Alternative end joining mechanisms exhibit varying dynamics in 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 multiple repair pathways and can give rise to different mutagenic processes. The propensity for activation directly affects genomic instability, with implications across health and evolution. However, the relative contribution of these mechanisms, their interplay and regulatory interactions remain to be fully elucidated. Here we present a new method to model the combined activation of non-homologous end joining, single strand annealing and alternative end joining. We use Bayesian statistics to integrate eight biological data sets of DSB repair curves under varying genetic knockouts. Analysis of the model suggests that there are at least three disjoint modes of repair, which we assign as fast slow and intermediate. Our results show that when multiple datasets are combined, the rate for intermediate repair is not constant amongst genetic knockouts. Further analysis suggests that the ratio between slow and intermediate repair depends on the presence or absence of DNAPKcs and Ku70. We outline how our insights can be directly tested using imaging and sequencing techniques and conclude that there is evidence of variable dynamics in alternative repair pathways. Our approach is an important 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 lesions in DNA that occur naturally by oxidative stress, DNA replication and exogenous sources [1, 2]. When left unprocessed or during erroneous repair, they cause changes to DNA structure creating mutations and potential genomic instability [3–8]. To repair DSBs, multiple mechanisms have evolved and are known to include 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 choice of mechanism depends on the structure of the break point, where simple breaks caused by restriction enzymes are different in structure from those caused by ionising radiation (IR) (reviewed in [25, 26]). This affects the probability of error prone repair because mutations are mechanism specific and depending on which mechanism is activated a cell might exhibit chromosome translocations [4, 5], small deletions or insertions [6, 7] and

recombination leading to loss of heterozygosity [8]. For example, in mouse, error by SSA causes chromosome translocations [4] and in *Saccharomyces cerevisiae*, NHEJ of simple DSBs is associated with small deletions or insertions [7]. *In vivo* studies of DSBs have suggested that in addition to structural activation arising from the break point, cell-cycle dynamics also play a role in repair mechanism activation (reviewed in [27]). In particular the choice of mechanism is not fixed at the time of damage and cells exhibit a pulse like repair in human U-2 OS cells [28], a behaviour supported by a molecular basis for cell cycle dependence in NHEJ, mediated by Xlf1 phosphorylation [29].

To understand how mutations are distributed in the genome, it is important to uncover the dynamic activation and interplay between different DSB repair mechanisms. This mutual activation is not fully understood, however the individual repair mechanisms and recruitment proteins of NHEJ, SSA and A-EJ have been documented. NHEJ requires little or no homology, is a mechanism of DNA end joining in both unicellular and multicellular organisms [7] and can exhibit fast repair by the binding of DNA-PK [15]. In vertebrates, NHEJ initiates the recruitment and binding of several proteins. 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]. Repair proceeds by Artemis facilitated overhang processing and end ligation via DNA Ligase IV [13,14]. Although well studied, new regulating components of NHEJ are still being discovered, for example the protein PAXX [17]. SSA is slower than NHEJ and in yeast works at almost 100% efficiency for homologous regions of at least 400bp [6]. First described in mouse fibroblast cells [19,20], during SSA two complementary sequences are exposed through a 5′ to 3′ exonuclease end resection and aligned. Remaining overhangs are then 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 [30]. Following resection, replication protein A (RPA) binds to the DNA and when phosphorylated, forms a complex with Rad52 to stimulates DNA annealing [31,32]. Similarly to NHEJ, following gap repair, SSA is terminated with end ligation by Ligase III [33]. 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 [34], suggesting that SSA may be active in the repair of DSBs caused by IR. 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 [35]. One interesting finding in genetic studies is that when NHEJ is compromised, DSBs are removed by alternative mechanisms that we collectively refer to as A-EJ [23,36], (reviewed in [37–39]). It is still unclear how A-EJ is regulated or interacts with other processes and there is evidence that the mechanisms are active in the repair of breaks with microhomology of 3-16 nucleotides, reviewed by Decottignies [39]. The mechanism has adopted various names in the literature, such as MMEJ in yeast [40] and back-up NHEJ (B-NHEJ) in higher eukaryotes [38]. Thought to act on break points with ends that are not complementary in the absence of NHEJ factors [36], an assortment of PARP-1, 53BP1, Lig3 and 1, Mre11, CtIP and Polθ have been proposed as regulators of A-EJ. PARP-1 is required and competes with Ku for binding to DNA ends through the PARP-1 DNA binding domain [24]. Other proteins are involved in initial binding, where activation of 53BP1 in MMEJ is dependent on Ku70 and independent of DNAPKcs [22] and CtIP has been associated through the use of microhomology [41]. The proteins required for end joining have been identified as Lig3 and Lig1 in the absence of XRCC1 [42–44]. 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 [45]. For an illustration of the three mechanisms (see Figure 1a).

Mathematical models of DSB repair have used biphasic [51], biochemical kinetic [52–57], multiscale [58, 59], and stochastic methods [60]. In a study by Cucinotta *et al.* [53], a set of coupled nonlinear ordinary differential equations were developed. The model was based on the law of mass action and stepwise irreversible binding of repair proteins to describe NHEJ rejoining kinetics and the phosphorylation of H2AX by DNA-PKcs. Similar studies have modelled repair kinetics and protein recruitment during SSA [54, 56], NHEJ [55] and other mechanisms including NHEJ, HR, SSA and two alternative pathways under a wide range of LET values and heavy ions [57]. These studies have been met with some controversy, for example with the argument that the biphasic model has never succeeded in providing definitive values for the repair components [48]. Recently however, the models have been further developed to model the complexity of a DSB by application to damage induced by ionising radiation of different qualities [46, 47]. This can be achieved because the spectrum of DSB-DNA damage can be computed by applying Monte Carlo track structure [49], for which various alternatives are reviewed in [50]. Previous biochemical kinetic models have been used to reproduce the experimental data observed. This approach often uses more parameters than 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, an issue known as identifiability. Consequently, predictions are not unique, which can be detrimental in the design of a biological experiment. Therefore the creation of models that provide a unique interpretation of repair dynamics is a challenge.

Here we develop a statistical model that can take DSB repair curve data, such as those generated from pulse field gel electrophoresis or comet assays, and infer repair mechanism activation. The method relies on training a simple model against multiple datasets of DSB repair under different genetic knockouts when multiple repair mechanisms are activated. Using the most probable set of parameter values, we can then simulate the model and make predictions on the activation of different rates of repair. Unlike previous modelling approaches, we do not model individual recruitment proteins. Instead we assign parameter values to different rates of repair. This has two benefits. Firstly it provides a method to uncover different rates of repair arising from different repair mechanisms that are implicit in the data. Secondly, it reduces the number of parameters required to describe the system, leading to a more identifiable model. 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. We use this method to integrate the data from eight repair curve assays under genetic knock outs including combinations of Ku70, DNA-PKcs, Rad52 and Rad54. We show that there are at least three disjoint dynamical repair mechanisms that explain the combined data and the dynamics depend on the regulating recruitment proteins. We propose that there are a number of alternative end joining dynamical processes that may or may not share a common genetic pathway. We also show that the activation of different repair processes over time depends on the speed of the dynamics.

2 Materials and Methods

A model of fast, slow and intermediate repair We assume DSBs caused by ionising radiation (IR) can be repaired at fast, slow and intermediate rates and propose that these could describe NHEJ, SSA and A-EJ respectively, see Figure 1b. Other mechanisms could be included, such as Rad54 dependent homologous recombination but this mechanism is not assumed to contribute substantially in the datasets that we use to assign parameter values in our model [61]. 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. Although we assume a set of fast, slow and intermediate repair, there is an extent to which the three pathways may overlap. For example, it is unclear whether alternative end joining depends on the presence of microhomologies and it has been proposed that the complex MRX may be required [62]. To address this, we model the data using distributions of the repair rates and allow for some overlap between the slow and intermediate mechanisms, which is known as a hierarchical model in statistical terminology. DNA repair is modelled with a stochastic reaction system consisting of six reactions on a population of DSBs (see Supplementary Data). In summary, the system models a population of DSBs, that in the presence of repair regulating components that we denote by E, can enter either fast, slow or intermediate repair, (Figure 1b). The recruitment reactions and three ligation reactions can be summarised:

$$\begin{aligned} \mathrm{DSB} + \mathrm{E} &\xrightarrow{K} \mathrm{Repair} \\ \mathrm{Repair} &\xrightarrow{K} \mathrm{Ligation}, \end{aligned} \tag{1}$$

where E represents one of the repair proteins Ku, MRN and PARP-1 and K represents the binding dependent rate for initial protein recruitment and end ligation (Figure 1b). Note that we constrain the parameter K to be equal in the recruitment and repair reactions. To model the limited resources available to the cell, we follow the approach of Cucinotta *et al.*, [53], and assume that the total amount of protein is conserved for each repair mechanism, which 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 Data. At any time, t, the total observed DSBs, $x(t)$, are given by the sum of all states for all DSBs in the population, where N_j is the number of DSBs in state j. The proportion of DSBs repaired by each mechanism are estimated by calculating the cumulative number of DSBs that enter each individual pathway (see Supplementary Data).

Experimental data 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 ob-

tained 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. The eight data sets are labelled D1-D8. Data D1 represents the wild type in this study and, since the cell cycle phase is unrestricted, we expect all three repair processes to be present. Data D2 and D3 are DNA-PKcs knockouts in G1 and G2 phase, where we expect NHEJ to be compromised but since Ku is present we still expect the recruitment process. Data D4 is a Rad52 knockout where we expect only NHEJ and A-EJ to be present. Data D5 and D6 are Ku knockouts and we assume that the whole of the NHEJ pathway compromised and only SSA and A-EJ remain. Data D7 and D8 are expected to have no repair by PARP-1 mediated A-EJ because both sets were treated with PARP-1 inhibitors. Data D7 comes from $Ku70^{-/-}$ mouse fibroblasts, where we expect to see no repair by NHEJ as well as a lack of A-EJ due to PARP-1 inhibition [24].

Parameter estimation and approximate Bayesian computation To build a model that can be used to obtain unique predictions, it is advantageous to minimise the number of parameters that describe the system. To do this, we developed a hierarchical model [67], 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 2a). By drawing parameters from one common hyper parameter across the datasets, the total number of parameters 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 approximate Bayesian computation sequential Monte Carlo (ABC SMC) [68–71]. This 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 methylation [72]. The method is used 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 the Supplementary Data.

3 Results

3.1 DSB repair requires 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 fit of the simulation to the data for all eight data sets is shown in Figure 2g. The fits capture the essential aspects of the repair curves and most points are consistent with the posterior distribution. The posterior distributions of the hyper parameters are shown in Figure 2b. 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 $D = 1$,

Data set	Dose (Gy)	Phase	Cell line, Mutant	Repair Mechanisms
D1	20	Asynchronous	WT MEFs* [61]	NHEJ, SSA, A-EJ
D2	20	G1	DNA-PKcs$^{-/-}$ MEFs* [61]	SSA, A-EJ
D3	20	G2	DNA-PKcs$^{-/-}$ MEFs* [61]	SSA, A-EJ
D4	80	Asynchronous	Rad52$^{-/-}$ DT40* [66]	NHEJ, A-EJ
D5	80	Asynchronous	Ku70$^{-/-}$/Rad54$^{-/-}$ DT40* [34]	SSA, A-EJ
D6	54	Asynchronous	Ku70$^{-/-}$ DT40* [34]	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 [66].

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 fitted using ABC SMC. For prior distributions on the hyper parameters, see Supplementary Data.

Data	Model Parameters			Mean (hrs^{-1}), Halftime (hrs)			Literature, (hrs^{-1})
	Fast	**|Slow**	**|A-EJ**	**Fast**	**|Slow**	**|A-EJ**	
D1, 20	K_{1d1}	K_{2d1}	K_{3d1}	3.3, 1	0.045, >8	0.51, 4	NHEJ: $k_1 = 0.5$
D2, 20	K_{1d2}	K_{2d2}	K_{3d2}	0.15	0.047, >8	0.74, 4	[55]
D3, 20	K_{1d3}	K_{2d3}	K_{3d3}	0.13	0.046, >8	1.3, 2	SSA: $k_1 = 0.057$
D4, 80	K_{1d4}	K_{2d4}	K_{3d4}	3.3, 1	0.17	0.39, 4	[56]
D5, 80	-	K_{2d5}	K_{3d5}	-	0.039, >8	0.14, 8	
D6, 54	-	K_{2d6}	K_{3d6}	-	0.04, >8	0.18, 8	
D7, 52	-	K_{2d7}	-	-	0.028, >8	-	
D8, 32	K_{1d8}	K_{2d8}	-	2.8, 1	0.047, >8	-	

Table 3: Mean parameter values and repair half times after ABC SMC. Values can be compared to initial binding rate constants from two models in the literature [55] and [56].

all tests $p < 2.2e^{-16}$). For each data set (D1-D8) the posterior interquartile 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 rates are defined as fast, slow and intermediate, there is variation in activation of the mechanisms among different mutants (Figure 2d). There is some overlap in parameter values K_1, K_2 and K_3 (Figure 2e) but the interquartile ranges of the parameters K_1, K_2 and K_3 are disjoint (Figure 2f) and this result was tested for an additional prior range, giving similar results (Supplementary Figure S3). This is also observed in all eight datasets (Supplementary Figure S1). For all posterior distributions of the parameters and a plot of individual DSBs and their repair in the wild type model, see Supplementary Figure S2. When individual DSBs are tracked in the model, the DSBs are quickly distributed amongst the three pathways and repaired according to the predicted rate. A plot of the repair of individual DSBs in the wild type model is presented in Supplementary Figure S4. To check that our model parameters were robust to adding additional data, we performed ABC SMC again on nine datasets. This new dataset consisted of the eight datasets listed in Table 1 and an additional repair time series of xrs-6 cells deficient in Ku80 inhibited of PARP-1 by DPQ [24], in which cells were deficient in NHEJ and A-EJ. The results of the total number of DSBs repaired and our predictions on the activation of A-EJ using this dataset were the same and a comparison between the eight data and nine data posterior showed a similar fit (Supplementary Figure S5 and S6). In summary, we conclude that the biological data can be explained by one fast, one slow and at least one intermediate rate of repair.

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

By re-simulating from our fitted model we were able to examine the dynamics of DSB repair across mechanisms and datasets. Repair and the cumulative repair were plotted for each data set (see Figure 3a,d). Data sets in which NHEJ is active exhibited a faster repair with the cumulative number of DSBs reaching to within 80% of the total within a period of 2 hours post irradiation. Next, we plotted the

number of DSBs entering each repair mechanism as a time series (Figure 3b,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 (D5,D6) than those without DNA-PKcs (D2,D3) (green curves, Figure 3b,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 3c,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 rates. In agreement with the results obtained from the time series plots (Figure 3b,e) there was a difference in the ratio of slow and intermediate rates between data sets D2,D3 and D5,D6. We also observed a difference in the number of DSBs repaired by A-EJ and slow rates between G1 and G2, corroborating with experimental results in the literature. Both rates increased and decreased significantly respectively (two sample t-test, $p<0.01$).

3.3 Fast and slow kinetics exhibit constrained rate parameters

To test the predictive capability of the model, we re-simulated repair curves using the posterior parameters for datasets D1-D8 and compared the simulated curves to new datasets from the literature (Figure 4). The model trajectories provide a good fit to wild type data at 40Gy (Figure 4a, [61]). More impressively, the simulations of cells defective in PARP-1 by application of 3'-AB and Ku80 deficient xrs-6 cells exposed to DPQ defective in PARP-1 and NHEJ show excellent agreement with the experimental curves (Figure 4b,c, [24]). This analysis suggests that small ranges of the rate paramaters for the fast and slow repair processes, when assigned to NHEJ and SSA, can predict multiple low LET datasets under different experimental conditions. This suggests that they are constrained across experimental and biological conditions. Next we investigated whether the intermediate rates from our model could potentially fit the experimental data when activation of fast and slow rates are inhibited. We re-simulated the model again but this time set the parameters for active NHEJ and SSA to zero across the models for D1-D8. The results are shown in Figure 4d along with the corresponding experimental data. In wild type data, we see that it is possible for all the DSBs to be removed with the predicted rates of A-EJ, although repair is slower, suggesting that the ability for A-EJ to repair DSBs is not saturated. When Rad52 is inhibited, because PARP-1 could still compete with MRN we see that it is not possible with the predicted rates for all the DSBs to be repaired in the absence of NHEJ. There was no clear difference in the repair of DNAPKcs mutants, however when Ku70 is inhibited (data D5, D6) we see that when competition by slow repair is removed the predicted rates of A-EJ, suggest that the activation of A-EJ is not saturated and is perhaps inhibited by competition with MRN.

3.4 Alternative end joining demonstrates variable dynamics

The time taken for over half the DSBs to be repaired by intermediate repair is shown in Figure 5a. 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. We also looked at the activation of A-EJ across the datasets by comparing the posterior distributions. The interquartile ranges of the posterior distributions for A-EJ are shown in Figure 5b. 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 DNA-PKcs is inhibited. These data suggest that A-EJ adopts a slow or fast repair and that the speed of repair depends on the presence or absence of DNA-PKcs 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 two ways in which this difference between Ku70 and DNA-PKcs 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 interpretation is that A-EJ is one repair mechanism that repairs at a slower rate when Ku70 is inhibited. We also modelled complete inhibition of intermediate repair when NHEJ is active to investigate how the assumptions of processes present within the datasets affects this conclusion. We found this does not change the range of the parameter distribution and we can recapture the dynamics observed, but the rate of fast repair is increased slightly in dataset D4 where there is inhibition of Rad52 (see Supplementary Figure S7.)

3.5 Total activation time of repair mechanisms depends on the rate of repair

Inspection of the time series data (Figure 3b,e) suggests that at time $t = 0.5$hrs, the majority of DSBs that are being processed are within the fast mode of repair. Figure 5c illustrates a typical distribution of DSBs over each mechanism at different points in time. At time $t = 0$, the cells are exposed to a single dose of ionising radiation. Quickly, for example at time $t < 1$hrs, fast and possibly faster alternative repair mechanisms dominate the DSB processing. Later, after all DSBs processed by the faster mechanisms have been repaired, the remaining DSBs fall within the category of breaks that require processing 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 of DSBs that remain in active repair mechanisms over time for the wild type data (see Figure 5d). By inspection, we can see that at 0.5 hours after irradiation around 50% and 5% of the remaining DSBs are within the fast and slow mechanisms respectively. At a time of $t = 8$hrs, the percentage changes to roughly zero for fast repair and around 60% for slow repair. The error bars in Figure 4d) are the standard deviation and this variation is due to the variation in repair rate K_i for each mechanism. Ultimately, it is the values of the parameters K_i that determine the rate of repair, so to confirm if the dynamics presented in Figure 5c 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 5e, where for each parameter K_i, a black line indicates the times at which the mechanism with rate K_i is greater than 30% active. There is a clear trend showing that the percentage of total activation decreases in time with an increase in repair rate K. In other words, the slower the repair process the longer it is actively repairing DSBs. When repair is extremely slow the repair mechanism never reaches 30 % of the current DSBs. In summary, these results predict 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 implied for NHEJ and HR (Figure 3 in [28]).

4 Discussion

In this study, we presented a new hierarchical model of DSB repair and applied 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. Bayesian computation constrains our simulated data to the biological data so that 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.

We have identified four major insights, each of which can be further tested experimentally. The first insight is that the data is explained by at least three independent mechanisms. Our results suggest that there are multiple dynamic regimes for the intermediate process. 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 out of DNA-PKcs). Another interesting finding is that intermediate repair is increased in G2 phase of the cell cycle. If we assume that intermediate repair corresponds to alternative 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 [61]. This marries with genetic studies that suggest two forms of alternative end joining depending on the presence of microhomology [39].

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 insight 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 fluorescent tagged 53BP1 [28], a protein that co-localises with alternative DSB markers and fluorescent tagged PARP1, a candidate protein for A-EJ [73]. In reference to existing evidence, this result is interesting because we observe a slower rate of intermediate repair when Ku70 is inhibited and a faster rate of intermediate repair when DNA-PKcs is inhibited (datasets D5-D6 and D2-D3 respectively, Figure 3). In the experimental literature, Ku deficient cells do not produce NHEJ products due to excessive degradation or inhibition of end joining [11] so in Ku knock outs (D5-D6) our model suggests that a slow A-EJ is probably active. In addition, inhibition of DNA-PKcs does not activate repair by PARP-1 mediated A-EJ [24] and leads to elevated levels of resection and more HR [74]. Together with our model, this suggests that an alternative mechanism that is faster than PARP-1 mediated A-EJ is activated when DNA-PKcs is inhibited (data sets D2-D3). In summary, our second insight suggests that PARP-1 mediated repair is slower than a faster alternative mechanism that becomes active when there are elevated levels of resection. Although, complementary studies in yeast have suggested that A-EJ is repressed by RPA which promotes error-free homologous recombination by preventing spontaneous annealing between microhomology which can lead to MMEJ [21].

The third insight is obtained by applying a density weighted integral to compute the total DSBs repaired by each mechanism. In the experimental literature, NHEJ is considered a fast repair process

because the local availability of DNA-PKcs leads to a fast rejoining process [15] and in mammalian cells, has been suggested to repair the majority of DSBs caused by IR [16]. This is upheld with our model predictions that when active, fast processes repair the most DSBs. A novel way in which we can use the number of DSBs repaired, is to estimate the proportion of mutations that are expected following DSB repair in wild type and mutant cells. Some cancers are deficient in at least one repair mechanism and in these cases, alternative mechanisms of repair have been observed to compensate [75]. This is apparent with an increase in chromosomal aberrations observed in cells compromised of NHEJ by loss of Ku80, suggesting a caretaker gene role for regulating components [3]. One mechanism that has been shown to play a role in cancer deficiency is A-EJ, where Polθ has been shown to be a necessary regulator for cell survival in homologous recombination deficient cancer [5]. 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, A-EJ 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 [44]. If we know how many DSBs are likely to be repaired by each mechanism, this information will be important in predicting the numbers and types of mutations that we expect to observe. Potentially, a better understanding of the interplay between DSB repair mechanisms could be applied to design potential synthetic lethal therapeutics in cancer [76].

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 [28] 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. This aspect of the model could be applied to general datasets. For example, if the rate of a repair mechanism is known by PFGE, then the time points at which the mechanism will dominate the repair after initial dose could be predicted.

Although we have made some predictions on the repair of DSBs quantified by PFGE, there are still some questions that will require further development before the model can be applied to general problems in DSB DNA repair. These limitations arise from the heterogeneity of the DNA damage spectrum, where complex breaks can account for 30-40% of the population [77]. Previous methods of Monte Carlo Track structure have been able to predict the number of single strand breaks, simple and complex DSBs that are created [49]. It is difficult however to determine the precise numbers of complex DSBs within a single repair curve that is generated following PFGE. By identifying proteins responsible for the processing of complex DSBs and analysing repair curves with and without their activity, it will be possible to gain understanding on the repair of heterogenous breaks. Recently, it has been suggested that the efficiency of ligations of DSB termini with proximal 8-oxoG lesions may be reduced relative to that for simple DSBs [77]. In the study by Dobbs et al., concerning base excision repair a potential role for NEIL-1 in the repair of complex DSBs is proposed. If it is the case that a substantial proportion of DSBs are complex, then inhibition of NEIL-1 should have an effect on the PFGE data. Since the datasets we used to fit our model used different doses we can't rule out that complexity of DSBs alters the dynamics of repair. However we have demonstrated that our model is predictive across multiple doses in the low LET regime.

In addition to the DNA damage spectrum associated with the complexity of the break region, clustered DSBs can also be prevalent [78] and similarly to the dependence of complex DSBs on the

presence of specific recruitment proteins, DSBs that reside in heterochromatic regions have been shown to require Artemis [79]. Clearly, methods of single DSB measurement would be directly useful in understanding repair of heterogenous populations of DSBs, such as the tracking and formation of γ-H2AX foci, however it has been shown that the normalised rate of repair resulting from physical methods of DSB quantification is different to the rate of repair in gamma-H2AX where there is a lag in the time for maximal DSBs to be recorded by foci. This could be because γ-H2AX is involved in the recruitment of repair complexes and because H2AX is present in clusters in bulk chromatin, the spreading of γ-H2AX is assumed to be non uniform [80]. Previous studies have modelled the formation of foci by summing the DSB enzymes involved in the phosphorylation of H2AX [53, 57] and using this approach, our model could be developed to take into account lower doses of radiation.

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

[2] Freifelder, D. (1966) Lethal Changes in Bacteriophage DNA Produced by X-Rays. *Radiation Research Supplement,* **6**, pp. 80–96.

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

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

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

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

[7] Moore, J. K. and Haber, J. E. (1996) 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.

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

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

[10] Rathmell, W. K. and Chu, G. (1994) 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.

[11] Liang, F., Romanienko, P. J., Weaver, D. T., Jeggo, P. A., and Jasin, M. (1996) Chromosomal double-strand break repair in Ku80-deficient cells. *Proceedings of the National Academy of Sciences,* **93**(17), 8929–8933.

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

[13] Ma, Y., Pannicke, U., Schwarz, K., and Lieber, M. R. (2002) 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.

[14] Grawunder, U., Wilm, M., Wu, X., Kulesza, P., Wilson, T. E., Mann, M., and Lieber, M. R. Activity of DNA ligase IV stimulated by complex formation with XRCC4 protein in mammalian cells. *Nature,* **388**(6641), 492–495.

[15] DiBiase, S. J., Zeng, Z.-C., Chen, R., Hyslop, T., Curran, W. J., and Iliakis, G. (2000) DNA-dependent Protein Kinase Stimulates an Independently Active, Nonhomologous, End-Joining Apparatus. *Cancer Research,* **60**(5), 1245–1253.

[16] Beucher, A., Birraux, J., Tchouandong, L., Barton, O., Shibata, A., Conrad, S., Goodarzi, A. A., Krempler, A., Jeggo, P. A., and Löbrich, M. (2009) ATM and Artemis promote homologous recombination of radiation-induced DNA double-strand breaks in G2. *The EMBO Journal,* **28**(21), 3413–3427.

[17] Ochi, T., Blackford, A. N., Coates, J., Jhujh, S., Mehmood, S., Tamura, N., Travers, J., Wu, Q., Draviam, V. M., Robinson, C. V., Blundell, T. L., and Jackson, S. P. (2015) PAXX, a paralog of XRCC4 and XLF, interacts with Ku to promote DNA double-strand break repair. *Science,* **347**(6218), 185–188.

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

[19] Lin, F. L., Sperle, K., and Sternberg, N. (1984) 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.

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

[21] Deng, S. K., Gibb, B., de Almeida, M. J., Greene, E. C., and Symington, L. S. (2014) RPA antagonizes microhomology-mediated repair of DNA double-strand breaks. *Nat Struct Mol Biol,* **21**(4), 405–412.

[22] Xiong, X., Du, Z., Wang, Y., Feng, Z., Fan, P., Yan, C., Willers, H., and Zhang, J. (2015) 53BP1 promotes microhomology-mediated end-joining in G1-phase cells. *Nucleic Acids Research,* **43**(3), 1659–1670.

[23] Mansour, W. Y., Rhein, T., and Dahm-Daphi, J. (2010) 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.

[24] Wang, M., Wu, W., Wu, W., Rosidi, B., Zhang, L., Wang, H., and Iliakis, G. (2006) PARP-1 and Ku compete for repair of DNA double strand breaks by distinct NHEJ pathways. *Nucleic Acids Research,* **34**(21), 6170–6182.

[25] Schipler, A. and Iliakis, G. (2013) 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.

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

[27] Branzei, D. and Foiani, M. (2008) Regulation of DNA repair throughout the cell cycle. *Nat Rev Mol Cell Biol,* **9**(4), 297–308.

[28] Karanam, K., Kafri, R., Loewer, A., and Lahav, G. (2012) 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.

[29] Hentges, P., Waller, H., Reis, C. C., Ferreira, M. G., and Doherty, A. J. (2015) Cdk1 Restrains NHEJ through Phosphorylation of XRCC4-like Factor Xlf1. *Cell Reports,* **9**(6), 2011–2017.

[30] Trujillo, K. M., Yuan, S.-S. F., Lee, E. Y.-H. P., and Sung, P. (1998) Nuclease Activities in a Complex of Human Recombination and DNA Repair Factors Rad50, Mre11, and p95. *Journal of Biological Chemistry,* **273**(34), 21447–21450.

[31] Deng, X., Prakash, A., Dhar, K., Baia, G. S., Kolar, C., Oakley, G. G., and Borgstahl, G. E. O. (2009) Human Replication Protein A Rad52 Single Stranded DNA Complex: Stoichiometry and Evidence for Strand Transfer Regulation by Phosphorylation. *Biochemistry,* **48**(28), 6633–6643 PMID: 19530647.

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

[33] Götlich, B., Reichenberger, S., Feldmann, E., and Pfeiffer, P. (1998) Rejoining of DNA double-strand breaks in vitro by single-strand annealing. *European Journal of Biochemistry,* **258**(2), 387–395.

[34] Wang, H., Zeng, Z.-C., Bui, T.-A., Sonoda, E., Takata, M., Takeda, S., and Iliakis, G. (2001) Efficient rejoining of radiation-induced DNA double-strand breaks in vertebrate cells deficient in genes of the RAD52 epistasis group. *Oncogene,* **20**(18).

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

[36] Wang, H., Perrault, A. R., Takeda, Y., Qin, W., Wang, H., and Iliakis, G. (2003) Biochemical evidence for Ku-independent backup pathways of NHEJ. *Nucleic Acids Research,* **31**(18), 5377–5388.

[37] Mladenov, E., Magin, S., Soni, A., and Iliakis, G. (2013) DNA Double-Strand Break Repair as determinant of cellular radiosensitivity to killing and target in radiation therapy. *Frontiers in Oncology,* **3**(113).

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

[39] Decottignies, A. (2013) Alternative end-joining mechanisms: a historical perspective. *Frontiers in Genetics,* **4**(48).

[40] Ma, J.-L., Kim, E. M., Haber, J. E., and Lee, S. E. (2003) 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.

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

[42] Audebert, M., Salles, B., and Calsou, P. (2004) 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.

[43] Liang, L., Deng, L., Nguyen, S. C., Zhao, X., Maulion, C. D., Shao, C., and Tischfield, J. A. (2008) 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.

[44] Soni, A., Siemann, M., Grabos, M., Murmann, T., Pantelias, G. E., and Iliakis, G. (2014) 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.

[45] Howard, S. M., Yanez, D. A., and Stark, J. M. (2015) DNA Damage Response Factors from Diverse Pathways, Including DNA Crosslink Repair, Mediate Alternative End Joining. *PLoS Genet,* **11**(1), e1004943.

[46] Taleei, R., Girard, P. M., and Nikjoo, H. (2015) DSB repair model for mammalian cells in early S and G1 phases of the cell cycle: Application to damage induced by ionizing radiation of different quality. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis,* **779**, 5 – 14.

[47] Taleei, R. and Nikjoo, H. (2014) Response to the Letter of Bodgi and Foray: On the coherence between mathematical models of DSB repair and physiological reality. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis,* **761**, 50 – 52.

[48] Bodgi, L. and Foray, N. (2014) On the coherence between mathematical models of DSB repair and physiological reality. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis,* **761**, 48 – 49.

[49] Taleei, R., Girard, P. M., Sankaranarayanan, K., and Nikjoo, H. (2015/11/02, 2013) The Non-homologous End-Joining (NHEJ) Mathematical Model for the Repair of Double-Strand Breaks: II. Application to Damage Induced by Ultrasoft X Rays and Low-Energy Electrons. *Radiation Research,* **179**(5), 540–548.

[50] Nikjoo, H., Uehara, S., Emfietzoglou, D., and Cucinotta, F. (2006) Track-structure codes in radiation research. *Radiation Measurements,* **41**(9-10), 1052 – 1074 Space Radiation Transport, Shielding, and Risk Assessment Models.

[51] L, M. and G, I. (1991) 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).

[52] Verbruggen, P., Heinemann, T., Manders, E., von Bornstaedt, G., van Driel, R., and Höfer, T. (2014) Robustness of DNA Repair through Collective Rate Control. *PLoS Comput Biol,* **10**(1), e1003438.

[53] Cucinotta, F. A., Pluth, J. M., Anderson, J. A., Harper, J. V., and O'Neill, P. (2008) Biochemical Kinetics Model of DSB Repair and Induction of gamma-H2AX Foci by Non-homologous End Joining. *Radiation Research,* **169**(2), 214–222.

[54] Taleei, R., Weinfeld, M., and Nikjoo, H. (2012) Single strand annealing mathematical model for double strand break repair. *journal of Molecular Engineering and Systems Biology,* **1**(1).

[55] Taleei, R. and Nikjoo, H. (2013) 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.

[56] Taleei, R., Weinfeld, M., and Nikjoo, H. (2011) A kinetic model of single-strand annealing for the repair of DNA double-strand breaks. *Radiation Protection Dosimetry,* **143**(2-4), 191–195.

[57] Belov, O. V., Krasavin, E. A., Lyashko, M. S., Batmunkh, M., and Sweilam, N. H. (2015) A quantitative model of the major pathways for radiation-induced DNA double-strand break repair. *Journal of Theoretical Biology,* **366**(0), 115 – 130.

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

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

[60] Li, Y., Qian, H., Wang, Y., and Cucinotta, F. A. (2012) A Stochastic Model of DNA Fragments Rejoining. *PLoS ONE,* **7**(9), e44293.

[61] Wu, W., Wang, M., Wu, W., Singh, S. K., Mussfeldt, T., and Iliakis, G. (2008) Repair of radiation induced DNA double strand breaks by backup NHEJ is enhanced in G2. *DNA Repair,* **7**(2), 329–338.

[62] Decottignies, A. (2013) Alternative end-joining mechanisms: a historical perspective. *Frontiers in Genetics,* **4**(48).

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

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

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

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

[67] Stryhn, H. and Christensen, J. (2014) The analysis–Hierarchical models: Past, present and future. *Preventive Veterinary Medicine,* **113**(3), 304 – 312.

[68] Toni, T., Welch, D., Strelkowa, N., Ipsen, A., and Stumpf, M. P. H. (2009) 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.

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

[70] Liepe, J., Barnes, C., Cule, E., Erguler, K., Kirk, P., Toni, T., and Stumpf, M. P. (2010) ABC-SysBio–approximate Bayesian computation in Python with GPU support. *Bioinformatics,* **26**(14), 1797–1799.

[71] Liepe, J., Kirk, P., Filippi, S., Toni, T., Barnes, C. P., and Stumpf, M. P. H. (2014) A framework for parameter estimation and model selection from experimental data in systems biology using approximate Bayesian computation. *Nature protocols,* **9**(2), 439–456.

[72] D., S. K., Paul, M., and Darryl, S. (2008) Modeling DNA Methylation in a Population of Cancer Cells. *Statistical Applications in Genetics and Molecular Biology,* **7**(1), 1–23.

[73] Haince, J.-F., McDonald, D., Rodrigue, A., Déry, U., Masson, J.-Y., Hendzel, M. J., and Poirier, G. G. (2008) PARP1-dependent Kinetics of Recruitment of MRE11 and NBS1 Proteins to Multiple DNA Damage Sites. *Journal of Biological Chemistry,* **283**(2), 1197–1208.

[74] Shibata, A., Conrad, S., Birraux, J., Geuting, V., Barton, O., Ismail, A., Kakarougkas, A., Meek, K., Taucher-Scholz, G., Löbrich, M., and Jeggo, P. A. (2011) Factors determining DNA double-strand break repair pathway choice in G2 phase. *The EMBO Journal,* **30**(6), 1079–1092.

[75] Ceccaldi, R., Liu, J. C., Amunugama, R., Hajdu, I., Primack, B., Petalcorin, M. I. R., O'Connor, K. W., Konstantinopoulos, P. A., Elledge, S. J., Boulton, S. J., Yusufzai, T., and D'Andrea, A. D. (2015) Homologous-recombination-deficient tumours are dependent on Polθ-mediated repair. *Nature,* **518**(7538), 258–262.

[76] Cho, N. W. and Greenberg, R. A. (2015) DNA repair: Familiar ends with alternative endings. *Nature,* **518**(7538), 174–176.

[77] Dobbs, T. A., Palmer, P., Maniou, Z., Lomax, M. E., and O'Neill, P. (2008) Interplay of two major repair pathways in the processing of complex double-strand DNA breaks. *DNA Repair,* **7**(8), 1372 – 1383.

[78] Sutherland, B. M., Bennett, P. V., Sidorkina, O., , and Laval, J. (2000) Clustered Damages and Total Lesions Induced in DNA by Ionizing Radiation:âĂĽ Oxidized Bases and Strand BreaksâĂă. *Biochemistry,* **39**(27), 8026–8031.

[79] Woodbine, L., Brunton, H., Goodarzi, A. A., Shibata, A., and Jeggo, P. A. (2011) Endogenously induced DNA double strand breaks arise in heterochromatic DNA regions and require ataxia telangiectasia mutated and Artemis for their repair. *Nucleic Acids Research,* **39**(16), 6986–6997.

[80] Kinner, A., Wu, W., Staudt, C., and Iliakis, G. (2008) γ-H2AX in recognition and signaling of DNA double-strand breaks in the context of chromatin. *Nucleic Acids Research,* **36**(17), 5678–5694.

7 Figure Legends

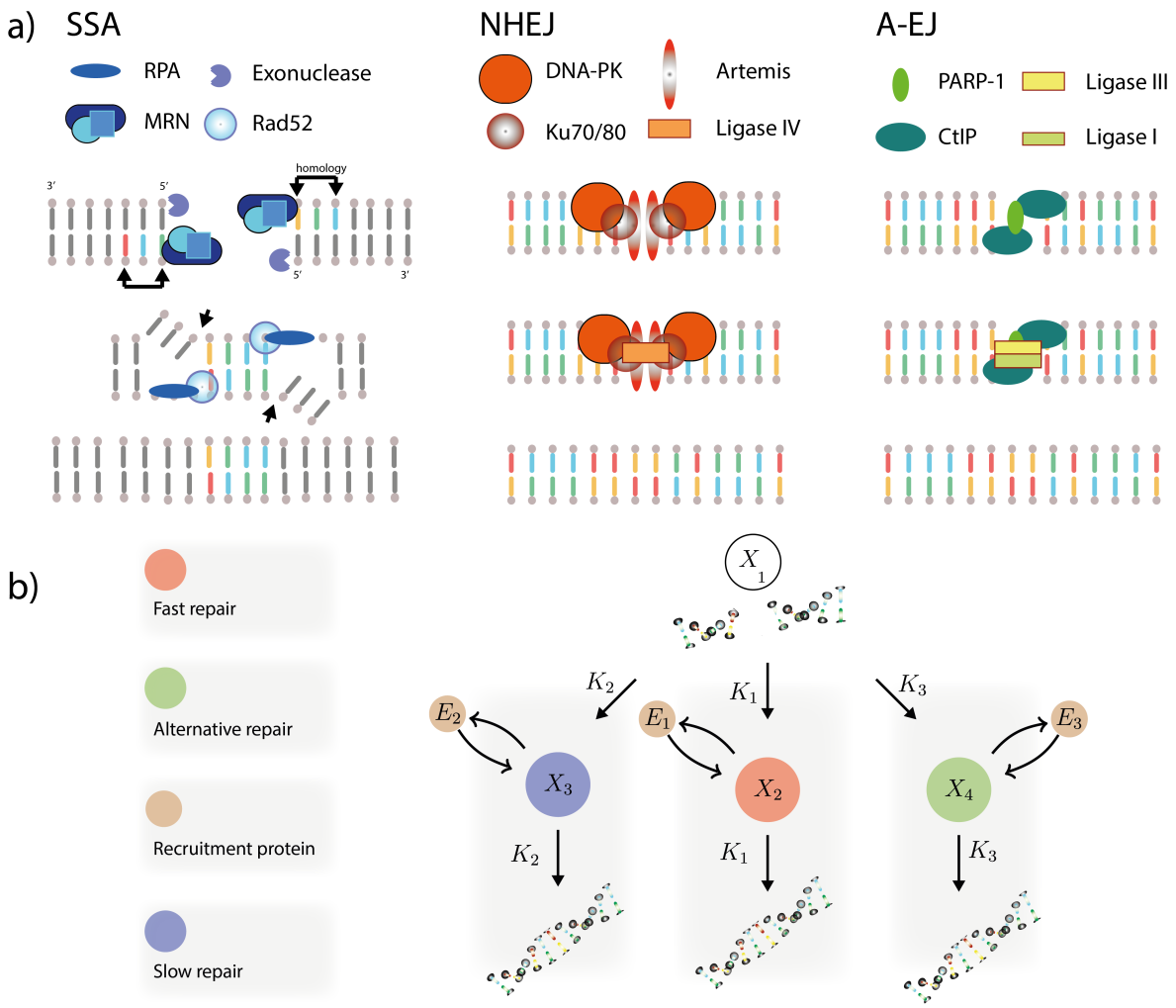

Figure 1: Modelling multiple repair mechanisms. a) Proteins and repair steps contributing to repair during SSA, NHEJ and A-EJ in mammalian cells (illustration). b) The model. Discs represent species and arrows represent reactions.

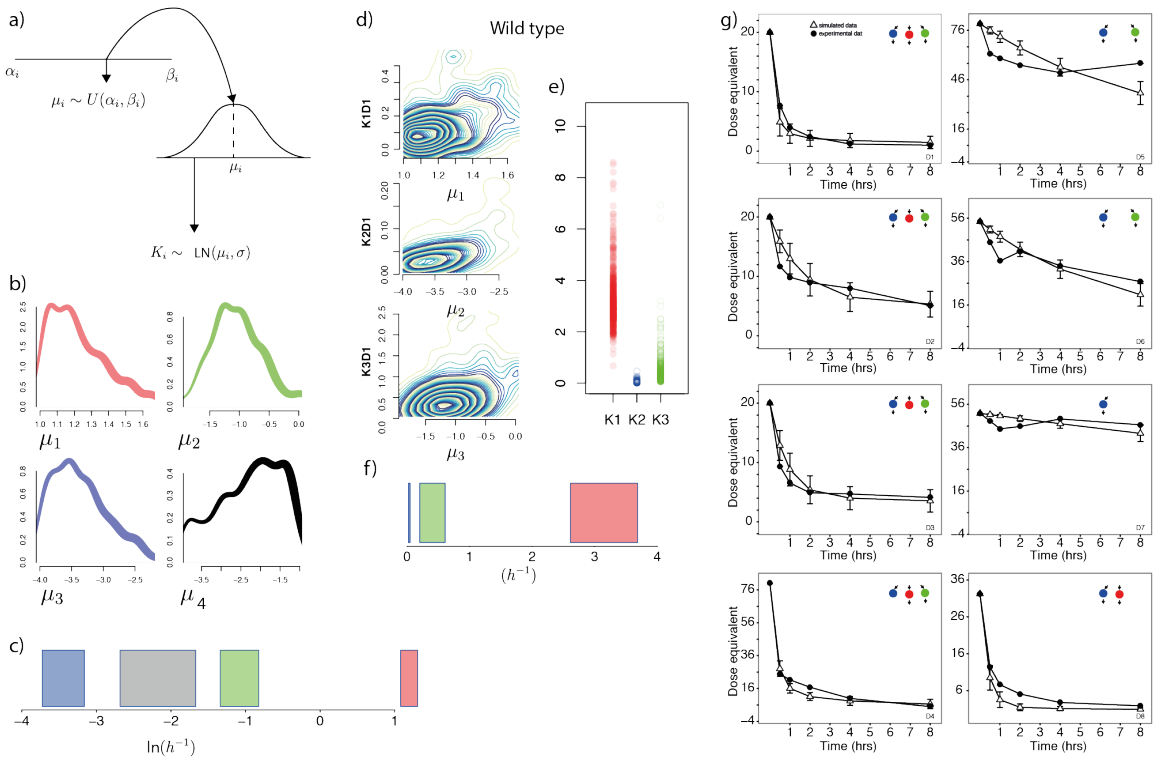

Figure 2: a) Diagram showing the parameter sampling process. Hyper parameters μ_i are drawn from a uniform distribution between α_i and β_i. Model parameters K_i are sampled from a lognormal distribution with mean μ_i. b) Posterior distributions for the hyper parameters μ_{1-3} and μ_4. c) Box plot showing 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). f) Interquartile range of the parameters $K1D1 - K3D1$. g) Time series plots of the experimental data and model simulation. Sub-figures on the top right represent the active repair mechanisms. Red, blue and green represent fast, slow and alternative repair.

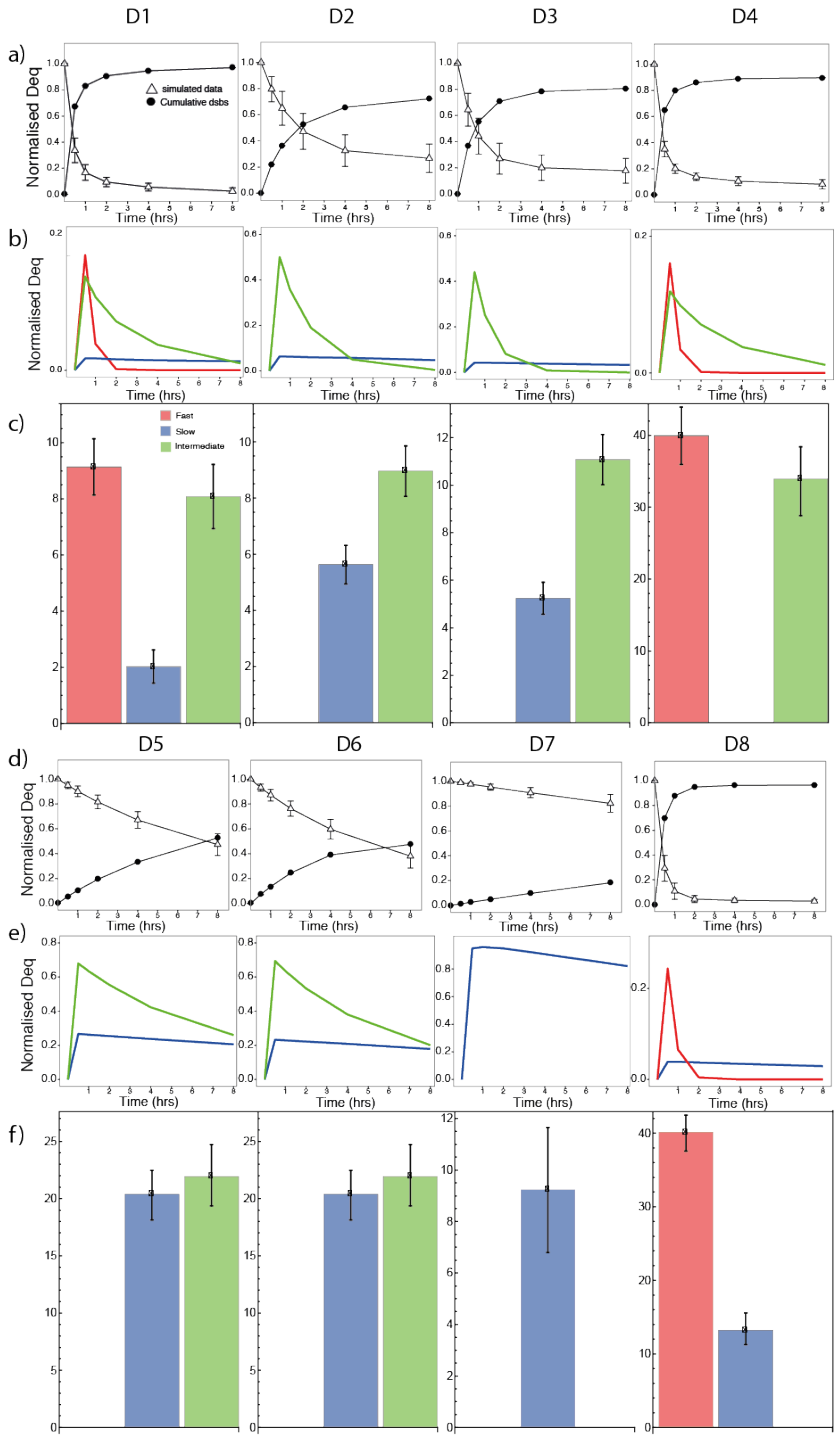

Figure 3: Identifying fast, slow and alternative mechanisms. a) Simulated data and Cumulative DSBs. b) DSBs entering each repair mechanism. c) Total amount of DSBs repaired by fast, slow and alternative repair. a-c) Results shown for datasets D1-D4. d) Simulated data and Cumulative DSBs. e) Time series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

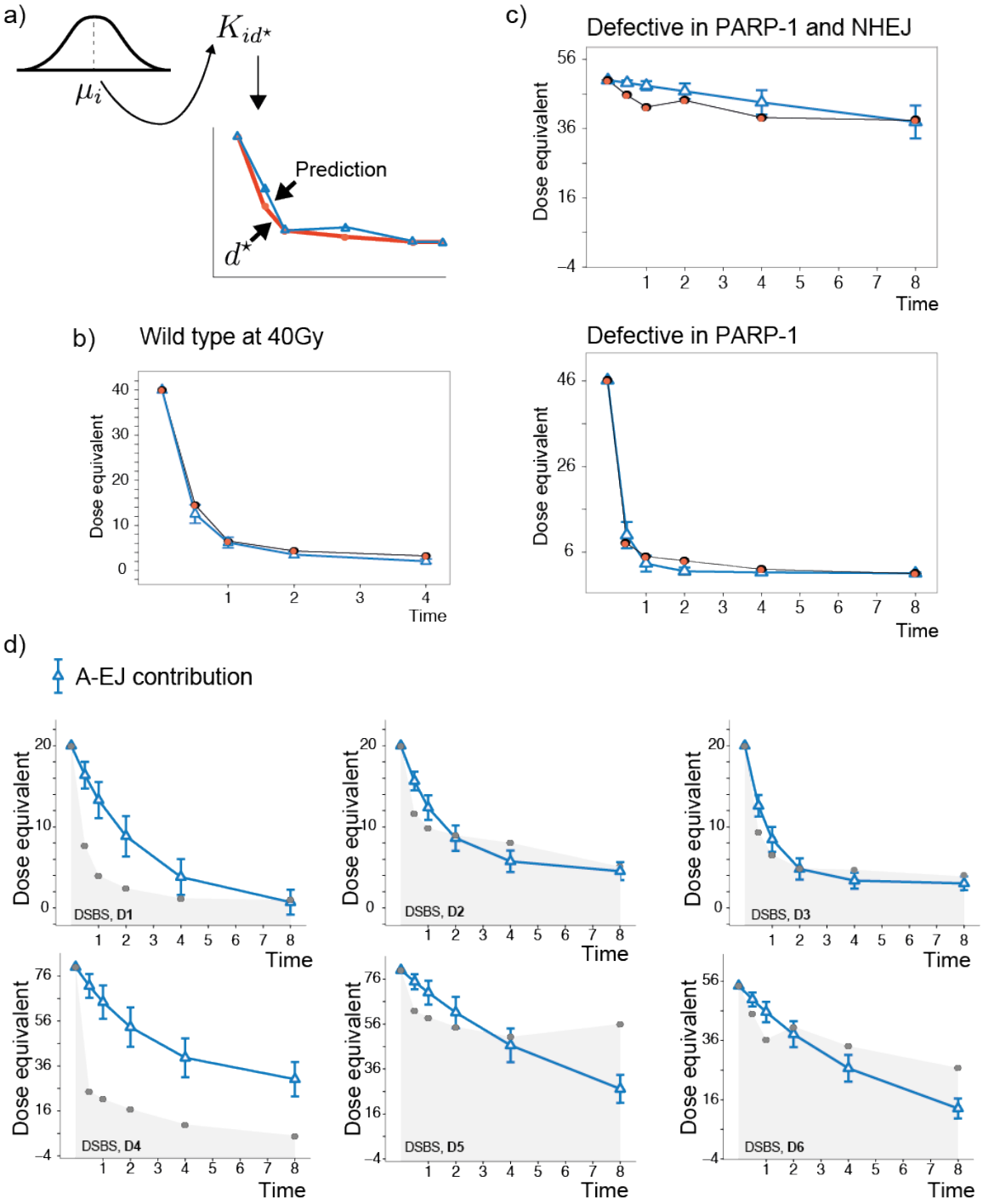

Figure 4: Model predictions. a). Parameters were drawn from the hyper parameter distribution and the code was simulated again under different initial conditions (blue curve) and compared to different datasets (red curve) b). Model prediction (blue curve) of wild type repair (red curve) from our initial parameter fit (Figure 2g). c). PARP-1 inhibited repair predictions. The initial parameter fit (blue curves) provided a good fit to the experimental data. d). By setting the parameters of NHEJ and SSA to zero in our predicted posterior, we were able to simulate the expected contribution of A-EJ in our datasets where A-EJ is assumed to be active. Number of DSBs remaining in each of the datasets are shown in the grey shaded region.

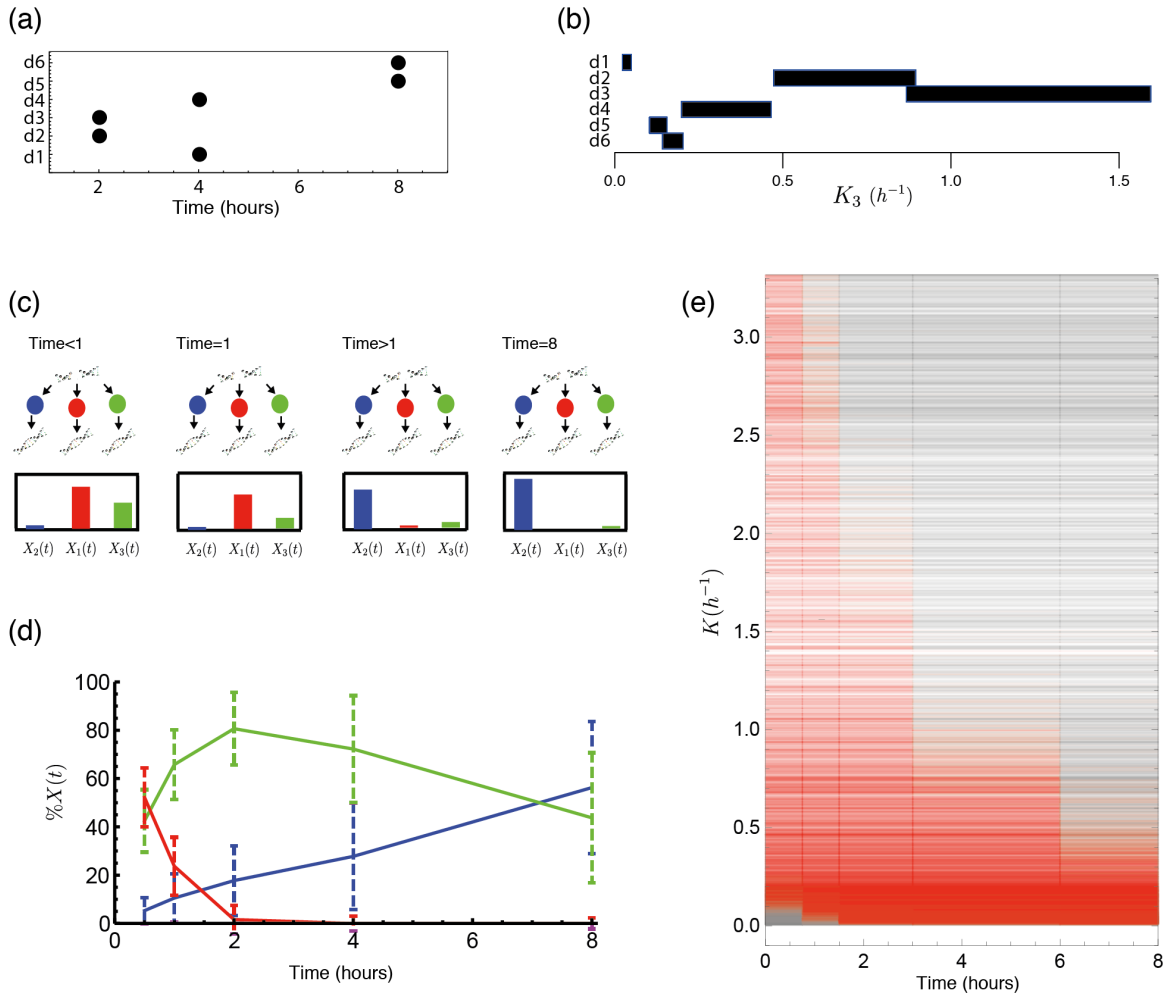

Figure 5: a). Time in which a repair curve has reached below half it's maximum value for each data set in which A-EJ is assumed to be active. The slowest mode of repair occurred in data sets 5 and 6, where Ku70 is inactive. b). Rectangle plot of the interquartile ranges of K_3 for all datasets where 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. Grey indicates that the mechanism is less than 30% active and red indicates the mechanism is greater than 30% active.