450, 1998.

[34] Xiaoyi Deng, Aishwarya Prakash, Kajari Dhar, Gilson S. Baia, Carol Kolar, Greg G. Oakley, and Gloria E. O. Borgstahl. Human Replication Protein A Rad52 Single Stranded DNA Complex: Stoichiometry and Evidence for Strand Transfer Regulation by Phosphorylation. *Biochemistry*, 48(28):6633–6643, 2009. PMID: 19530647.

[35] U H Mortensen, C Bendixen, I Sunjevaric, and R Rothstein. DNA strand annealing is promoted by the yeast Rad52 protein. *Proc Natl Acad Sci U S A*, 93(20):10729–34, 1996.

[36] Bernd Götlich, Susanne Reichenberger, Elke Feldmann, and Petra Pfeiffer. Rejoining of DNA double-strand breaks in vitro by single-strand annealing. *European Journal of Biochemistry*, 258(2):387–395, 1998.

[37] J Fishman-Lobell, N Rudin, and J E Haber. Two alternative pathways of double-strand break repair that are kinetically separable and independently modulated. *Molecular and Cellular Biology*, 12(3):1292–1303, 1992.

[38] Huichen Wang, Zhao-Chong Zeng, Tu-Anh Bui, Eiichiro Sonoda, Minoru Takata, Shunichi Takeda, and George Iliakis. Efficient rejoining of radiation-induced DNA double-strand breaks in vertebrate cells deficient in genes of the RAD52 epistasis group. *Oncogene*, 20(18), 2001.

[39] Huichen Wang, Ange Ronel Perrault, Yoshihiko Takeda, Wei Qin, Hongyan Wang, and George Iliakis. Biochemical evidence for Ku-independent backup pathways of NHEJ. *Nucleic Acids Research*, 31(18):5377–5388, 09 2003.

[40] Emil Mladenov, Simon Magin, Aashish Soni, and George Iliakis. DNA Double-Strand Break Repair as determinant of cellular radiosensitivity to killing and target in radiation therapy. *Frontiers in Oncology*, 3(113), 2013.

[41] George Iliakis. Backup pathways of NHEJ in cells of higher eukaryotes: Cell cycle dependence. *Radiotherapy and Oncology*, 92(3):310 – 315, 2009. Special Issue on Molecular and Experimental Radiobiology Including papers from the 11th International Wolfsberg Meeting on Molecular Radiation Biology/Oncology.

[42] Jia-Lin Ma, Eun Mi Kim, James E. Haber, and Sang Eun Lee. Yeast Mre11 and Rad1 Proteins Define a Ku-Independent Mechanism To Repair Double-Strand Breaks Lacking Overlapping End Sequences. *Molecular and Cellular Biology*, 23(23):8820–8828, 2003.

[43] Aashish Soni, Maria Siemann, Martha Grabos, Tamara Murmann, Gabriel E. Pantelias, and George Iliakis. Requirement for Parp-1 and DNA ligases 1 or 3 but not of Xrcc1 in chromosomal translocation formation by backup end joining. *Nucleic Acids Research*, 42(10):6380–6392, 2014.

[44] Yu Zhang and Maria Jasin. An essential role for CtIP in chromosomal translocation formation through an alternative end-joining pathway. *Nat Struct Mol Biol*, 18(1):80–84, 01 2011.

[45] Marc Audebert, Bernard Salles, and Patrick Calsou. Involvement of Poly(ADP-ribose) Polymerase-1 and XRCC1/DNA Ligase III in an Alternative Route for DNA Double-strand Breaks Rejoining. *Journal of Biological Chemistry*, 279(53):55117–55126, 2004.

[46] Li Liang, Li Deng, Son C. Nguyen, Xin Zhao, Christopher D. Maulion, Changshun Shao, and Jay A. Tischfield. Human DNA ligases I and III, but not ligase IV, are required for microhomology-mediated end joining of DNA double-strand breaks. *Nucleic Acids Research*, 36(10):3297–3310, 2008.

[47] Sean M. Howard, Diana A. Yanez, and Jeremy M. Stark. DNA Damage Response Factors from Diverse Pathways, Including DNA Crosslink Repair, Mediate Alternative End Joining. *PLoS Genet*, 11(1):e1004943, 01 2015.

[48] Metzger L and Iliakis G. Kinetics of DNA double-strand break repair throughout the cell cycle as assayed by pulsed field gel electrophoresis in CHO cells. *Int J Radiat Biol*, 59(1), 1991.

[49] Francis A Cucinotta, Janice M Pluth, Jennifer A Anderson, Jane V Harper, and Peter O'Neill. Biochemical Kinetics Model of DSB Repair and Induction of gamma-H2AX Foci by Non-homologous End Joining. *Radiation Research*, 169(2):214–222, February 2008.

[50] Reza Taleei, Michael Weinfeld, and Hooshang Nikjoo. Single strand annealing mathematical model for double strand break repair. *journal of Molecular Engineering and Systems Biology*, 1(1), 2012.

[51] Reza Taleei and Hooshang Nikjoo. The Non-homologous End-Joining (NHEJ) Pathway for the Repair of DNA Double-Strand Breaks: I. A Mathematical Model. *Radiation Research*, 179(5):530–539, May 2013.

[52] Reza Taleei, Michael Weinfeld, and Hooshang Nikjoo. A kinetic model of single-strand annealing for the repair of DNA double-strand breaks. *Radiation Protection Dosimetry*, 143(2-4):191–195, 2011.

[53] Werner Friedland, Pavel Kundrát, and Peter Jacob. Stochastic modelling of DSB repair after photon and ion irradiation. *International Journal of Radiation Biology*, 88(1-2):129–136, 2012.

[54] Werner Friedland, Peter Jacob, and Pavel Kundrát. Stochastic Simulation of DNA Double-Strand Break Repair by Non-homologous End Joining Based on Track Structure Calculations. *Radiation Research*, 173(5):677–688, January 2013.

[55] Yongfeng Li, Hong Qian, Ya Wang, and Francis A. Cucinotta. A Stochastic Model of DNA Fragments Rejoining. *PLoS ONE*, 7(9):e44293, 09 2012.

[56] Oleg V. Belov, Eugene A. Krasavin, Marina S. Lyashko, Munkhbaatar Batmunkh, and Nasser H. Sweilam. A quantitative model of the major pathways for radiation-induced DNA double-strand break repair. *Journal of Theoretical Biology*, 366(0):115 – 130, 2015.

[57] Paul Verbruggen, Tim Heinemann, Erik Manders, Gesa von Bornstaedt, Roel van Driel, and Thomas Höfer. Robustness of DNA Repair through Collective Rate Control. *PLoS Comput Biol*, 10(1):e1003438, 01 2014.

[58] Tina Toni, David Welch, Natalja Strelkowa, Andreas Ipsen, and Michael P H Stumpf. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society, Interface / the Royal Society*, 6(31):187–202, February 2009.

[59] Tina Toni and Michael P H Stumpf. Simulation-based model selection for dynamical systems in systems and population biology. *Bioinformatics*, 26(1):104–110, January 2010.

[60] Juliane Liepe, Chris Barnes, Erika Cule, Kamil Erguler, Paul Kirk, Tina Toni, and Michael P.H. Stumpf. Abc-sysbio approximate bayesian computation in python with gpu support. *Bioinformatics*, 26(14):1797–1799, 2010.

[61] Juliane Liepe, Paul Kirk, Sarah Filippi, Tina Toni, Chris P Barnes, and Michael P H Stumpf. A framework for parameter estimation and model selection from experimental data in systems biology using approximate Bayesian computation. *Nature protocols*, 9(2):439–456, February 2014.

[62] Siegmund Kimberly D., Marjoram Paul, and Shibata Darryl. Modeling DNA Methylation in a Population of Cancer Cells. *Statistical Applications in Genetics and Molecular Biology*, 7(1):1–23, 2008.

[63] Daniel T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, 81(25):2340–2361, December 1977.

[64] Frank Windhofer, Wenqi Wu, and George Iliakis. Low levels of DNA ligases III and IV sufficient for effective NHEJ. *Journal of Cellular Physiology*, 213(2):475–483, 2007.

[65] Y Yamaguchi-Iwai, E Sonoda, J M Buerstedde, O Bezzubova, C Morrison, M Takata, A Shinohara, and S Takeda. Homologous recombination, but not DNA repair, is reduced in vertebrate cells deficient in RAD52. *Mol Cell Biol*, 18(11):6430–5, 1998.

[66] Henrik Stryhn and Jette Christensen. The analysis–Hierarchical models: Past, present and future. *Preventive Veterinary Medicine*, 113(3):304 – 312, 2014. Special Issue: Schwabe Symposium 2012.

[67] G. Iliakis, H. Wang, A. R. Perrault, W. Boecker, B. Rosidi, F. Windhofer, W. Wu, J. Guan, G. Terzoudi, and G. Pantelias. Mechanisms of DNA double strand break repair and chromosome aberration formation. *Cytogenetic and Genome Research*, 104(1-4):14–20, 2004.

[68] Jean-François Haince, Darin McDonald, Amélie Rodrigue, Ugo Déry, Jean-Yves Masson, Michael J. Hendzel, and Guy G. Poirier. PARP1-dependent Kinetics of Recruitment of MRE11 and NBS1 Proteins to Multiple DNA Damage Sites. *Journal of Biological Chemistry*, 283(2):1197–1208, 2008.

7 Figure Legends

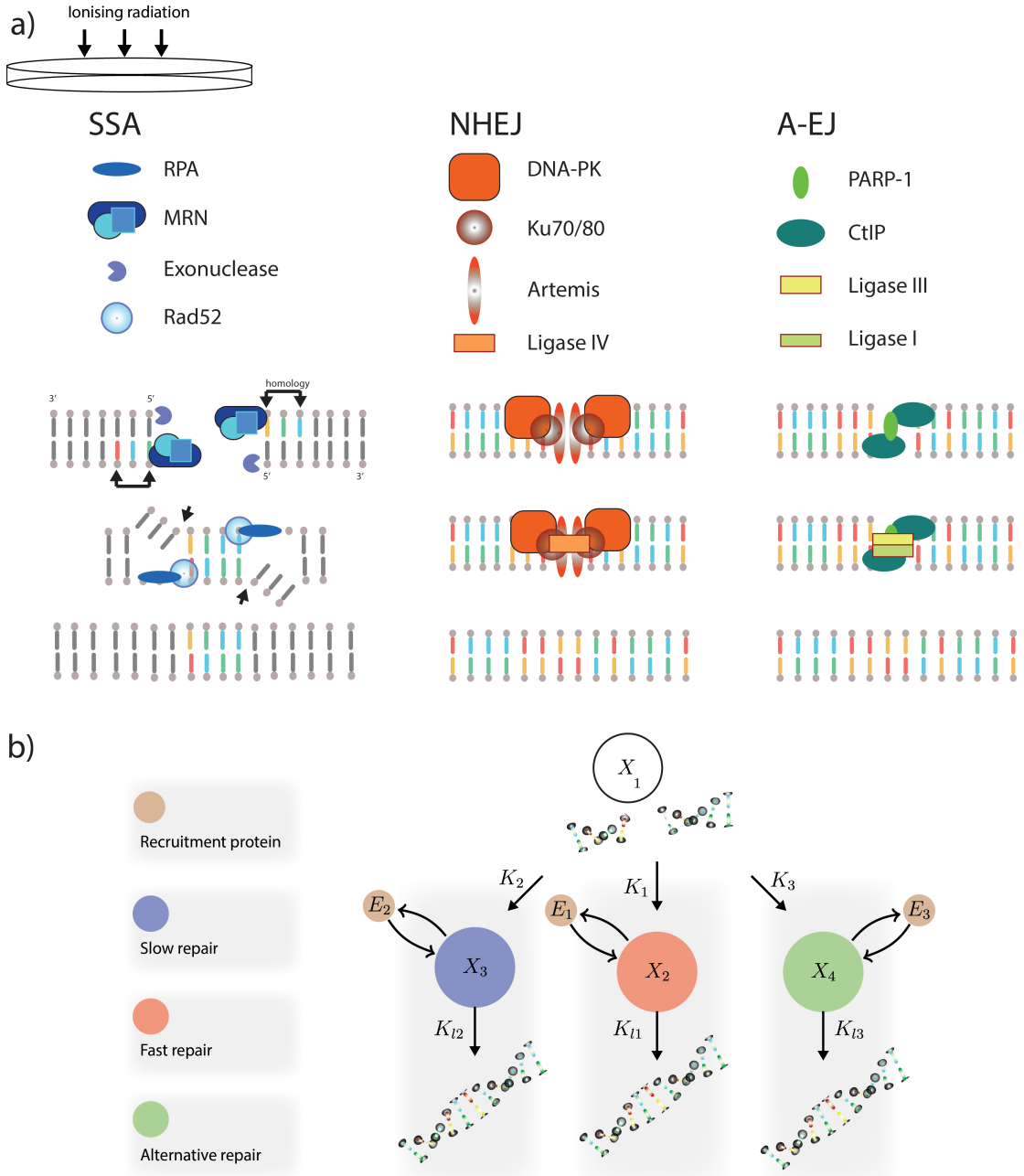

Figure 1: Modelling multiple repair mechanisms. a) Proteins and repair steps contributing to repair during SSA, NHEJ and A-EJ in mammalian cells (illustration). b) The model. Discs represent species and arrows represent reactions.

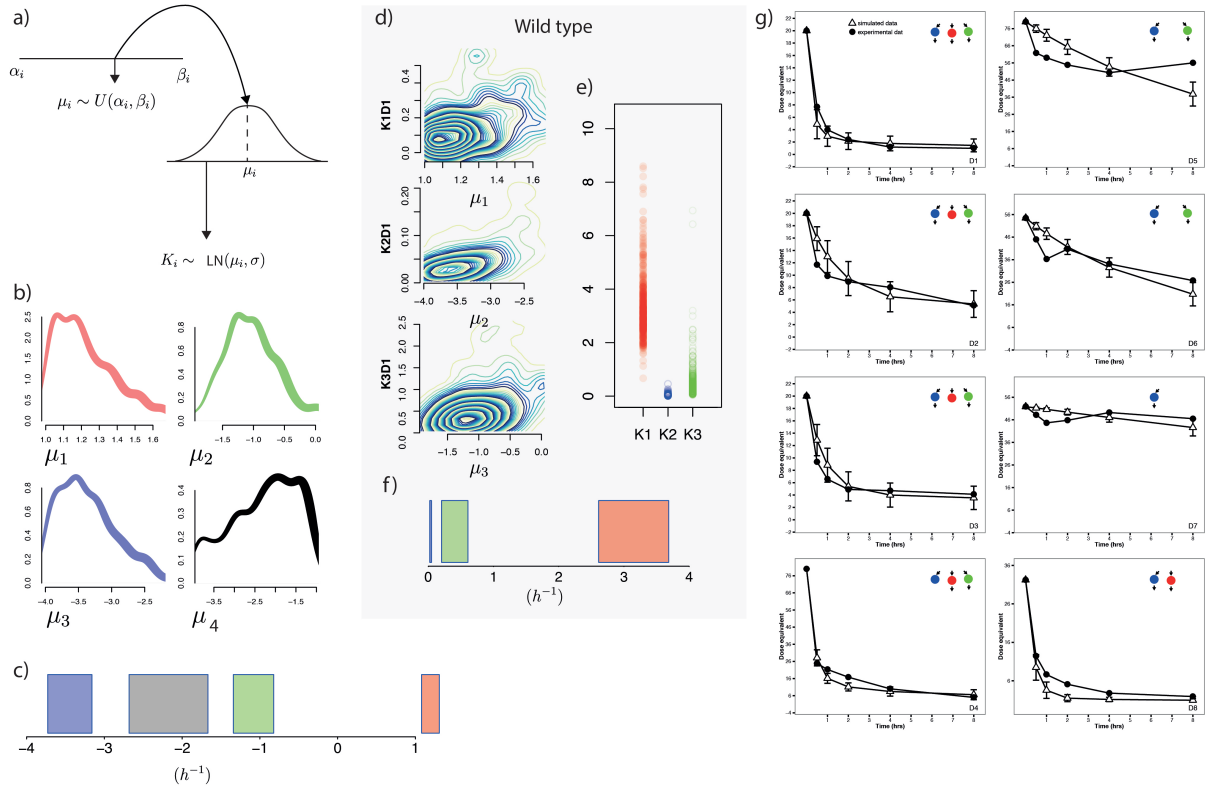

Figure 2: a) Diagram showing the parameter sampling process. Hyper parameters μ_i are drawn from a uniform distribution between α_i and β_i. Model parameters K_i are sampled from a lognormal distribution with mean μ_i. b) Posterior distributions for the hyper parameters μ_{1-3} and μ_4. c) Box plot showing the interquartile ranges of the hyper parameters. d) Posterior analysis for dataset $D1$. Marginal distributions of the parameters $K1D1 - K3D1$ against the hyper parameters, (top left). Posterior distributions of the parameters $K1D1 - K3D1$, showing some overlap (top right). Interquartile range of the parameters $K1D1 - K3D1$ (bottom). e) Time series plots of the experimental data and model simulation. Figures on the top right represent the active repair mechanisms. Red, blue and green represent fast, slow and alternative repair.

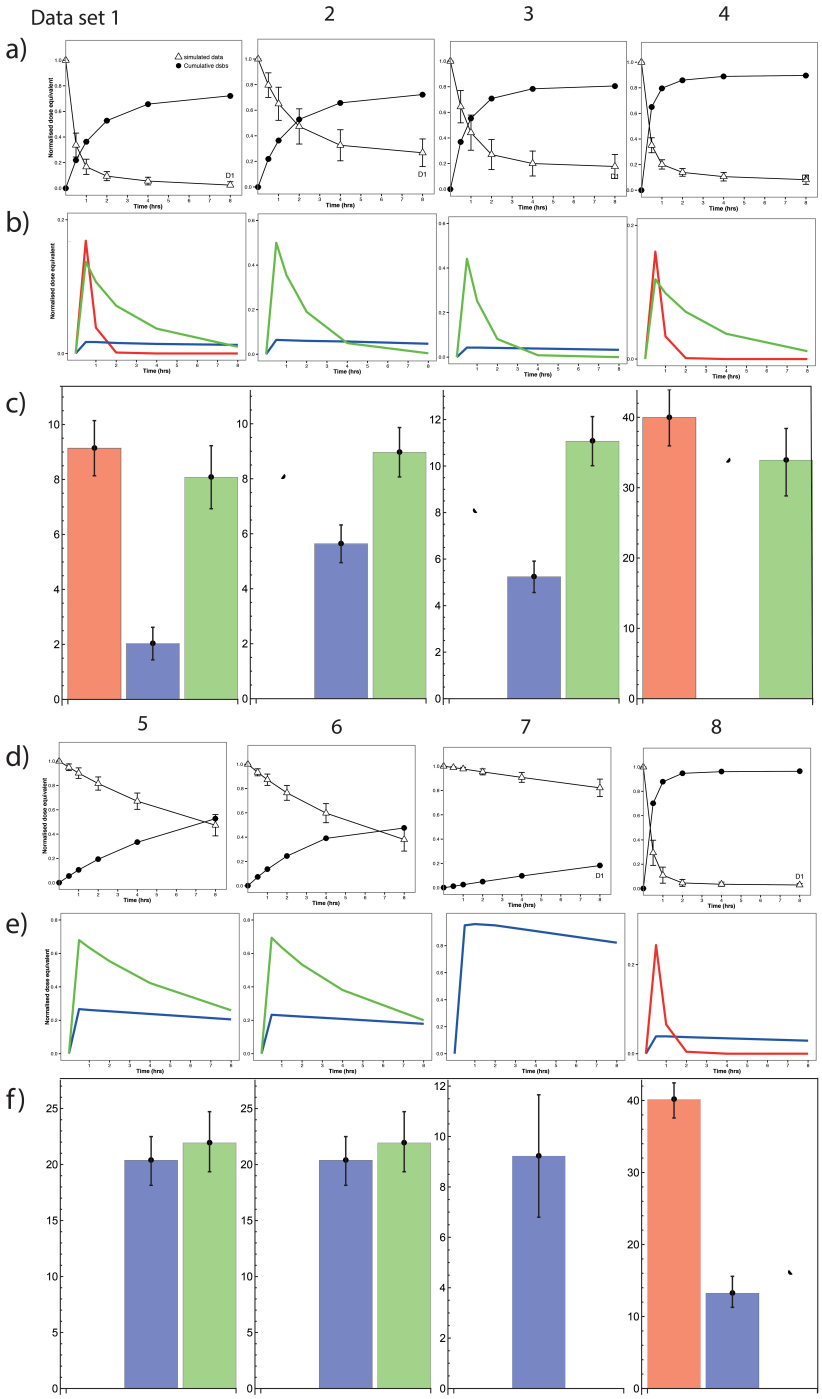

Figure 3: Identifying fast, slow and alternative mechanisms. a) Simulated data and Cumulative DSBs. b) DSBs entering each repair mechanism. c) Total amount of DSBs repaired by fast, slow and alternative repair. a-c) Results shown for datasets D1-D4. d) Simulated data and Cumulative DSBs. e) Time series of DSBs entering each repair mechanism. f) Total DSBs repaired by fast, slow and alternative repair. d-f) Results shown for datasets D4-D8. Red, blue and green represent fast, slow and alternative repair.

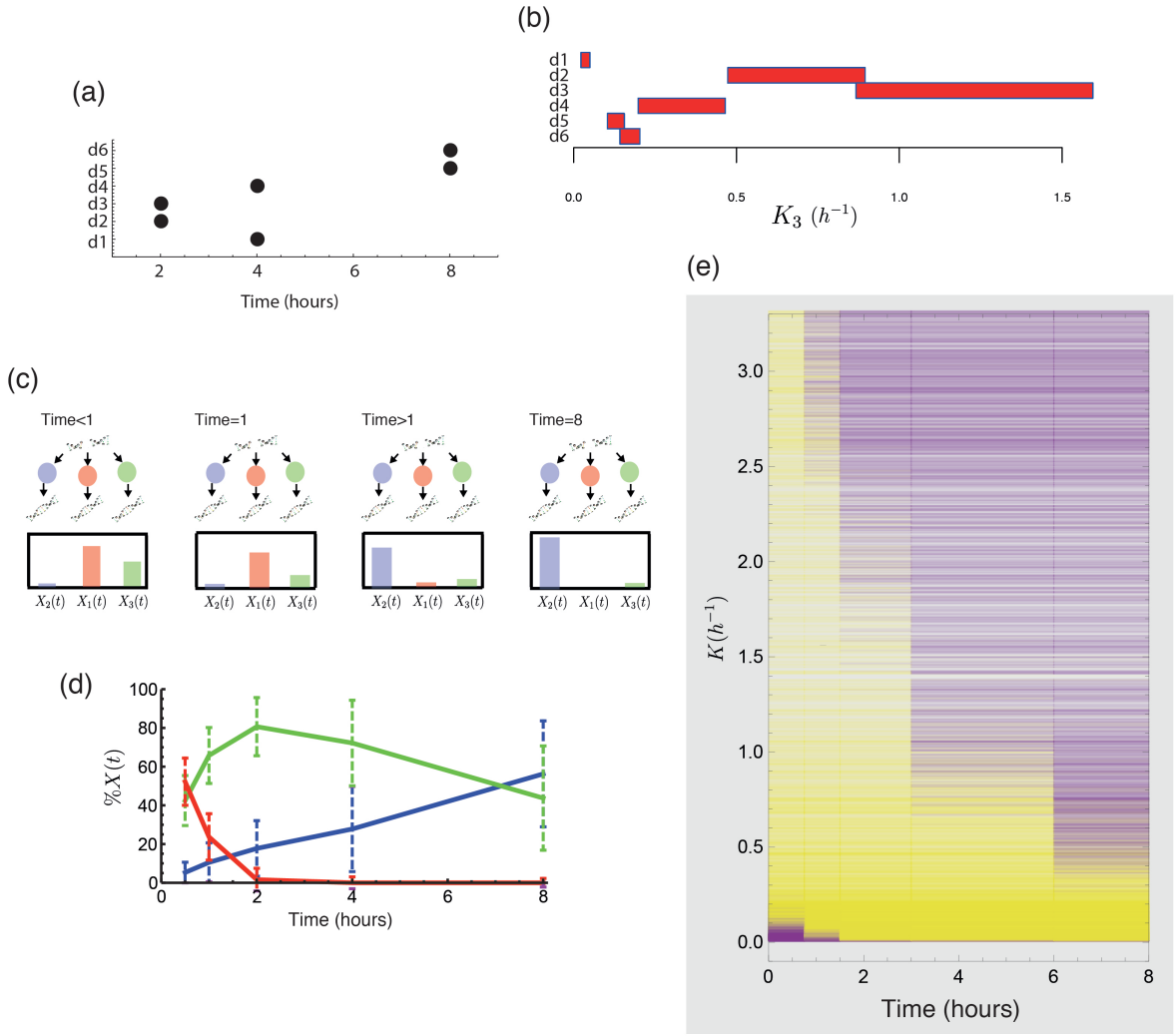

Figure 4: a). Time in which a repair curve has reached below half it's maximum value for each data set in which A-EJ is assumed to be active. The slowest mode of repair occurred in data sets 5 and 6, where Ku70 is inactive. b). Rectangle plot of the interquartile ranges of K_3 for all datasets where A-EJ is assumed to be active. c). Illustration, showing a typical distribution of the DSBs that remain to be repaired over time. For times $<$ 1hour a large proportion of DSBs are being repaired by fast NHEJ and faster A-EJ mechanisms, whereas at later times, the majority of DSBs reside in slower HR mechanisms. d). Time series showing the percentage of remaining DSBs in each repair pathway for the wild type data D1. e). Plot showing the time at which each repair mechanism is greater than 30% active for different parameter values. Purple indicates that the mechanism is less than 30% active and yellow indicates the mechanism is greater than 30% active.