Dynamical processes in Ageing, Gene regulation & Communication

Ph.D. Thesis

Kristian Moss Bendtsen

Acknowledgements
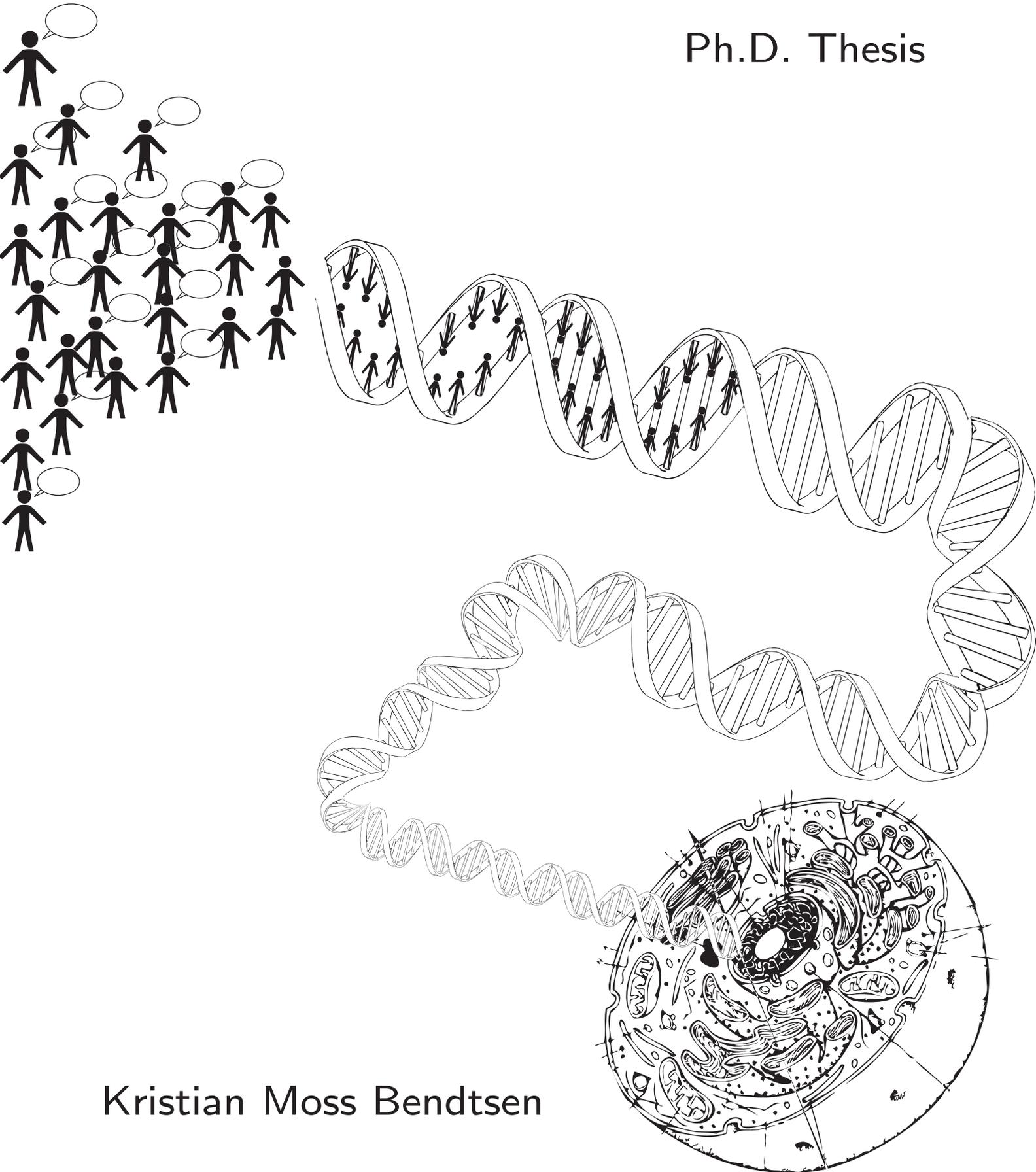
This work was primarily performed at Center for Models of Life (CMOL) at the Niels Bohr Institute. These past four years have been nothing short of amazing, and a primary reason is the working environment at CMOL. I would like to thank **Kim Sneppen** for intellectual discussions during these four years and for strongly contributing to the very inspiring atmosphere at CMOL. During my time at CMOL I have collaborated with several people. Among those I would like to acknowledge **Szabolcs Semsey**, who I have collaborated with on two projects and without his biological insight they would have been very meaningless. I would also like to thank **Jeppe Juul** my former officemate, for enjoyable hours spend working and travelling together. I am very grateful for the scientific inputs but especially for the encouragement **Ala Trusina** gave me during the rough patches of my Ph.D. I want to thank **Jan Haerter** and **Florian Uekermann** for the scientific discussions we had which were extremely stimulating. **Sandeep Krishna** has been a great host during my visits to Bangalore, and I know I have learned at lot from his intuitive explanations to complex questions. Lastly I would like to thank my supervisor **Mogens Høgh Jensen**, who gave me the chance to work at CMOL. Mogens has throughout my Ph.D. allowed me the freedom to do research on my own but at the same time offered valuable guidance when I needed it. I am very thankful to Mogens for valuing the well-being of his students above everything else.

Contents

Publications

Published

1. KM Bendtsen, J Erdössy, Z Csiszovszki, SL Svenningsen, K Sneppen, S Krishna and S Semsey
   Direct and indirect effects in the regulation of overlapping promoters
   Nucleic Acid Research vol. 39(16), page 6879–6885, **2011**

2. KM Bendtsen, J Juul and A Trusina:
   Fragile DNA Repair Mechanism Reduces Ageing in Multicellular Model
   PloS One vol. 12 (5), page e. 36018, **2012**

3. S Holbek, KM Bendtsen and J Juul
   Moderate Stem-cell telomere shortening rate postpones cancer onset in a stochastic model
   Physical Review E vol. 88(4), **2013**

4. C Strandkvist, J Juul and KM Bendtsen
   Asymmetric Segregation of Damaged Cellular Components in Spatially Structured Multicellular Organisms
   PloS one 9 (2), **2014**

5. KM Bendtsen, MB Jensen, A May, L Rasmussen, A Trusina, V Bohr and MH Jensen
   Dynamics of the DNA repair proteins WRN and BLM in the nucleoplasm and nucleoli
   European Biophysics Journal, 1-8 **2014**

Submitted

6. KM Bendtsen, MH Jensen, S Krishna and S Semsey:
   The Role of mRNA and protein stability in the function of coupled positive and negative feedback systems in eukaryotic cells
   Submitted to *Scientific reports*

In preparation

7. KM Bendtsen, F Uekermann , J Haerter:
   The Expert Game - Spontaneous emergence of trust and social capital

Abbreviations

FRAP	Flourescent Recovery After Photobleaching
GFP	Green Fluorescent Protein
DSB	Double Stranded Breaks
WRN	Werner Syndrome helicase
BLM	Bloom Syndrome helicase
R-D	Reaction-diffusion
AIC	Akaike Information Criteria
BIC	Bayesian Information Criteria
ID	Identities kept between games
NO-ID	Identities randomized between games
SC	Social Capital
RNAP	RNA polymerase
ROS	Reactive oxygen species
FBM	Frustrated bistability motif
NAR	Negative autoregulation
PAR	Positive autoregulation
NAF	Negative autoregulated Frustated bistability

Abstract

My thesis consists of three parts. The first part covers ageing phenomena. In the first project I measured the mobility of two DNA repair proteins. Contrasting diffusion coefficients from literature I was able to classify DNA repair protein into either "scanners" or "responders". In a second project we constructed a mathematical model and showed that if DNA damage is primarily caused by geno-toxic agents, it would be advantageous for cells to have a fragile DNA repair mechanism.

The second part of my Ph.D. thesis covers gene regulation. In the first project we show how RNA polymerase can be used as a transcription factor. This requires that promoter regions overlap, which 15% of promoters in E.coli do. In the second project I analyse a negative auto regulated transcription motif coupled to a positive auto regulation transcription motif. I find that a general feature of this motif is that unstable activation and stable repression is a requirement for the motif to produce oscillations.

The last part of this thesis studies the emergence of communication networks. In this study we constructed a simple e-mail game. E-mails from two session with 16 players, who had never met before, showed how players develop favourite communication partners. We observed how this dynamic caused a communication network to form. By quantifying the information flow in this network, we were able to shown how that the network functions as an anti-exploration mechanism against "information leeches".

Resumé

Min afhandling består af tre dele. Den første del omhandler aldringsfænomener. Her har jeg målt mobiliteten af to DNA reparationsproteiner. Ved at sammenligne diffusionsko-efficienter fra tidligere studier, var det muligt at klassificere DNA reparationsproteiner som enten "scanners" eller "responders". I et andet projekt lavede vi en matematisk model, og viste at hvis DNA skade primært stammer fra omgivelserne, kan det være en fordel for celler at have en skrøbelig DNA reparationsmekanisme.

Den anden del af min Ph.D. omhandler gen-regulering. I det første af mine projekter viste vi hvordan RNA polymerase kan fungere som en transskriptionsfaktor. Et krav er at promoterområder overlapper med hinanden, hvilket 15 % af promotorne i E.coli gør. Mit andet projekt omhandler et negative selvregulerende transskriptionsmotiv, koblet til et positivt selvregulerende transskriptionsmotiv. Jeg fandt at et generelt krav for at oscillationer opstår, er at motivet har ustabile aktivatorer og stabile repressorer.

Den sidste del af min Ph.D. beskæftiger sig med, hvordan kommunikationsnetværk opstår. Vi opfandt et simpelt e-mail spil for at undersøge opbyggelsen af kommunika-tionsnetværk i tid. E-mails fra to forsøg med 16 personer, der aldrig havde mødt hinanden før, viste hvordan spillerne udviklede foretrukne kommunikationspartnere. Vi observerede hvordan denne dynamik med tiden fik et stabilt kommunikationsnetvær-ket til at opstå. Ved at kvantificere mængden af information på netværket, viste vi at netværket var et redskab til at undgå informationssnyltere.

Ageing

1.1 Introduction to ageing

The ageing process in humans is the gradual deterioration of physiological functions. The physiological function can be thought of as an emergent phenomenon originating from the interactions between multiple cells. It is believed that cellular senescence promotes ageing in humans [1, 2]. Senescent cells are cells which, after several divisions, have lost their ability to further proliferate [3] (see Figure 1.1). Mammalian cells have a finite number of cell divisions, roughly 50 *in vitro*; this limit is called the Hayflick limit [4]. The molecular reason for replicative senescence is that the chromosome ends, named telomeres, shorten after each division. When the telomeres have reached a critical length a DNA damage response is triggered causing permanent cell cycle arrest (senescence) [2].
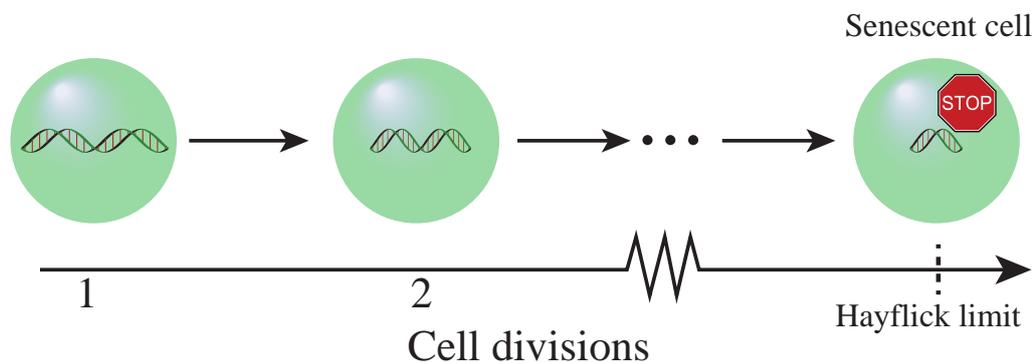
Figure 1.1: **Hayflick limit:** The chromosome ends are called telomeres. After each cell division the telomere lengths are reduced. When a cell has gone through roughly 50 division, cell proliferation is inhibited. The inability of the cells to further proliferate is called senescence. The number of divisions, that a cell can undergo before it becomes senescent is called the Hayflick limit.

It has been shown that senescent cells accumulate in humans with age [3]; for primates the accumulation has been shown to be exponential [5]. The fraction of senescent cells in old tissue has been estimated to be 1-15% [3]. Thus, as tissues have a fairly constant number of cells, the increased fraction of senescent cells in the tissues compromises the tissue's ability to renew and repair regions of damage. Another way senescent cells might promote ageing is through the secretion of proteins, which alters the micro environment of the tissue[3].

A 2007 study by Xue et al. has shown that the immune system removes senescent cells [6]. It is still not known why senescent cells accumulate. Is it because the immune system clearance becomes less effective? Or because the rate of cells becoming senescent

increases? [3].

Besides cellular senescence induced by the shortening of telomeres, premature senescence can be triggered by DNA damage [2].
Both mitochondrial DNA (mtDNA) and nuclear DNA have been shown to be linked to the ageing process [7, 8]. DNA damage can be induced either by intrinsic factors (e.g., reactive oxygen species (ROS)) or by external factors (e.g., exposure to UV-light or toxic chemicals) [9]. ROS is primarily produced in the mitochondria as a by-product of ATP synthesis [2]. Damaged and dysfunctional mitochondria have been thought to create a vicious circle where dysfunctional mitochondria create more ROS, which creates more damage both to DNA and mtDNA, which again creates more dysfunction [7].

Yet another candidate for promoting ageing is the decline of stem cell function [10]. The capacity for regeneration and renewal of tissue is dependent on the population of somatic stem cells. Somatic stem cells have activated telomerase, a protein which elongates the telomeres [11]. However, the level of telomerase is not high enough to completely eliminate telomere shortening [12], and thus somatic stem cells have larger - although still limited - proliferation capacity [13, 14].

In order to reduce ROS and thereby DNA damage somatic stem cells are in a low-activity quiescent state. The low-activity state minimizes the use of ATP and thereby reduces the production of ROS [15]. Reactive oxygen species production is further reduced in quiescent stem cells by synthesizing ATP in an anearobic fashion [15]. The downside of the anearobic ATP synthesis is that it is less efficient than the aerobic/mitochondria pathway. When tissue renewal is needed to maintain the tissue, stem cells transition from the quiescent state into the proliferating state, thereby allowing renewal; but at the same time, however increasing the risk of acquiring DNA damage is increased [16, 15] (see Figure1.2).

Accumulation of senescent cells and decline of stem cells are probably not the only mechanisms that contribute to ageing. Neither of the two mechanisms seem to account for ageing in a slow turnover tissue. The turnover rate of a tissue is a measure of how fast a tissue renews its cells. For example, heart cells (cardiomyocytes) are replaced 11 times during a human life (100 years), giving the heart cells a turnover rate of roughly 10 years [17]. In neurons, which are also a slow turnover tissue, it seems that accumulation of damage to either nuclear or mitochondrial DNA causes functional decline of the brain [18]. A proliferating cell with persistent DNA damage is programmed to go apoptotic. However, neurons and other slow turnover cells seem to simply silence the gene affected by the DNA damage [7]. Thus, for slow turnover tissue, the idea seems to be that it is better to have less functional cells than fewer cells.

A commonality between the three causes of ageing mentioned above (namely senescence, stem cell depletion and gene silencing) is DNA damage. Damage to both mtDNA and nuclear DNA plays a major role in the process of ageing. Further evidence of DNA damages as a cause of ageing is that several diseases where DNA repair is compromised cause premature ageing. Such diseases include Werner syndrome, ataxia telangiectasia, and Bloom syndrome [19, 20]. In Figure 1.3 I have summarized some of the causes of ageing and the links between them. However, ageing is a very complex phenomenon, and the simple schematic presented here does not capture all of the causes of ageing.

In the following chapters I present three projects in which we have investigated

Quiescent stem cell   Proliferating stem cell
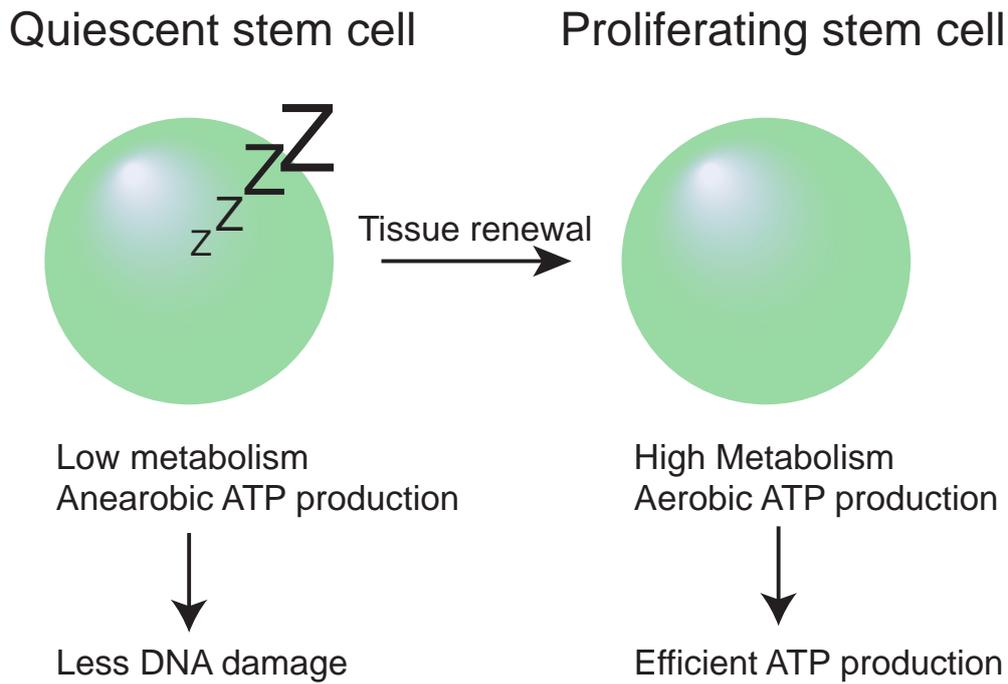
Figure 1.2: **Quiescent stem cells:** In order to reduce the level of DNA damaging agents, stem cells are in a quiescent state. Both the reduced need for ATP synthesis and using a anaerobic pathway reduce the production of ROS. When tissue renewal is needed the stem cells transition from the quiescent state into the proliferating state. The proliferating state requires a higher metabolism and therefore increases the risk of acquiring DNA damage.

parts of the ageing process either via phenomenological models or through mechanistic models and biological experiments.
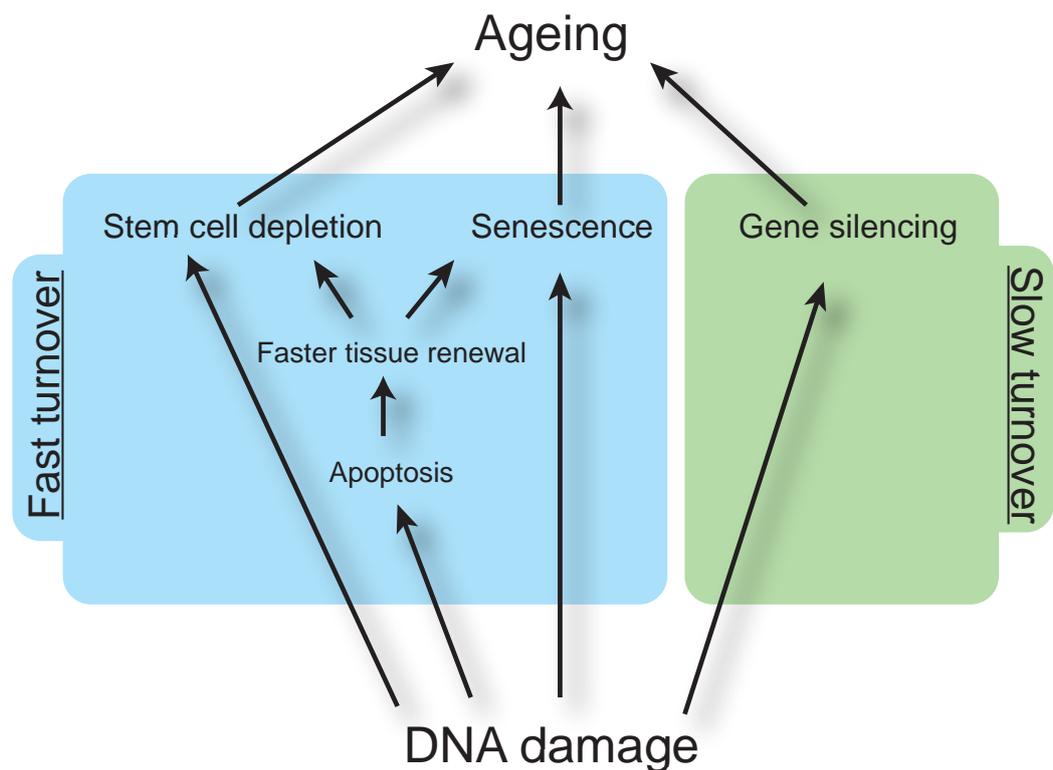
Figure 1.3: **Schematic summary of ageing:** DNA damage can cause cells to go into damage-induced senescence. Increased levels of DNA will result in more cells going apoptotic, which will require either stem cells or tissue cells to proliferate. The increased demand for divisions will result in either a faster deterioration of the stem cell pool or increased number of senescent cells. For slow turnover tissue (e.g., neurons) DNA damage induces gene silencing and thereby less functional cells.

1.2 Possible advantages of fragile DNA repair mechanisms

Deleterious mutations to certain DNA repair proteins have severe effects on the expected lifespan. Consider, for example, proteins involved in the base excision repair (BER) pathway. These proteins can be classified into either completely non-essential ones that have no phenotypical effect (no decrease in life expectancy when mutated) or completely essential ones that are embryonically lethal [21]. The lack of gradual worsening in the BER pathway is puzzling and suggests that if complete redundancy is not possible, then the second best "strategy" is no redundancy at all. In addition, research has shown that in several organisms the capacity of DNA repair declines with age [2] and as a consequence the mutation frequency increases [22]. This decline in DNA repair capacity could be caused by damage to the DNA repair genes. We propose that a fragile DNA repair mechanism where a decline in repair capacity is drastic instead of gradual could be beneficial for an organism. The reason is that cells with compromised DNA repair will go apoptotic faster than healthy cells.

To investigate this process we constructed a simple mathematical model in which DNA damage leads to mutations that impair the ability to repair future genotoxic damage. Since ageing is a phenomenon that involves a whole tissue, we simulated a population of highly proliferating cells exposed to genotoxic damage. In our model, cells continuously acquire DNA damage. The result of the damage is modelled by three possible outcomes:
a) Repair, b) Apoptosis or c) Mutation (see Figure 1.4). For simplicity we chose not to model senescence or the influx of new cells from the stem cell pool explicitly. However, senescence and influx of new cells are considered implicitly by monitoring the number of cell replications and the total number of cells at any point in time.

1.2.1 Simulation steps

First a random cell out of N acquires DNA damage. After damage there are three possible outcomes (see Figure 1.4):

1. The damage can be fully repaired with probability R. The repair rate R is specific to the individual cell.

2. If the cell is unable to repair the damage, it can go apoptotic with probability a. In this case another cell divides to replace it, keeping the total number of cells constant.

3. Otherwise, the cell accumulates the damage through a mutation in the DNA, which reduces its ability to repair future genotoxic damage from R to $R - \Delta$.

When a cell divides to take the place of a cell that has gone apoptotic, the daughter cell inherits the repair rate from the parent cell (see Figure 1.5). When investigating two- or three-dimensional systems, the cells are located on a square lattice and apoptotic cells can only be replaced by neighbouring cells dividing.

One time step is when N cells have received damage. Therefore on average each cell will acquire one damage per time step, regardless of the system size. All cells are initiated with the same repair rate R_0. The parameter Δ can be interpreted as the fragility of the DNA repair mechanism. If Δ is large, cells with unrepaired DNA damage will have a greatly reduced ability to repair future genotoxic damage (see Figure 1.4).
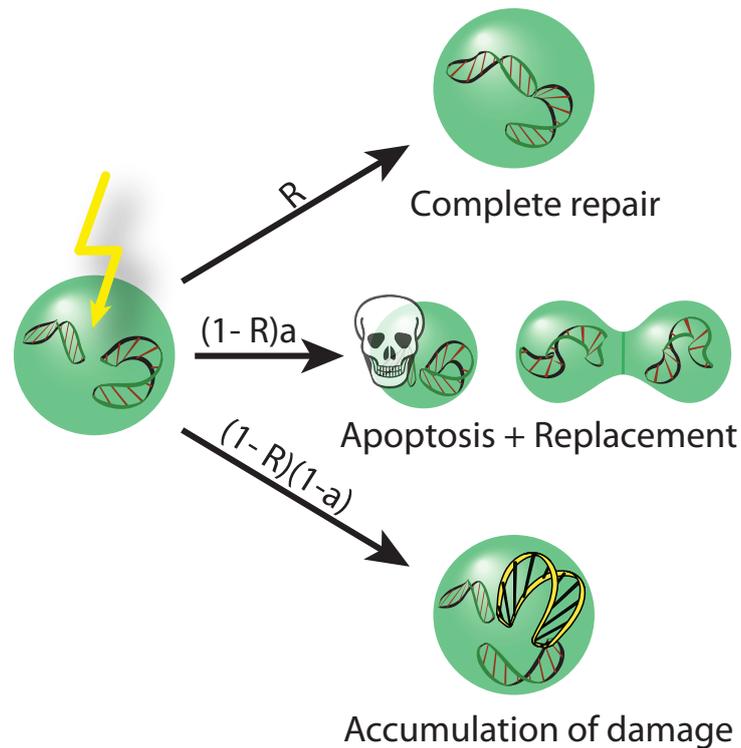
Figure 1.4: **Fragile DNA repair mechanism model:** When a cell acquires DNA damage, there are three possible outcomes: i) Complete repair of the damage with probability R, ii) Apoptosis with probability $(1 - R)a$, in which case another cell divides to keep the number of cells constant; or iii) Accumulation of the damage through a mutation, reducing the rate of repair to $R - \Delta$.
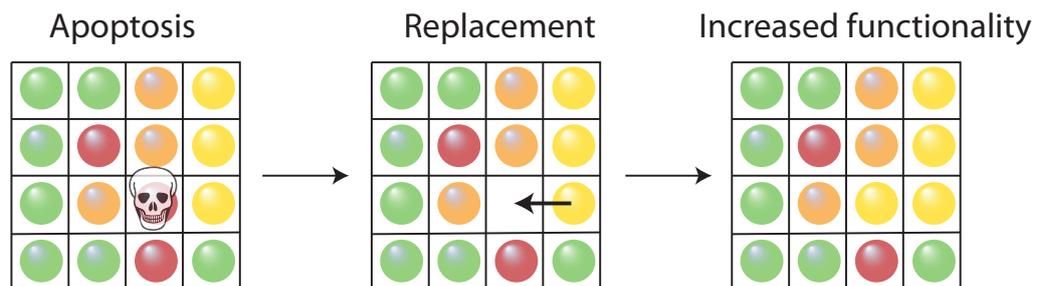
Figure 1.5: **Schematics of the simulation:** In every time step N cells acquire damage. If a cell goes apoptotic, another cell is chosen to replicate. In the well-mixed system any cell can replicate, but in the spatial models only neighbouring cells are allowed to replicate. When a cell gets a mutation, the repair capability of that cell decreases. Cells are not allowed to revert the mutations. The average tissue repair rate is effectively maintained by the fact that the cells with impaired repair capability have a higher probability of going apoptotic. They are then replaced by a random cell, which might have a higher repair rate, thereby effectively increasing the average repair rate of the tissue.

1.2.2   Existence of a temporal steady state

To determine the functionality of the tissue we monitored the average repair rate of the system, $\langle R \rangle$. It starts out at the maximum value R_0 (no cells with mutations) and then decreases as cells accumulate mutations. Since mutations are irreversible, the mean repair rate of each cell will inevitably drop to zero. Surprisingly, a temporary steady state exists where the repair rate of the system fluctuates around an average value of R^* (see Figure 1.6). The temporary steady state is maintained because cells with a low repair rate are more likely to go apoptotic, and when they do they are replaced by cells with a higher repair rate. The system leaves the temporary steady state after a time τ, at which point the average rate of repair decreases to zero very fast.

The biological interpretation of τ is the time a tissue can self-sustain as a functional tissue. After time τ, an introduction of new cells from the stem cell pool is needed to maintain a high rate of repair. To link τ with the ageing of the tissue we need to remember the link between ageing and fast turnover tissue. A higher τ means less introduction of stem cells, which means less metabolic activity of the stem cells, which again means less accumulation of damage in the stem cell pool and consequently fewer mutations [3, 23]. From this rather convoluted argument a large τ corresponds to a slow ageing of the tissue.
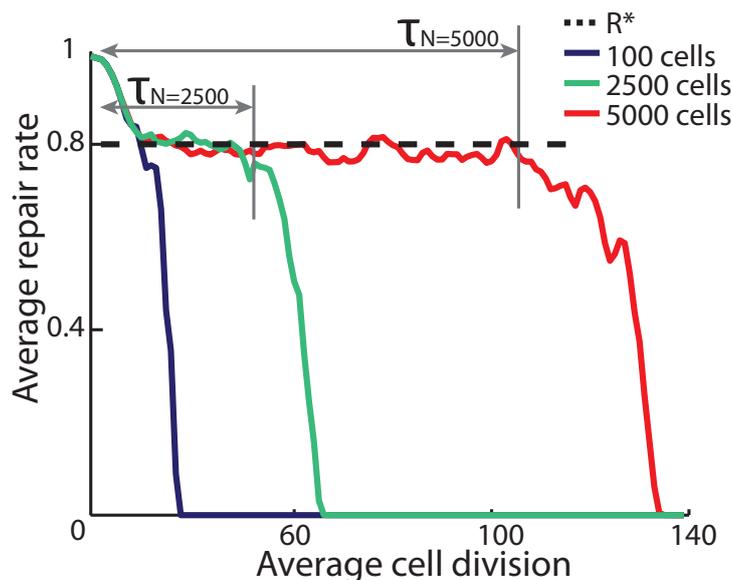
Figure 1.6: **Temporal steady state:** The average repair rate of the system drops as cells irreversibly accumulate mutations. Since cells with a low rate of repair are more likely to go apoptotic, a temporary steady state exists where the repair rate of the system fluctuates around an average value of R^*. The length of time, τ, before the system leaves this temporary state increases drastically with the size N of the system. If τ is large, a high rate of repair may be maintained without renewal from the stem cell pool, and consequently the organism will age more slowly. The simulation is carried out with parameters $R_0 = 0.99$, $a = 0.05$, $\Delta = 0.0056$. The median of 20 simulation runs is shown.

1.2.3   High probability of repair, apoptosis and a fragile mechanism is beneficial

When the initial rate of repair R_0 is high, both the steady state rate of repair R^* and the time spent in this state τ increases (see Figure 1.7 A). Since new, undamaged cells are produced by somatic stem cells, the initial rate of repair can be thought of as the stem cell repair rate. Thus, the repair rate of stem cells is not only important for avoiding DNA damage in the stem cell pool, but also for retaining a high level of function in tissue cells.

Another parameter that we tried varying was the rate of apoptosis, a. An increase in the rate of apoptosis a decreases the risk of accumulating damage and thereby increases both R^* and τ (see Figure 1.7). This is not that surprising since apoptosis is thought of as a fail-safe mechanism to hinder mutation accumulation, which in the worst case leads to cancer [24]. What is surprising, though, is that increasing the rate of apoptosis does not cause more cells in the system to go apoptotic. The increase of the average rate of repair in the steady state exactly balances out the increased risk of apoptosis when the repair fails. Put in another way, a higher apoptosis rate does not increase the rate at which cells need to divide in order to keep tissue homeostasis. Consequently, the Hayflick limit in the tissue is not reached faster for a higher apoptosis rate.

Varying the last model parameter, the fragility (Δ), shows that a more fragile repair mechanism increases the self-sustaining time τ (see Figure 1.7). The reason for the increase in τ is that the repair rate decreases drastically when a cell accumulates a mutation. A second DNA damage will therefore often cause the cell to go apoptotic and become replaced. This means that cells with mutations are quickly removed from the system and healthy cells take their place. Thus even though fragile DNA repair is disadvantageous for the individual cell, the tissue as a whole is better off.

1.2.4   More neighbouring cells is beneficial

In addition to the three parameters shown in Figure 1.6, we also investigated how the size and dimension of the system would affect the stability of the tissue. For a well-mixed system, the time τ increases fast with the number of cells N, but in a two- or three-dimensional system, where apoptotic cells can only be replaced by neighbouring cells dividing, the time τ increases linearly with increasing N (see Figure 1.7 B).

1.2.5   The fragile DNA repair model

We investigate the well-mixed system analytically. Here, well-mixed means that each cell interacts with all of the other cells. The state of the system can be characterized by the number of times i each cell has accumulated DNA damage. Each time a cell accumulates damage, the cell will decrease its repair rate with Δ. The number of cells with the repair rate $R_i = R_0 - i\Delta$ is denoted N_i. When the cells fail to repair after being exposed to DNA damage, N_i will decrease. Conversely, N_i will increase when these cells replicate after another cell has gone apoptotic. The rate that a cell with repair rate
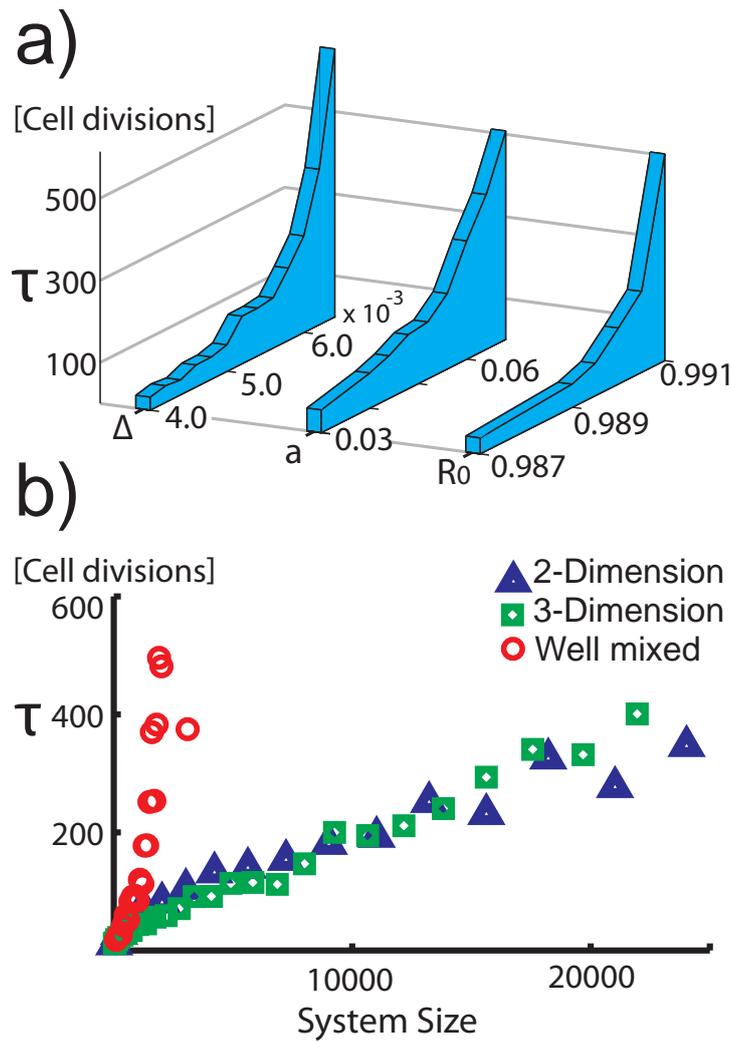
Figure 1.7: **Effect of parameters: A)** The time τ spent in steady state increases with the initial repair rate R_0, rate of apoptosis a, and fragility of the repair mechanism Δ. (Parameters that are not varied are set to the values given in the caption of Figure 1.6) **B)** For the well mixed system, the time spent at steady state, τ, increases drastically with system size N. For spatially structured systems the increase is linear. (Parameters used: $R_0 = 0.99$, $a = 0.05$, $\Delta_{\text{Well mixed}} = 0.01$, $\Delta_{\text{2dimension}} = 0.01$, $\Delta_{\text{3dimension}} = 0.008$, $N = 5000$.

R_i goes apoptotic is given by:

$$\text{Rate of apoptosis} = a \underbrace{(1 - R_i)}_{\text{P not repairing}} ,$$

$$\text{Average rate of apoptosis} = \sum_i a(1 - R_i)\frac{N_i}{N} = a(1 - \sum_i R_i \frac{N_i}{N}) = (1 - R^*)a.$$

Therefore the average rate that a cell goes apoptotic is given by: $(1 - R^*)a$. Where R^* is the average repair rate of the temporary steady state. Cells with $i > 0$ mutations will, in addition, increase in numbers when cells of repair rate R_{i-1} mutate one more time.

These considerations lead to the dynamical equations of the system:

$$\dot{N}_0 = (1 - R^*)aN_0 - (1 - R_0)N_0, \tag{1.1}$$

$$\dot{N}_i = (1 - R^*)aN_i - (1 - R_i)N_i + (1 - R_{i-1})(1 - a)N_{i-1}. \tag{1.2}$$

In order to investigate the temporal steady state of our simulation, we solve the steady state equations. We set the time derivatives to zero and solve for the average repair rate R^* in equation (1.1).

$$\boxed{R^* = 1 - \frac{1 - R_0}{a}} \tag{1.3}$$

R^* can then be inserted into (1.2) and using $R_i = R_0 - i\Delta$ we get the recurrence relation for the temporary steady state.

$$N_i = \frac{1}{i}\left(\frac{1 - R_0}{\Delta} - 1 + i\right)(1 - a)N_{i-1}. \tag{1.4}$$

This recurrence relation can be solved, using *Wolfram-Alpha*:

$$\boxed{N_i = N_0(1 - a)^i \binom{\frac{1 - R_0}{\Delta} - 1 + i}{i}} \tag{1.5}$$

where the last factor is a binomial coefficient and N_0 is the number of undamaged cells. Summing all N_i gives the total number of cells in the system N, and using the identity $\sum_{x=0}^{\infty}(1 - a)^x \binom{b+x}{x} = a^{-(b+1)}$ we get a relationship between all cells N and the number of unmutated cells N_0:

$$N = N_0 \sum_{i=0}^{\infty}(1 - a)^i \binom{\frac{1 - R_0}{\Delta} - 1 + i}{i}, \Leftrightarrow \tag{1.6}$$

$$N = N_0 a^{-\frac{1 - R_0}{\Delta} - 1 + 1}, \Leftrightarrow \tag{1.7}$$

$$N_0 = N a^{\frac{1 - R_0}{\Delta}}. \tag{1.8}$$

Notice that we here have summed over all i. However the sum should have been truncated at $i_{max} = R_0/\Delta$, since there are only $\frac{R_0}{\Delta}$ repair states that the system can be in. For small Δ's, states with $i > R_0/\Delta$ contain virtually no cells, so the truncated sum and the infinite sum becomes roughly the same (see Figure 1.8 A).

1.2.6   Collapse of system due to fluctuations

The steady state mean repair rate R^* (equation (1.3)) is in perfect agreement with simulations for the well-mixed system (see Figure 1.8). In addition to the mean repair rate, the theoretical steady state distribution (equation (1.5)) shows a perfect overlap with the actual distribution of the simulation, shown for two time points in Figure 1.8 A. Figure 1.8b shows the development of the distribution over time. Note the drastic collapse of the distribution. For this simulation the population of cells (N) was 5000; the other parameters are given in Figure 1.6.

We can show that the steady state distribution is attractive by introducing the pertubation $\widetilde{N_i} = N_i + \delta$ in (1.2):

$$
\begin{aligned}
\dot{\widetilde{N_i}} =& (1 - R^*)a\,(N_i + \delta) - (1 - R_i)\,(N_i + \delta) + (1 - R_{i-1})\,(1 - a)N_{i-1}, \\
=& (1 - R^*)aN_i - (1 - R_i)N_i + (1 - R_{i-1})\,(1 - a)N_{i-1} \\
& + (a(1 - R^*) - (1 - R_i))\,\delta, \\
=& \dot{N_i} + \left(a(\frac{1 - R_0}{a}) - (1 - R_i) \right)\delta, \\
=& \dot{N_i} + (R_i - R_0)\delta.
\end{aligned}
$$

Since $R_0 > R_i$ the perturbation will be reduced, making the steady state attractive for all states except $i = 0$.

The reason the distribution collapses is the finite size of the system. The number of cells that have accumulated damage i times will fluctuate around the distribution given by equation (1.5). If the number of undamaged cells randomly fluctuates to zero, new undamaged cells can never be reintroduced.

The time τ the system spends in the temporary steady state can be understood as the time passing before random fluctuations cause the number of undamaged cells to go to zero. After this, a new steady state with a lower maximum repair rate can be found, corresponding to the substitution $R_0 \to R_0 - \Delta$ in equation (1.8). From Figure 1.6, however, we see that this steady state will be even more unstable than the previous. This results in a fast collapse of the system to the absorbing state where all cells have accumulated the maximum amount of DNA damage (see Figure 1.8).

If the dynamics of the undamaged cells were a Poisson process the standard deviation of undamaged cells would be $\sqrt{N_0}$. However, in our simulation we saw larger fluctuation, although we were not able to quantify the magnitude of them. What we were able to show was that τ increases with system size N. The relative fluctuations of the undamaged cells decreases when the system size increases. Since the relative fluctuations decrease, so does the probability that the number of undamaged cells fluctuates to zero. That is, if N_0 is high, the time spent in the steady state τ increases drastically (see Figure 1.7 A). From equation (1.8) we see that N_0 increases with both R_0, a, Δ, and N, which explains figure 1.7 A.

1.2.7 Turnover rate of an average cell

From a biological point of view, an interesting measure is the time an average cell lives in the system, which is denoted by $\langle t_C \rangle$. The lifespan of a cell in our simulation is the number of time steps the average cell has gone through at the time it goes apoptotic. Remember that a time step is when N cells have acquired damage (see section *Simulation steps* Simulation steps). On average a cell acquires one damage per time step and has the probability $(1 - R^*)a$ of going apoptotic. Thus, the average cells will be alive for

$$
\langle t_C \rangle = \frac{1}{a(1 - R^*)} = \frac{1}{a\left(1 - \left(1 - \frac{1 - R_0}{a}\right)\right)} = \frac{1}{1 - R_0}. \tag{1.9}
$$

Notice that the average lifetime only depends on the initial repair rate R_0, meaning that increasing the rate of apoptosis a does not increase the number of cells going apoptotic. This is in agreement with our simulations.
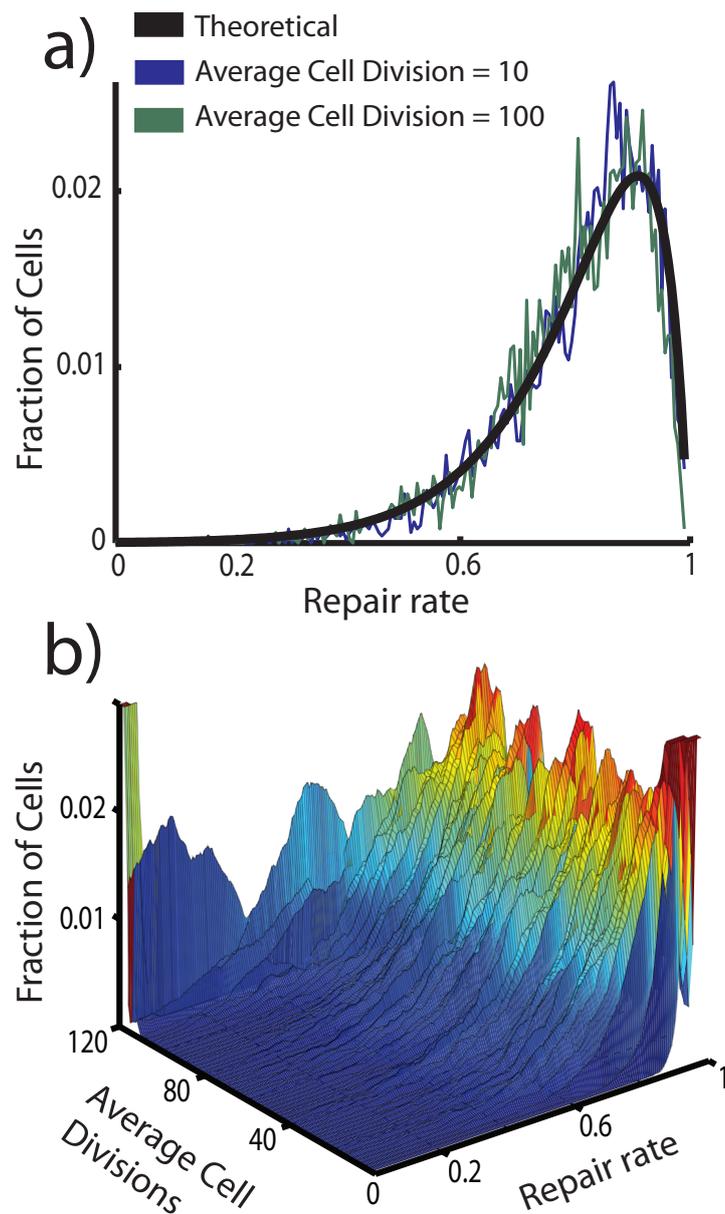
Figure 1.8: **Steady state distributions A)** The theoretically predicted distributions of repair rates in the well mixed systems, given by (1.5), is seen to agree well with the actual distribution during the temporary steady state. **B)** The distribution of repair rates as a function of time. (All parameters are set to the values given in the caption of Figure 1.6.)

1.2.8 Discussion

Ageing in fast turnover tissue of a multicellular organism can be seen as the progressive deterioration of cell function caused by wear and tear on somatic stem cells. Therefore, tissue that can maintain its function for a long time without renewal from stem cells will have a slower ageing process.

We have made a simple mathematical model to investigate how damage that compromises DNA repair capacity in single cells might affect a whole tissue. The model in its

present form captures two important experimental observations, namely the decreased repair rate and increased mutation frequency with age [2].

There are two interesting and non-intuitive predictions resulting from our model. Firstly, the model predicts that an initial rapid decline in repair capacity should be followed by a rather long period of persistently high repair capacity R^*. In this temporary steady state, the average number of mutations to genes coding for DNA repair should remain approximately constant. Experimental studies on mutation frequencies in mice by Busuttil et al. [22] seem to suggest similar temporal behaviour; the initial rapid increase in mutation frequencies (young mice) is followed by constant, or nearly constant mutation frequencies (older mice).

Secondly, the time of high repair capacity is longer when the repair mechanism is fragile. This could explain why a repair pathway such as the BER pathway has completely essential proteins [21]. Essential proteins are proteins that, when deleted, cause the mice to die as a fetus. In our model essential proteins, is modelled as a high Δ, since mutation to those proteins would severely decrease the repair rate. Therefore we would interpret the BER repair pathway as fragile. A congenital DNA repair mutation would in our model reduce the initial repair rate $R_0 \rightarrow R_0 - \Delta$. A lower initial repair rate drastically reduces τ (see Figure 1.6 A), which means faster ageing and eventual death. The evidence that BER mutations are embryonically lethal therefore supports our model. Other premature ageing diseases (e.g. Werner syndrome and Cockayne syndrome) could also be interpreted this way.

We find that the average lifetime of cells in the model is independent of the rate at which damaged cells go apoptotic. It should be noted that since our model only takes DNA damage affecting the DNA repair mechanism into account, it is not able to distinguish a successful repair from damage to the genome that does not directly or indirectly influence DNA repair pathways. If the rate of apoptosis is increased when damages occur to other parts of the DNA, the average lifetime of cells will decrease. It is still an open question to what extent apoptosis affects the process of normal ageing [25].

It should be pointed out that the model completely neglects DNA damages that occur during the replication process. The number of DNA damages inflicted in every cell every day may be as high as 100,000 [26]. However, it was not possible for us to find a number for the ratio of DNA damages occurring between DNA replication and DNA damages caused during the replication process. If a large amount of the DNA damage is due to the replication process, then keeping a highly functioning tissue by a high apoptosis rate will be impossible, simply because each replication would cause large amounts of DNA damage to the dividing cell.

1.3   Telomere shortening

Telomeres are the noncoding ends of chromosomes that prevent loss of genomic information during DNA replication [27]. Each cell division leads to a telomere shortening of 50–100 base pairs, partly due to what is known as the end-replication problem [28, 29]. When the telomeres reach a critical length the cell goes senescent, which refers to a state of permanent replication arrest that prohibits any further proliferation [3]. As explained in the introduction this number is called the Hayflick limit [1].

Short telomeres have been linked to increased mortality and age-related diseases [30] and accumulation of senescent cells is seen as one of the causes of aging [1, 31, 32]. However, the proliferation limit associated with telomere attrition is thought to work as a fail-safe mechanism against cells that divide in an uncontrolled fashion, particularly cancer cells [33]. It has therefore been suggested that telomere shortening is a trade-off between oncogenesis (onset of cancer) and physiological ageing [34, 35, 36].

Telomerase, which mainly consists of the two components TERT and TERC, elongates the telomeres such that the telomere length of cells is maintained during replication [37]. Telomerase is also found in stem cells and germ line cells [12]. The amount of telomerase in germ line cells is sufficient to maintain telomere length [27]. For stem cells, however, the level of telomerase is lower and the telomeres are therefore shortened after each cell division, but at a lower rate than for somatic cells [38]. Somatic cells are all of the cells that are not germ or stem cell. Here we particularly think of somatic cells as tissue cells, which are the cells performing the physiological function.

It is an open question why stem cells express telomerase, and why the level of telomerase is not high enough to avoid a shortening of the telomere length throughout life. The capability of cells to tune their telomere shortening rate by varying the expression of telomerase suggests the possibility of optimal telomere shortening strategies.

In the following sections we present a stochastic model that tries to answer if and why shortening of stem cell telomeres could increase the lifespan.

1.3.1   The simplistic oncogenesis versus ageing trade-off model

The goal of this model is not to describe every detail in the biology behind renewal of somatic cells by stem cells. The complete cell renewal process is very complicated, involving multiple biochemical signalling pathways [39], spatial distribution of stem cell niches and their mobility [10], several differentiation steps (stem cell $\rightarrow$ progenitor cell $\rightarrow$ ... $\rightarrow$ somatic cell) [40], and the interplay between immune system (clearance of senescence cells) and stem cell division [36].

Instead we create a higher level, coarse-grained and simplistic model to describe the trade-off between oncogenesis and physiological ageing in a multicellular organism. In our simplistic model the organism consists of cell types, stem cells and tissue cells.

The stochastic model describes the accumulation of cancerous mutations in the whole organism. Each cell division causes the telomeres to shorten, thereby decreasing the proliferation potential, while at the same time a cell division has the potential to cause mutations. Mutations can arise both in somatic cells and in stem cells.

The model is initialized by an unmutated stem cell dividing, and thereby creating a somatic cell and a daughter stem cell. The somatic cell then proliferates H_0 times, where the Hayflick limit is reached and the cell undergoes senescence. Each cell division, either

somatic or stem, has a probability of acquiring a cancerous mutation. For somatic cells this probability is denoted p, and for stem cells p_{sc}. Mutations acquired by somatic cells are lost when that lineage reaches its Hayflick limit H_G and the somatic cell goes senescent (see Figure 1.9). Mutations acquired in the stem cells are, however, permanent and inherited by all daughter stem cells. Note that in order to maintain homeostasis (i.e.m same tissue size), only one daughter cell survives after each replication. Each stem cell division causes the telomeres to shorten, which we model by the Hayflick limit H_G that declines after each stem cell division with a constant amount $\alpha \geq 0$. The constant reduction of the Hayflick limit will eventually cause the stem cell pool to go senescent. Since the somatic cells cannot be renewed from a senescent stem cell pool, this sets an upper limit to the lifespan. Reaching this upper limit could be considered as "dying due to ageing". We define oncogenisis as accumulation of more than a fixed number of cancerous mutations C_m in any cell. The trade-off is then between the maximal lifespan set by the stem cells going senescent versus the probability of accumulating more than C_m in either stem cells or somatic cells. The lifespan is measured as the total number of cell divisions that the system undergoes before the onset of cancer or stem cell senescence. A schematic illustration of the model is shown in Figure 1.9.

Please note that in order for us to analytically analyse the system, we have a stem cell population of one and a somatic cell population of one, which of course is very unrealistic. In the appendix to our paper we expand the model and computationally explore the case where a larger somatic cell population is maintained by a smaller stem cell pool; the results are qualitatively similar to those of this simpler model [41].
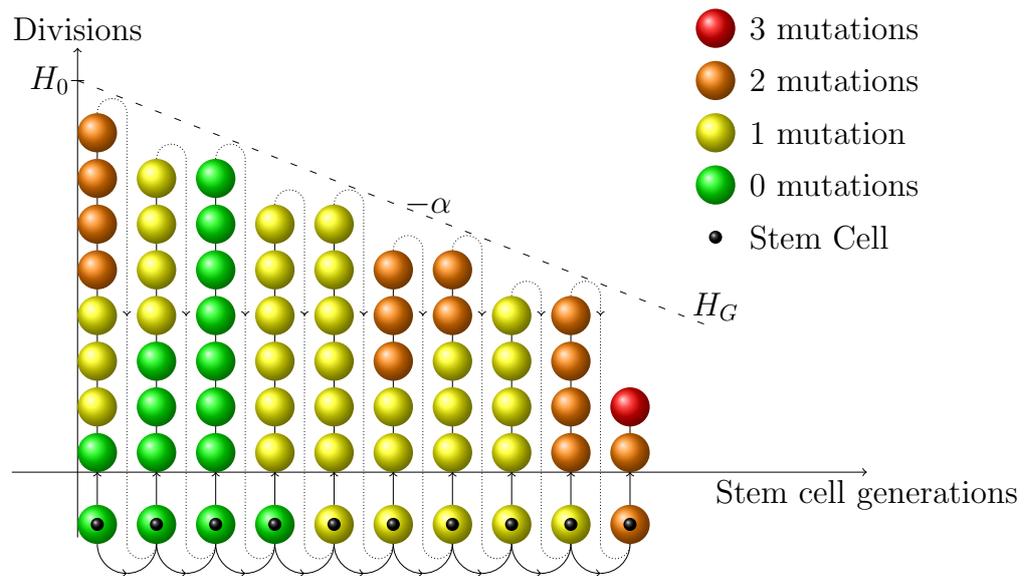
Figure 1.9: **Schematic illustration of the model:** The flow of time follows the dashed arrows. Starting at the bottom left corner, initially a stem cell divides. One daughter cell specializes into a somatic cell, which proliferates until it reaches the Hayflick limit and goes senescent (first vertical column). In order to keep a constant number of somatic cells, the daughter stem cell then divides to produce a new daughter stem cell and a new generation of somatic cells (subsequent vertical column). For each generation the Hayflick limit H_G is reduced by a constant amount α. At every cell division, somatic cells and stem cells have the probabilities p and p_{sc}, respectively, to acquire cancerous mutations. Mutations in stem cells are permanent, while mutations in somatic cell lineages are lost when the cell line reaches the Hayflick limit. When the system has accumulated more than C_m mutations (which in this illustration is two), the organism will have developed cancer. The total cancer-free lifespan is thus the sum of all somatic divisions.

1.3.2 Theoretical analysis

In this section we derive a mathematical expression for the average cancer-free lifespan obtained for different telomere shortening strategies.

Each time a stem cell divides its telomeres shorten. This decreases the Hayflick limit of the somatic cell by α. The initial Hayflick limit is H_0 so after G stem cell divisions (generations), the Hayflick limit H_G is given by:

$$H_G = \lfloor H_0 - \alpha G \rceil \tag{1.10}$$

Cell divisions are integers, so $\lfloor \rceil$ denotes rounding to the nearest integer. As explained before, stem cell telomere shortening ($\alpha > 0$) sets the upper limit for the lifespan of the system. The upper limit is simply $\frac{H_0 - 0.5}{\alpha}$ generations.

The somatic cells can divide H_G times, and each time the somatic cells have a probability p of acquiring a cancerous mutation. The mutations are for simplicity considered independent, which makes each mutation event equivalent to a biased coin toss. Since the sequence of mutations are unimportant, the probability $F(x)$ that the somatic cell lineage acquires x or fewer mutations before it reaches the Hayflick limit is given by the cumulative binomial distribution.

$$F(x) = \sum_{i=0}^{x} \binom{H_G}{i} p^i (1-p)^{H_G - i}. \tag{1.11}$$

With each stem cell division there is a risk p_{sc} of acquiring a permanent mutation in the stem cell lineage. The probability for the organism to have j mutations in the stem cells after G generations, but still be cancer-free, is denoted $S_{G,j}$. Since the number of mutations only increases, two events can lead to the state $S_{G+1,j}$. First, if the stem cell in generation G has j mutations, no mutation occurs during stem cell division and no oncogenesis happens during somatic cell division. Second, the stem cell in generation G has $j-1$ mutations, a mutation occurs during stem cell division, and again no oncogenesis happens during somatic division.

The probability $S_{G+1,j}$ to survive to generation G+1 with j stem cell mutations is therefore given by the recurrence relation:

$$S_{G+1,j} = \overbrace{\left[S_{G,j}(1 - p_{sc}) + S_{G,j-1} p_{sc} \right]}^{\text{Probability for j mutation in SC}} F(C_m - j), \tag{1.12}$$

where the expression in square brackets is the probability of having j mutations in the stem cell, and $F(C_m - j)$ is the probability of not exceeding the critical number of C_m mutations during the somatic divisions.

The initial stem cell is unmutated, $S_{0,0} = 1$, so in terms of the Kronecker delta the starting condition is

$$S_{0,j} = \delta_{0j} \tag{1.13}$$

Using Eqs. (1.12) and (1.13), the system can be solved numerically. The overall probability that the organism will not have developed cancer at a given generation is defined as:

$$S_G = \begin{cases} \sum_j S_{G,j} & \text{for } G \leq \frac{H_0 - 0.5}{\alpha} \\ 0 & \text{for } G > \frac{H_0 - 0.5}{\alpha} \end{cases} \tag{1.14}$$

$F(C_m - j)$ is zero for $j > C_m$. The average cancer-free lifespan $\langle L(H_0, \alpha) \rangle$ is the mean number of cell divisions the organism will undergo before exceeding the critical number of mutations. It can be found by multiplying the probability of getting cancer at each generation with the number of somatic cell divisions it takes to reach that generation.

$$\langle L(H_0, \alpha) \rangle = \sum_{G=1}^{\infty} P(G) \sum_{G'=1}^{G} H_{G'-1}$$

Probability $P(G)$ of oncogenesis in generation G is given as $(S_G - S_{G+1})$. Inserting this and substituting $H_{G'-1}$ with equation 1.10.

$$\langle L(H_0, \alpha) \rangle = \sum_{G=1}^{\infty} (S_G - S_{G+1}) \sum_{G'=1}^{G} H_0 - \alpha \left(G' - 1 \right)$$

Using the identity $\sum_{n'=1}^{n} n' = \frac{n(n+1)}{2}$ yields.

$$\langle L(H_0, \alpha) \rangle = \sum_{G=1}^{\infty} (S_G - S_{G+1}) G H_0 - \alpha \frac{G(G+1)}{2} - G$$

$$= \sum_{G=1}^{\infty} (S_G - S_{G+1}) G \left(\frac{2H_0}{2} - \alpha \frac{G+1-2}{2} \right)$$

$$= \sum_{G=1}^{\infty} (S_G - S_{G+1}) G \left(\frac{2H_0}{2} - \alpha \frac{G-1}{2} \right)$$

Rewriting using equation 1.10.

$$\langle L(H_0, \alpha) \rangle = \sum_{G=1}^{\infty} (S_G - S_{G+1}) G \frac{H_0 + H_{G-1}}{2} \tag{1.15}$$

When $\alpha = 0$ the Hayflick limit H_0 is constant

$$\langle L(H_0, \alpha) \rangle = \sum_{G=1}^{\infty} (S_G - S_{G+1}) G H_0$$

The sum written out is $(S_1 - S_2) + 2(S_2 - S_3) + 3(S_3 - S_4) + ...$ rearranging the terms yields $S_1 + S_2 + S_3 + S_4 + ...$ so eq. (1.15) becomes:

$$\langle L(H_0, 0) \rangle = H_0 \sum_{G=1}^{\infty} S_G \tag{1.16}$$

Note that for this case S_G will never reach 0 and a cut-off is necessary. This is made when $S_G < 10^{-5}$ and has negligible influence on the cancer-free lifespan. A simpler way of calculating $\sum_{G'=1}^{G} H_{G'-1}$ the number of somatic division when reaching generation G, is by using elementary school geometry. The number of divisions can be interpreted as the area under the curve up until generation G in Figure 1.10. By calculating the area of the squares and triangles we recover the same expression as in equation 1.15. This serves as a good sanity check.

1.3.3 Results

Interestingly, the mutation rate in stem cells seems to be a roughly a factor 100 lower than in somatic cells [42, 43]. This fixes the ratio of mutation rates between stem cells and somatic cells $\frac{p}{p_{sc}} = 100$ and we set $p = 10^{-2}$. The number of mutations necessary
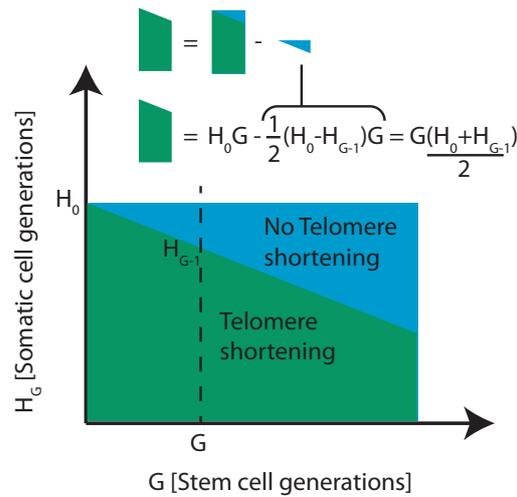
Figure 1.10: **Geometric sanity check** Part of the average lifespan equations can be derived by noticing that the number of cell divisions is simply the area under the curve. The blue and green area until G is the lifespan without telomere shortening after G generations. The green area is with telomere shortening and using simple geometry we can calculate part of equation 1.15.

to develop cancer and the mutations interplay is an active topic of research, but a recent study using sequencing data from roughly 3000 different cancer types estimates the number to be between two-six mutations [44], which matches the estimate before [45]. We set the number of mutations high: $C_m = 6$. The choices of p, $p_s c$ and C_m have quantitative but not qualitative influence on the results (see the appendix to our paper [41]). These parameters were chosen such that the optimal lifespan found was close to the experimentally observed Hayflick limit of 50 [4]. The average cancer-free lifespan is then calculated from Eqs. (1.15) and (1.16).

The goal of this model is to investigate the trade-off between ageing processes and oncogenesis. Setting α to zero means no telomere shortening during stem-cell division, and in the context of our model no ageing. Under the assumption of no ageing ($\alpha = 0$) the initial Hayflick limit/telomere length can be optimized for maximal average cancer-free lifespan $\langle L(H_0, 0) \rangle$. We denote this set of parameters, that is, the optimal lifespan in a non-ageing organism, as strategy 1 (see point 1 in Figure 1.11),

We scan the model in the parameter space H_0 and α, contrasting with strategy 1 (see Figure 1.11), picking the following three strategies for further study:

1. **Optimal non-ageing:** The strategy where stem cells express a telomerase activity, such that the telomere length of stem cells is constant over time ($\alpha = 0$). The initial Hayflick limit is chosen such that the cancer-free lifespan is maximized.

2. **Optimal ageing:** The strategy that gives the longest cancer-free lifespan when allowing stem cell telomeres to shorten at each division ($\alpha > 0$). Again, the initial Hayflick limit is chosen such that the cancer-free lifespan is maximized. This results in a larger initial Hayflick limit than for strategy 1.

3. **Evolutionary optimal:** The strategy that combines the initial Hayflick limit from strategy 1 and the stem cell telomere shortening rate from strategy 2. Note that
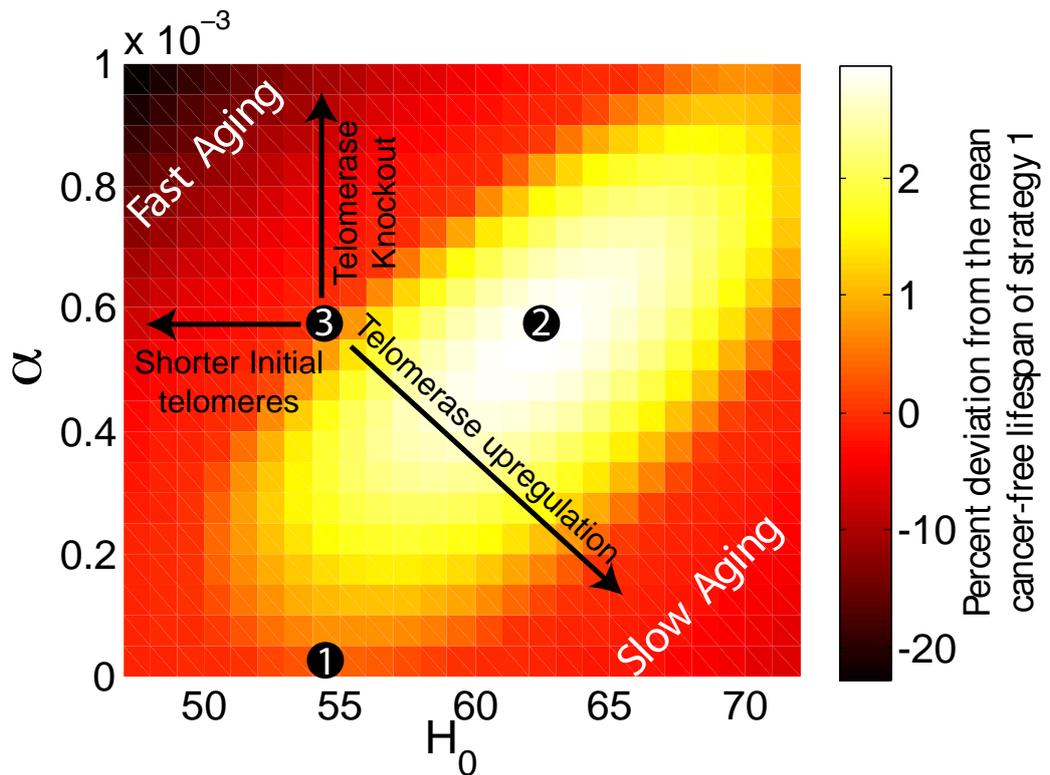
Figure 1.11: **Three strategies:** Deviation in average cancer-free lifespan for different initial Hayflick limits and shortening factors α compared to the longest possible cancer-free lifespan for a constant Hayflick limit (strategy 1). The overall longest cancer-free lifespan (strategy 2) is obtained for a higher initial Hayflick limit, which is then slowly reduced through telomere shortening in stem cells. Strategy 3 has the same initial Hayflick limit as strategy 1 and the same telomere shortening rate as strategy 2. Black arrows show how different experimental setups can change the position in parameter space through modifications of the initial Hayflick limit of stem cells and the telomere shortening rate. A low initial Hayflick limit and high telomere shortening rate cause premature ageing due to increased accumulation of senescent cells. The parameters are $C_m = 6$, $p = 10^{-2}$, $p_{sc} = 10^{-4}$.

this strategy yields a similar cancer-free lifespan as obtained for strategy 1.

First off, note that either increasing α or decreasing H_0 increases the turnover of stem cells thereby accelerating ageing (decreasing the maximal lifespan). Second, the faster turnover of stem-cells increases the risk of accumulating permanent mutations in the stem cells. On the other hand, this decreases the number of somatic cell divisions, thereby decreasing the risk of mutations arising in the somatic cells.

1.3.4   Optimal ageing

Our model predicts that the optimal ageing strategy (e.i., strategy 2) balances the risk of accumulating mutations and the ageing process by having longer initial telomeres (higher H_0) than strategy 1 but decreasing the telomere length of stem cells ($H_G < H_0$). The mechanism behind strategy 2 is that the decreased proliferation potential later in life balances the increased risk of accumulating the last fatal mutation.
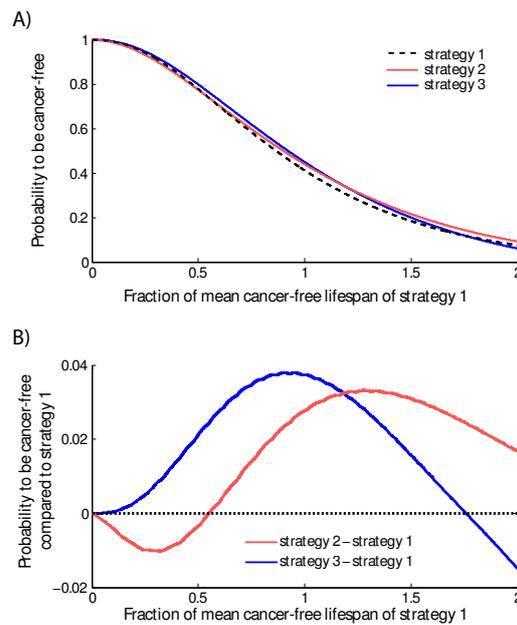
Figure 1.12: **Comparison of the three strategies:** Probability of not yet having developed cancer as a function of time for different telomere shortening strategies. **A)** For all three strategies, the probability slowly decreases with time. Only very few individuals will develop cancer later than twice the mean cancer-free lifespan of the population. **B)** Probability of not yet having developed cancer for strategies 2 and 3 compared to strategy 1. Strategy 2, which has the longest average cancer-free lifespan, will have a higher probability of cancer development early in life. Strategy 3, which has a low initial Hayflick limit, is able to postpone cancer until later in life.

1.3.5 Evolutionary optimal ageing

Investigating the probability of being cancer-free as a function of time for the three strategies (see Figure 1.12), we find that even though strategy 2 is the optimal strategy for an average cancer-free lifespan, strategy 3 (which has the same average life expectancy as strategy 1) has a lower probability of oncogenesis early in life. We speculate that from an evolutionary point of view it is advantageous to keep the reproductive population cancer-free. The negative effect on the old portion of the population is however not optimized for. Therefore, strategies that solely optimize cancer-free lifespan may not be advantageous from an evolutionary point of view if they increase the risk of cancer early in life when reproduction takes place, as seen with strategy 2.

1.3.6 Temporal regulation

Since the trade-off seems to be between optimizing against early life oncogenisis versus slow ageing in late life, we investigate what would happen if telomerase is upregulated later in life. We compare strategy 3, the evolutionary optimal strategy with strategy 3*. Strategy 3* has the same initial conditions as strategy 3, but will, when reaching $H_G = 39$, double its telomerase production (halving α) for the remaining cancer-free lifespan. With this up-regulation in telomerase production, longevity can be increased (see Figure 1.13). The increased average cancer-free lifespan relies solely on the ability to postpone stem cell senescence by up-regulating telomerase. The steep peak in Figure (1.13), which

is seen after a time that is 2.1 longer than the mean cancer-free lifespan, is caused by the stem cells of strategy 3 going senescent so S_G immediately drops to 0.
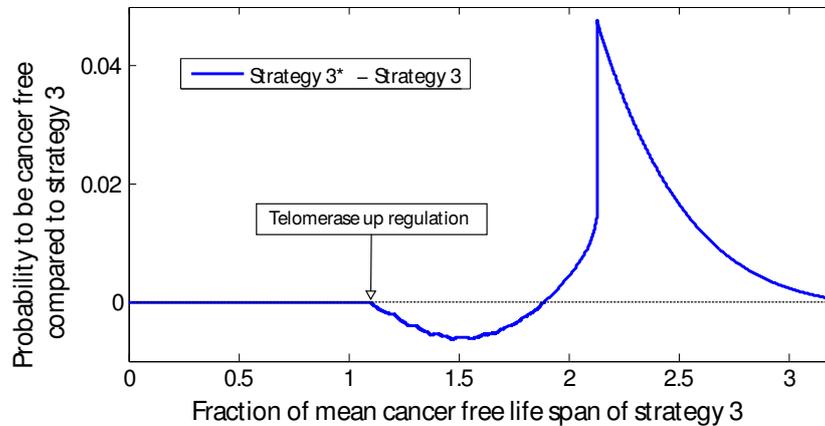
Figure 1.13: **Temporal telomerase regulation** Increased probability to be cancer-free for two strategies with same initial conditions but with strategy 3* having an up-regulation of telomerase in stem cells later in life. The up-regulation (lower α) leads to a slightly increased cancer risk, but the strategy turns advantageous later on, as the increased telomerase production postpones the time at which the stem cells go senescent. The stem cells of strategy 3 go senescent shortly after a time 2.1 times longer than the mean cancer-free lifespan, whereas this is postponed for strategy 3*.

1.3.7   Discussion and conclusion

The effect of optimizing the initial telomere length (H_0) and stem cell telomerase activity α seems very small, namely a 2.5% increase in average lifespan or roughly two years. Both the simplicity of our model and lack of directly measured parameters limit the credibility of the model's quantitatively predictions. However, for comparison, the effect on increased average lifespan if all cancers were eradicated is estimated to be two-and-a-half years [46], also when introducing population sizes the advantage of limited telomerase activity increases (see appendix of our paper [41]). The advantage of the simplicity/abstraction level of the model is that it can easily connect several otherwise disconnected experiments.

Assuming that strategy 3 is the evolutionary optimal strategy, the model predicts that removing telomerase, which is the same as increasing α, will increase risk of cancer because of increased turnover of stem cells and cause faster ageing because of stem cell senescence (see Figure 1.11). Indirectly the model also predicts that the effect is most pronounced in highly proliferating tissue, since the turn-over of stem cells would also be faster. In telomerase (TERC) deficient mice, González et al. (2000) [47] showed a slight increase in oncogenesis in highly proliferating tissue. Also in TERC deficient mice, Liu et al. (1998) showed premature ageing symptoms, especially in highly proliferating tissue [48].

In our model, up-regulating telomerase activity means that α decreases, which corresponds to less telomere shortening during stem cell division. However, since the telomerase would also be up-regulated in somatic cells, the proliferation potential would increase, which means H_0 would increase (see Figure 1.11). Our model predicts that an increase in telomerase activity will cause the organism to be more prone to cancers. This

effect was experimentally observed by González et al. (2002) [49]. However, the model
also predicts slower ageing, which was not checked for in the study.

Our model predicts that decreasing H_0 leads to a faster attrition of the stem cell pool
and therefore faster ageing. Comparing two types of telomerase deficient mice - one
with initially shorter telomeres (low H_0) and one with normal initial telomere length -
the mice with shorter telomeres show decreased viability [50, 51]. The lower viability
was not due to increased cancer incidents, and premature aging signs such as graying
of hair were noticed. However, the studies also showed that several ageing signs such
as increase in osteoporosis were not observed.

Lastly, the model predicts that a temporal induction of telomerase in stem cells could
increase cancer-free lifespan (see Figure 1.13). If the organism starts with a moderate
telomerase activity and then increases the activity in stem cells later in life, the organism
could thereby prevent cancer onset at an early stage of life and postpone ageing, thereby
increasing longevity. A recent study by Bernandes de Jesus et al. (2012) supports this
result. By inducing telomerase activity in adult mice via a virus treatment, they were
able to increase longevity without increasing the risk of cancer [52].

The major discrepancy between our model and data is the shape of the survival
curve. When contrasting real survival curves with our model it is clear that the shape
of the curves only show a very slight resemblance to each other (see Figure 1.14).
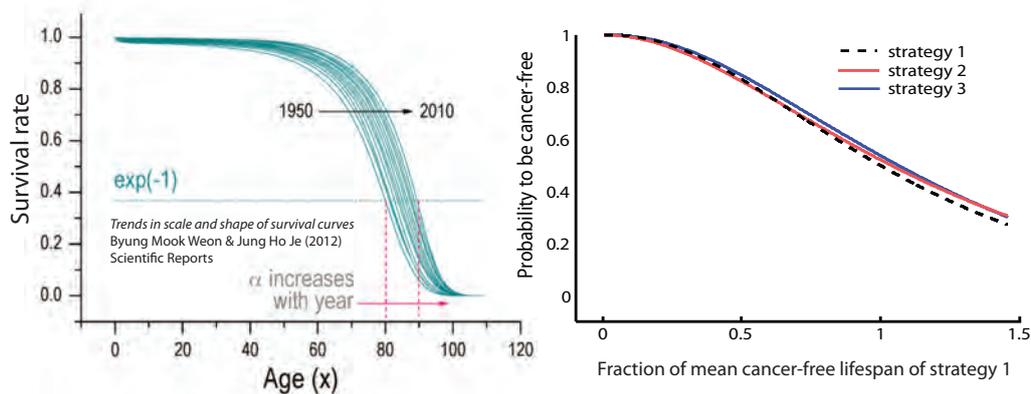
Figure 1.14: Survival curves for swedish females over six decades (1950-2010) figure
published by Byung Mook Weon and Jung Ho Je [53]. When contrasting the survival
curves from real data and our model there is a clear discrepancy between the shape
of the curves. This Indicates that our model does not capture overall survival in a
satisfactory way.

In conclusion, not all phenomena of physiological ageing can be attributed to telom-
ere length and stem cell turn-over rates, and our model unfortunately does not capture
the shape of real survival curves. Our model simply tries to give a rationale as to why
moderate telomerase activity in stem cells might be advantageous, and at the same
time capture some aspects of physiological ageing observed in otherwise disconnected
experiments.

1.4   Dynamics of DNA repair proteins

DNA repair proteins are crucial for maintaining genomic stability through a variety of pathways [54]. DNA repair is a term that covers several distinct biochemical pathways; among these are the homologue and non-homologue DNA repair pathways, both of which are capable of repairing double-stranded breaks (DSB). DNA repair can be visualized as a biochemical cut and paste mechanism, where the DNA damage is cut out of the DNA strand and a new piece of DNA is synthesized and glued back into the DNA strand. All of the DNA repair processes involve multiple steps and several classes of proteins. For each pathway most of the proteins can be classified as follows [55]:

1. Endo/exo-nucleases and glycosidases (*cutting DNA*)

2. DNA helicases (*unwinding DNA*)

3. DNA ligases (*glueing DNA backbone*)

4. DNA polymerases and nucleotidyltransferases (*adding nucleotides*)

5. Adaptors (*Protein–protein interactions*)

Mutations in these proteins lead to devastating diseases with symptoms that include premature ageing [20] and increased risk of cancer [56]. Individuals who lack either Werner syndrome helicase (WRN) or Bloom syndrome helicase (BLM) show symptoms of accelerated ageing [20].

Most DNA repair proteins (e.g., WRN and BLM) tend to accumulate at the site of DNA damage caused by genotoxic stresses. While extensive research has been done on the functions of DNA repair proteins, the process of DNA repair protein recruitment to the site of DNA damage is not well understood. The repair of DNA damage is a sequential process and the recruitment of DNA repair proteins is thought to be somewhat hierarchical [57]. This means that a subset of DNA repair proteins detects the damage and promotes recruitment of other proteins that then again could promote recruitment of yet other proteins and so on. A minimal requirement for a DNA repair protein to detect damage is that the protein interacts with the DNA. Most DNA repair proteins have DNA binding domains and can therefore potentially interact directly with the DNA [55]. However, it is possible that a structural change is needed for the DNA binding domain to be exposed and thereby to allow for protein-DNA binding.

Following this logic we expect that proteins where the DNA binding domain is always exposed will constantly participate in binding/unbinding reactions to exposed stretches of DNA and therefore have the potential to scan for DNA damages. The binding/unbinding events of the DNA damage scanning proteins will hamper the diffusion of these proteins. Oppositely, if the protein's DNA binding domain is only exposed during the DNA repair process, the accumulation at the DNA damage site will be in response to the detection of damage by the scanner. We expect that these proteins will have higher mobility in the nucleoplasm.

The classification of DNA repair proteins into either scanners or responders has previously been suggested by Houtsmuller et al. [58]. Here we try to use their mobility to classify them. Performing Fluorescent Recovery After Photobleaching (FRAP) experiments, we have investigated the spatial dynamics of two DNA repair proteins (namely

WRN and BLM). The dynamics allowed us to classify the two proteins as either (possible) scanners or responders. Comparison of diffusion coefficients for other DNA repair proteins (obtained from literature) shows a separation of proteins into either scanners or responders.

Another question that we tried to answer is whether the accumulation of WRN and BLM at DNA damage is limited by the time it takes the proteins to find the damage (i.e., the search time) or if the accumulation is due to reaction steps of other DNA repair proteins. In order to answer this question I created regions of DNA damage in cells and monitored the accumulation of WRN and BLM in these regions (illustrated in Figure 1.15). Lastly, since neither WRN nor BLM are uniformly distributed throughout the nucleus but highly concentrated at nucleoli (nucleus sub-compartment) [59, 60], we also measured the diffusion coefficient at the nucleoli. WRN and BLM are both involved in the repair of double-stranded breaks [61], therefore, unless specified otherwise, all mentions of DNA damage in this section refer to DSBs.

1.4.1   Experiments

WRN and BLM proteins were fused with Green Fluorescent Protein (GFP). All measurements and theoretical calculations where the abbreviations WRN and BLM are used to refer to the fusion proteins WRN-GFP and BLM-GFP, respectively. To examine the dynamics of WRN and BLM we used FRAP. In a FRAP experiment fluorescent molecules in a small area of the nucleus become irreversible photobleached [62] (see Figure 1.15). The photobleaching was induced by a focused (micropoint) laser. The surrounding fluorescent proteins then diffuse into the small area. The mobility of the fusion protein can be obtained from the recovery of fluorescent intensity.

If the observed proteins are very mobile the intensity is quickly recovered; however, for immobile proteins no recovery will occur [63]. In FRAP experiments where the intensity is not completely recovered, the proteins are separated into: "the mobile fraction" and "the immobile fraction". The immobile fraction is the fraction of proteins that is stably bound to immobile (or almost immobile) cellular components (e.g. DNA) [63]. Immobile means stably bound within the timespan of a measurement (usually a couple of minutes): as such, "slow exchanging protein fraction" would have been a more correct phrase. The mobile fraction, on the other hand, is the fraction of proteins that exchanges position with the surrounding fluorescent proteins within the timespan of a measurement.

We performed FRAP in the nucleoplasm and nucleoli, both for WRN and BLM. FRAP at the site of DNA damage was performed when the recruitment of WRN and BLM to the site of damage had saturated. This took roughly two hours. The low laser power used for FRAP did not induce damage (see Figure A.4 in the appendix). For details on the microscope setup and cell culture preparation,s see the Material and methods section of our paper [64].
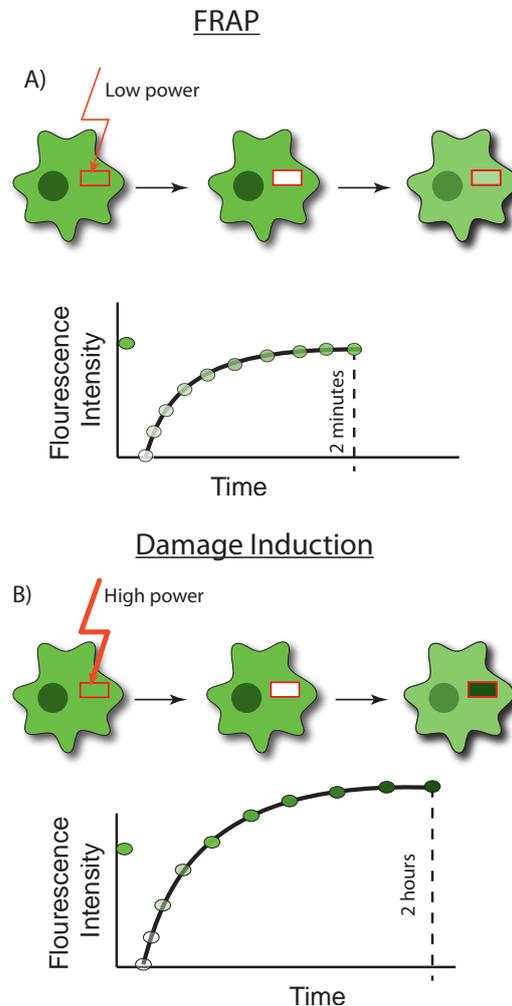
Figure 1.15: **Flourescent Recovery After Photobleaching and Damage induction experiments:**
A) FRAP: Fluorescent molecules are irreversibly photobleached in a small area by a focused laser with a low laser power (0.6 μW). Because fluorescent molecules diffuse into the bleached area, the fluorescence in the small area is recovered. Mobility of the fluorescent molecule can be estimated from fitting the FRAP curve. Intensity for both WRN and BLM were recovered within a couple of minutes. **B) Damage Induction:** A high power (1.8μW) focused laser is used to generate DNA damage in a small region. DNA repair protein assemble at the site of damage and we monitored the GFP-tagged WRN and BLM that accumulated at the damage site. Accumulation of WRN and BLM at the damage site saturated after roughly two hours.

1.4.2 FRAP models

The spatio-temporal dynamics of proteins in a cell are often well described by the reaction-diffusion model (R-D model) [65]. By fitting solutions of the R-D model to fluorescent recovery curves, we can identify whether there are rate-limiting steps in the dynamic of interest; furthermore if there is a rate limiting step, we can distinguish whether the dynamics are dominated by diffusion or reaction.

To establish the nature of the dominant dynamics, we compared three different mathematical models: a full R-D model, a model limited by diffusion and a model limited by reaction.

Reaction-diffusion model:

$$\frac{\partial F}{\partial t} = D\nabla^2 F - k_{on}^* F + k_{off} B \tag{1.17}$$

$$\frac{dB}{dt} = k_{on}^* F - k_{off} B \tag{1.18}$$

Where F and B are the concentration of free and bound protein, respectively. k_{off} is the off-rate and $k_{on}^* = k_{on} S$ is the pseudo on-rate where S is the number of binding sites. D is the diffusion coefficient. We have assumed here that the bound proteins are spatially fixed. The fluorescent signal ($Frap(t)$) is given as the total fluorescent intensity inside the bleach spot of both free and bound proteins:

$$Frap(t) = T(t) = F(t) + B(t) \tag{1.19}$$

Where T is the total amount of protein, $T = F + B$. Since intensity is normalized prior to bleaching, we assume that in equilibrium $F_{eq} + B_{eq} = 1$. For the derivations in this thesis we assume that the equilibrium after FRAP bleaching is the same (see appendix A), the correction to this is described in the *FRAP analysis: Correction for initial bleaching profile and finite geometry* section. We further assume rotational symmetry. There is no closed form for the Frap(t) given the R-D model; however, for the Laplace transform the solution is:

$$\mathcal{L}\left(Frap(t)\right) = \frac{1}{s} - \frac{F_{eq}}{s}\left(1 - K_1(qr_c)I_1(qr_c)\right)\left(1 + \frac{k_{on}^*}{s + k_{off}}\right) - \frac{B_{eq}}{s + k_{off}} \tag{1.20}$$

Where $\mathcal{L}\left(Frap(t)\right)$ is the Laplace transform of the intensity signal $Frap(t)$. s is the complex frequency. r_c is the radius of the bleaching spot. I_1 and K_1 are modified Bessel functions of respectively the first and second kind. q is given by $q^2 = \frac{s}{D}\left(1 + \frac{k_{on}^*}{s + k_{off}}\right)$. We obtained the intensity signal $Frap(t)$ by numerically inverting the Laplace transform.

Effective diffusion model: In the limit where the dynamics of binding/unbinding are fast, B(r,t) can be considered in equilibrium. The dynamics are then dominated by diffusion and the model then becomes.

$$Frap(t) = F(t) + B(t) = F(t) + B_{eq} \tag{1.21}$$

Where the equation governing the fluorescent signal can be written as (see appendix A.2 for derivation):

$$\frac{\partial Frap(t)}{\partial t} = D_{eff}\nabla^2 F \tag{1.22}$$

The effective diffusion is given by $D_{eff} = D \frac{1}{1+\frac{k_{on}^*}{k_{off}}}$. Solving for the bound proteins in equilibrium and using that $T = F + B$, we find that in equilibrium $F = \frac{T}{1+\frac{k_{on}^*}{k_{off}}}$. This gives an intuition about the effective diffusion coefficient since it can be rewritten as $D_{eff} = D \frac{F}{T}$ so intuitively the reduction of the diffusion coefficient is simply proportional to the fraction of free proteins. The solution to 1.22 is given as:

$$Frap(t) = \left[I_0 \left(\frac{r_c^2}{2tD_{eff}} \right) + I_1 \left(\frac{r_c^2}{2tD_{eff}} \right) \right] \exp \left(-\frac{r_c^2}{2tD_{eff}} \right) \tag{1.23}$$

r_c^2 is the radius of the bleach spot. I_0 and I_1 are modified bessel functions of the first kind. From this we get that if the data is best fitted by the diffusion model, we can estimate the fraction of free proteins based on the fraction between the theoretically expected diffusion coefficient $D_{theoretical}$ and the measured effective diffusion coefficient D_{eff}.

Reaction model: In the other limit where the dynamics are dominated by the binding/unbinding events, the equations reduce to (see appendix A.3 for derivations):

$$Frap(t) = F(t) + B(t) = F_{eq} + B(t) \tag{1.24}$$

$$= 1 - B_{eq} \exp \left(-k_{off}t \right) \tag{1.25}$$

When the data can be fitted by the reaction model the unbinding rate (k_{off}) can be calculated directly from the fit.

The two simpler models, namely diffusion and reaction, are approximations of the full model. This means that for any dataset the full R-D model will give the best fit, but at the cost of an extra parameter. Models are therefore chosen based on their Akaike Information Criteria value, (see 1.4.3 *Model selection criteria* for details.)

1.4.3 Model selection criteria

We use the Akaike Information criteria (AIC) to differentiate between the three models. AIC is a measure that evaluates a model based both on the goodness of fit and the number of variables. The AIC is given as [66]:

$$AIC = 2k - \ln(L) \tag{1.26}$$

Where k is the number of variables and L is the maximum likelihood of the model. The likelihood of a non-linear fit is given as[67]:

$$\ln(L) = -\frac{n}{2} \cdot (\ln(2\pi) + 1 + \ln(\langle RSS \rangle)) \tag{1.27}$$

Where $\langle RSS \rangle = \frac{\sum_{i=1}^{n}(y_i - f(x_i))^2)}{n}$ is the average sum of squares of the residuals, n is the number of data points, y_i is the i-th data point and $f(x_i)$ is the predicted value of the model. Taking small sample size into account, the corrected AIC is given as [68]:

$$AIC_c = AIC + \frac{2k(k+1)}{n-k-1} \tag{1.28}$$

The model with the lowest AIC_c is the preferred model. AIC_c can be considered a correction of the log-likelihood for a model. This also means that the relative likelihood between models can be calculated[68, 67].

$$p_i = \exp\left(\frac{AIC_{c,min} - AIC_{c,i}}{2}\right) \tag{1.29}$$

Where p_i is the probability that the i-th model is as good as the preferred model. A similar model selection criteria is the Bayesian Information Criteria (BIC), which is given as:

$$BIC = k \cdot ln(n) - 2\ln(L) \tag{1.30}$$

It should be noted that neither BIC nor AIC is an absolute measure since they only evaluates the models relative to each other. The $\langle RSS \rangle$ is an absolute measure, but it does not take number of variables into account and therefore does not penalize for over fitting. We therefore indicate, the AIC, BIC and $\langle RSS \rangle$ which shows that the choice of selection criteria does not affect the preferred model, and allows us to have an "absolute" value for how good the models are.

1.4.4 FRAP analysis: Correction for initial bleaching profile and finite geometry

Circular spots were bleached in the nucleoplasm, nucleolus and damage site. The spot diameter was 1μm in all cases except for WRN in the nucleoplasm where it was 2 μm. The size of the spot is a trade-off between a good signal to noise ration and recording the first time points. At least five cells were used to create the recovery curves. The images were background corrected and the intensity of the FRAP region was divided by the average intensity of the nucleus to correct for photobleaching. The derivation of the models (as seen in appendix A), assumes that the bleaching spot is cylindrical and that the nucleus is infinite. As work by Mueller et al. [69] has shown that these two assumptions can affect the estimation of the diffusion constant, we therefore correct for these two. First the initial bleaching profile is taken into account by fitting it to a modified Gaussian, as described by Mueller et al. [69].

$$I_0(r) = \begin{cases} \theta & r < r_c \\ 1 - (1-\theta)\exp\left(-\frac{(r-r_c)^2}{2\sigma^2}\right) & r \geq r_c \end{cases} \tag{1.31}$$

The shape of the profile is a "flat-headed" Gaussian profile, where r_c determines the size of the flat (constant) part.

Second, the finite geometry of the nucleus is taken into account. We assume that the nucleus is circular and need to estimate the radius of the nucleus. This radius R is an effective radius based on the final recovery level ϕ and the initial bleaching profile $I_0(r)$. The total fluorescent intensity of the nucleus after bleaching is given as:

$$F_A = 2\pi \int_0^R rI_0(r)dr$$

Since the images are corrected for photo bleaching, F_A is constant over time. Because the images were normalized before bleaching, the total intensity inside the nucleus before

bleaching F_B is simply the area of the nucleus.

$$F_B = 2\pi \int_0^R rI(r)dr = 2\pi \int_0^R r1dr$$
$$= \pi R^2$$

The ratio between the intensity after (F_A) and before bleaching (F_B) is then given as:

$$\phi = \frac{F_A}{F_B} \frac{2\pi \int_0^R rI(r)dr}{\pi R^2} \qquad (1.32)$$

ϕ the loss of intensity from FRAP can be directly measured in the images and $I_0(r)$ is found by fitting equation 1.31 to the data. Using equation 1.32 allows us to find the effective radius R of the nucleus. The initial profile $I_0(r)$ and the effective radius of the nucleus R is then used to correct the FRAP models as described in Mueller et al. [69].

1.4.5   Diffusion coefficient of fusion proteins

Given the relation between molecular weight and hydrodynamic radius[70]:

$$R_h \propto M^{\frac{1}{3}}. \qquad (1.33)$$

and the "Stokes-Einstein" equation:

$$D = \frac{k_B T}{6\pi \eta R_h}. \qquad (1.34)$$

We get that the diffusion coefficient D scales with the mass M as:

$$D \propto M^{-\frac{1}{3}}. \qquad (1.35)$$

Using the scaling relationship of the corresponding molecular weights $\frac{D_{\text{fusion protein}}}{D_{GFP}} = \sqrt[3]{\frac{M_{GFP}}{M_{\text{fusion protein}}}}$. With M_{GFP} = 27 kDa [71], M_{WRN} = 165 kDa [72] and M_{BLM} 170 kDa [73] and D_{GFP} = 28 [74]. We can estimate the corresponding diffusion constants of the fusion protein: $D_{WRN-GFP}$ = 14.7 $\frac{\mu m^2}{s}$ and $D_{BLM-GFP}$ = 14.4 $\frac{\mu m^2}{s}$

This assumes spherical shapes of the proteins, which can be corrected for according to Erickson (2009) [70]. The deviation from a spherical confirmation to a globular protein structure corresponds to a 20% reduction of the diffusion coefficient.

$$D_{theoretical} = \frac{D_{fusion}}{1.2} \qquad (1.36)$$

This results in $D_{WRN,theoretical}$ = 12.2 $\frac{\mu m^2}{s}$ and $D_{BLM,theoretical}$ = 12.0 $\frac{\mu m^2}{s}$.

1.4.6   WRN and BLM undergo fast binding/unbinding reactions in nucleoplasm and nucleoli with at least 90% of proteins bound to DNA at any time

We analysed the WRN mobility in the nucleoplasm. The results are shown in Figure 1.17. Using the AIC we find that the diffusion model is the best model among the three we tested (see table 1.16). From the diffusion coefficient of GFP (28 $\frac{\mu m^2}{s}$) [74], the size of WRN [72] and shape corrections [70] we calculated the theoretical diffusion coefficient

of the fluorescent fusion protein WRN-GFP to be $12.2\frac{\mu m^2}{s}$. Note that this method where we use GFPs measured diffusion in nucleoplasm and scale by protein size accounts for the viscosity inside the nucleoplasm. The shape corrections are used when converting between protein mass and size. The calculations are shown in the Methods subsection 1.4.5 *diffusion coefficient of fusion proteins*.

The experimental WRN diffusion coefficient in nucleoplasm obtained from the FRAP data is much slower ($D_{WRN_{eff}} = 1.62\frac{\mu m^2}{s}$) than the theoretical one ($12.2\frac{\mu m^2}{s}$). Using the slower diffusion coefficient with the fact that the best model is the diffusion model, we can calculate the fraction of free protein from the effective diffusion constant (experimentally measured) and the theoretical diffusion constant. Surprisingly we find that 90% of WRN is bound to chromatin at any time. Since the FRAP data shows full recovery of intensity (see Figure 1.17 C), this large bound fraction must undergo fast binding/unbinding events.

From similar calculations we find that the fluorescent fusion protein BLM-GFP is expected to have a diffusion coefficient of $12.0 \frac{\mu m^2}{s}$ (see Methods *diffusion coefficient of fusion proteins*). However, from our experimental data we find that BLM also has a much lower effective diffusion coefficient $D_{BLM_{eff}} = 1.34\frac{\mu m^2}{s}$. The data is shown in appendix A.1. Both WRN and BLM are constantly and very rapidly binding and unbinding to the chromatin, and at all times 90 % is bound.

We performed the same analysis in the nucleoli and found similar results (see Figure1.18). For both proteins FRAP dynamics were best described by diffusion. A large reduction of the diffusion coefficients ($D_{WRN_{eff,nucleoli}} = 0.12\frac{\mu m^2}{s}$ and $D_{BLM_{eff,nucleoli}} = 0.13\frac{\mu m^2}{s}$) is expected to stem from roughly 99 % of the proteins being bound to DNA but undergoing fast binding/unbinding events.

	Models / Quality	WRN			BLM		
		D	R	R-D	D	R	R-D
Nucleoplasm	AIC_c	**-55.7**	-40.1	-50.7	**-41.3**	-25.3	-35.9
	BIC	**-316.1**	-227.6	-309.8	**-194.0**	-163.9	-188.1
	<RSS>	$4 \cdot 10^{-5}$	$1.4 \cdot 10^{-3}$	$4 \cdot 10^{-5}$	$6 \cdot 10^{-4}$	$2.5 \cdot 10^{-3}$	$6 \cdot 10^{-4}$
	P_{AIC}		$4.1 \cdot 10^{-4}$	0.08		$3.4 \cdot 10^{-4}$	0.07
	P_{BIC}		$6 \cdot 10^{-20}$	0.04		$2.7 \cdot 10^{-7}$	0.05
Nucleoli	AIC_c	**-58.5**	-37.6	-53.3	**-44.6**	-37.1	-39.4
	BIC	**-254.3**	-202.9	-248.2	**-213.4**	-200.9	-207.4
	<RSS>	$1 \cdot 10^{-4}$	$1 \cdot 10^{-3}$	$1 \cdot 10^{-4}$	$7 \cdot 10^{-4}$	$1.1 \cdot 10^{-3}$	$7 \cdot 10^{-4}$
	P_{AIC}		$2.9 \cdot 10^{-5}$	0.07		0.02	0.07
	P_{BIC}		$6 \cdot 10^{-12}$	0.05		$1.9 \cdot 10^{-3}$	0.05
Damage	AIC_c	-31.6	-30.8	-30	-34.3	-30.8	-31.6
	BIC	-181.2	-180.3	-183.8	-193.6	-188.4	-192.8
	<RSS>	$2 \cdot 10^{-3}$	$1.8 \cdot 10^{-3}$	$1.3 \cdot 10^{-3}$	$1.8 \cdot 10^{-3}$	$2 \cdot 10^{-3}$	$1.4 \cdot 10^{-3}$
	P_{AIC}		0.67	0.45		0.17	0.26
	P_{BIC}	0.27	0.17			0.07	0.67

Figure 1.16: **Summary of model selection:** The models highlighted in blue are the optimal models. The optimal model is based on the lowest AIC_c value. Both AIC_c and BIC give similar predictions about the optimal model. P_{AIC} or P_{BIC} indicated among the three models the probability that the less optimal model could be the better model. The diffusion model is the optimal model for both WRN and BLM in nucleoplasm and in nucleoli. Our data on the dynamics at DNA damage is not able to discriminate between the three models. Since the amount of data is the same, this could be interpreted as the dynamics at DNA damage being more complicated than the dynamics in nucleoplasm and at nucleoli. Both AIC and BIC are only relative measures between models; a more absolute measure is $\langle RSS \rangle$, although it does not take the number of parameters into account.
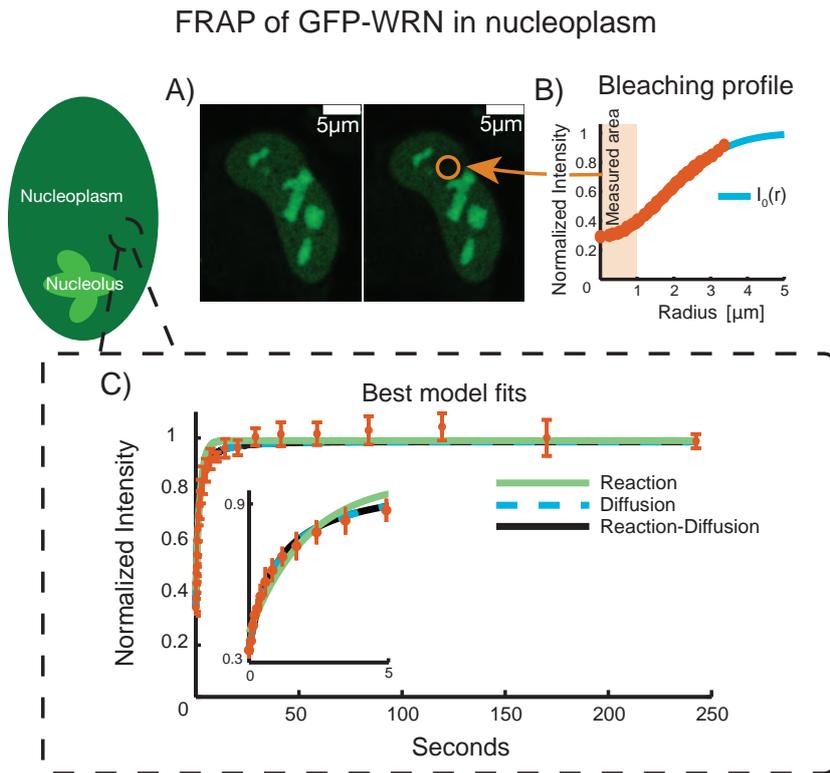
Figure 1.17: **In the nucleoplasm the dynamics of WRN are governed by effective diffusion with** $D_{WRN-eff} = 1.62 \frac{\mu m^2}{s}$ **A)** In cells expressing EGFP-WRN, a 2μm spot is bleached in the nucleoplasm, as indicated by the orange arrow. **B)** To identify the shape of the initial loss of intensity, $I_0(r)$, a Gaussian profile (blue line) is fitted to the measurements of the initial intensity at a given radius from the centre of the bleached region (circles). **C)** Fluorescence recovery curves (circles) and corresponding best fits using: Reaction (solid green line), Diffusion (dashed blue line) and Reaction-Diffusion (solid black line) models. The inset is a zoom-in of the first five seconds. Data points show the averages of five cells and the error bars represent the standard deviations. For BLM we also found that the diffusion model gave the best fit with $D_{BLM-eff} = 0.07 \frac{\mu m^2}{s}$ (see appendix A.1).

FRAP of GFP-WRN in nucleolus
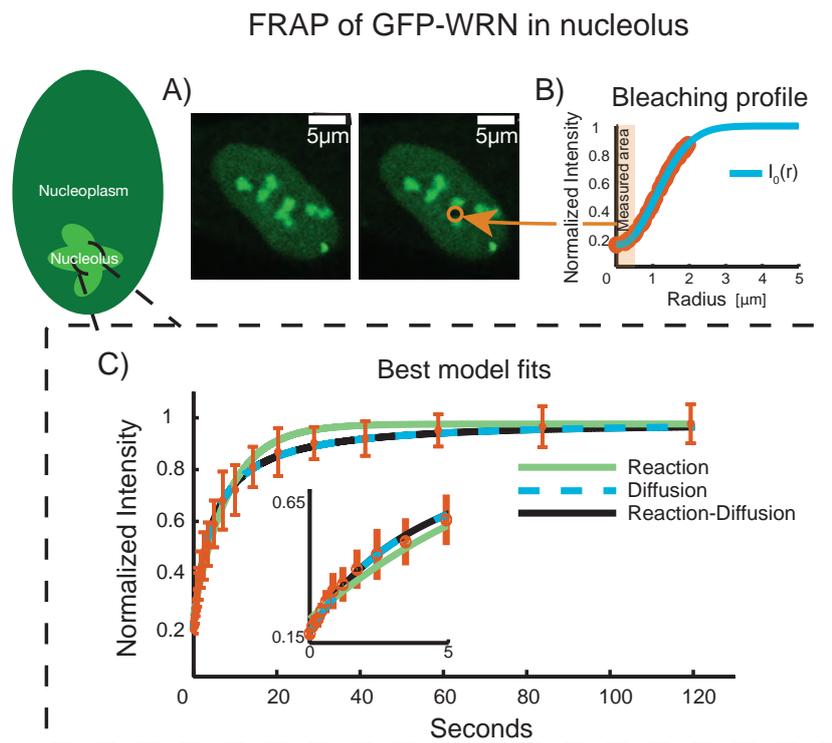
Figure 1.18: **In the nucleoli the dynamics of WRN are governed by effective diffusion with** $D_{WRN-eff} = 0.12\frac{\mu m^2}{s}$ **A)** In cells expressing EGFP-WRN, a $1\mu m$ spot is bleached in the nucleoli, as indicated by the orange arrow. **B)** To identify the shape of the initial loss of intensity,$I_0(r)$, a Gaussian profile (blue line) is fitted to the measurements of the initial intensity at a given radius from the centre of the bleached region (circles). **C)** Fluorescence recovery curves (circles) and corresponding best fits using: Reaction (solid green line), Diffusion (dashed blue line) and Reaction-Diffusion (solid black line) models. The inset is a zoom-in of the first five seconds. Data points show the averages of five cells and the error bars represent the standard deviations.

1.4.7   WRN and BLM dynamics at the damage site have two different timescales

A distinct feature of the data for both WRN and BLM at the site of DNA damage is that the fluorescent recovery curves reach saturation at 80% of the original fluorescent signal. This means that for both WRN and BLM, 20% of the proteins are so tightly bound in the region of DNA damage that they do not unbind within the 1 1/2 minute we imagined. The immobile fraction could represent proteins that are actively involved in DNA repair. The mobile fraction (80%) reaches saturation within less than a minute. This indicates that the time it takes a protein to localize at the site of damage (e.i., the search time) is less than a minute.
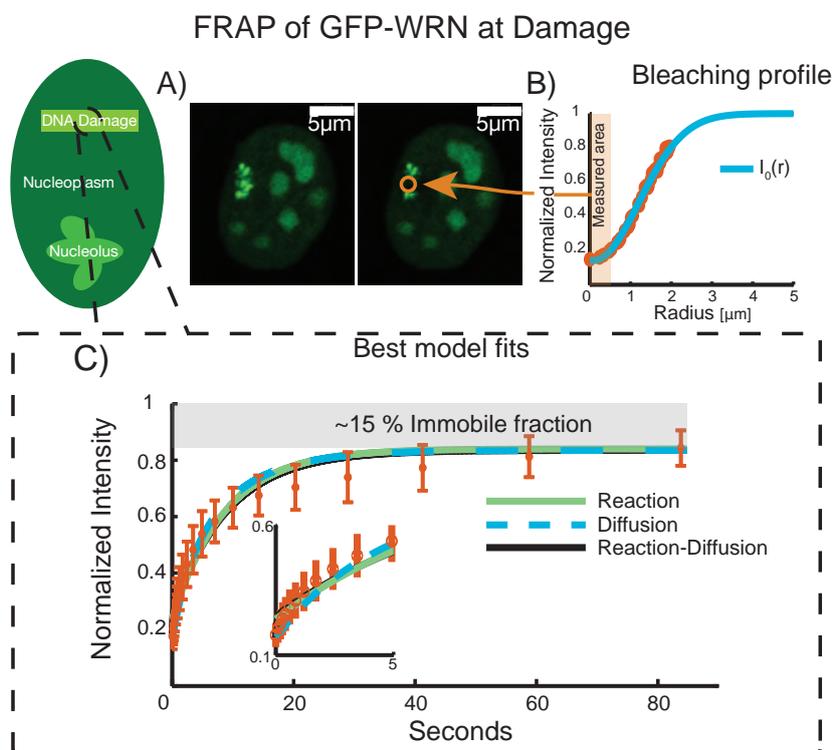
Figure 1.19: **Analysis of WRN dynamics at the site of damage.** At the site of DNA damage WRN (and BLM) have at least two distinct binding modes. We performed the FRAP measurements after the accumulation at the damage sites had saturated, which is roughly two hours (see Figure 4). **A)** In cells expressing EGFP-WRN, a 1 μm spot is bleached in the DSB, as indicated by the orange arrow. **B)** To identify the shape of the initial loss of intensity, I0(r), a Gaussian profile (blue line) is fitted to the measurements of the initial intensity at a given radius from the centre of the bleached region (circles). **C)** Fluorescence recovery curves (circles) and corresponding best fits using: Reaction (solid green line), Diffusion (dashed blue line) and Reaction-Diffusion (solid black line) models. The inset is a zoom-in of the first five seconds. Data points show the averages of five cells and the error bars represent the standard deviations. We were not able to discriminate between the three models (see table 1.16).

So far we have found two dynamics at the damage site with different time scales: the immobile fraction of WRN and BLM that indicates there are dynamics on the scale of minutes or longer, and a diffusion time scale of seconds. In addition to these two dynamics, there is also the accumulation of WRN and BLM at the site of damage.

After DNA damage induction we monitor the accumulation of WRN and BLM to the site of DNA damage. From our data we find that it takes WRN and BLM roughly two-three hours to saturate at the site of damage (see Figure 1.20 A).

The time it takes for WRN and BLM to accumulate cannot be explained by the proteins diffusing and binding directly to DNA damage. The DNA damage is created rapidly after laser irradiation, and our FRAP measurements show that the search time is less than one minute (see Figure 1.19 D). Since the accumulation at damage sites does not saturate after roughly one minute, a gradual creation of additional binding sites for WRN and BLM at the damage site could be dominating the accumulation dynamics. The additional binding sites are in such close proximity to the target of irradiation that they are indistinguishable at the resolution of our microscope ($\approx$250 nm).

Possible biological scenarios for the production of binding sites could be either histone modification or complex formation of repair proteins assembling at the damage site. WRN has already been shown to bind to γH2AX, a stretch of 2000 histones that get phosphorylated upon DSBs [75]. While the histone phosphorylation occurs on a timescale of minutes, other slower histone modifications could be a possible mechanism for the recruitment to the DSBs [76].
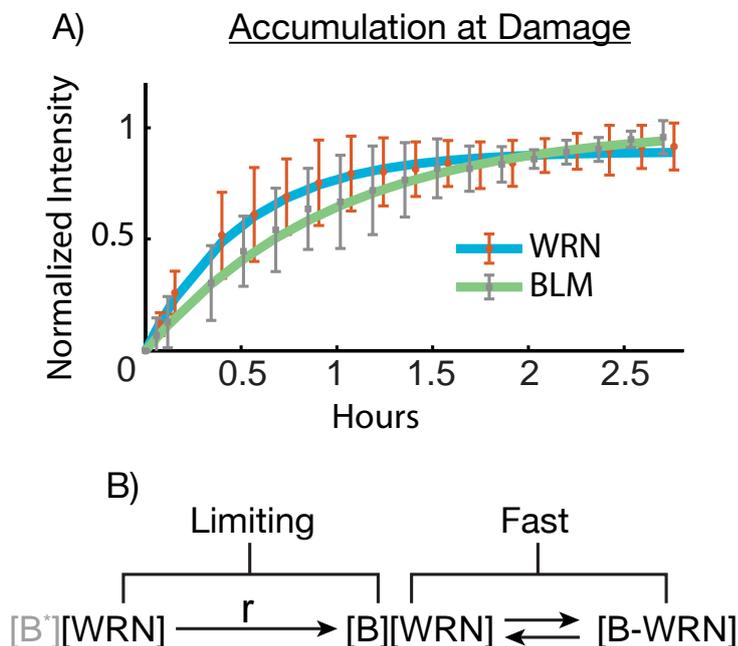
Figure 1.20: **Accumulation of WRN and BLM at DNA damage. A)** Accumulation of WRN and BLM at the site of damage takes two-three hours to saturate. The data is from 21 WRN cells and 9 BLM cells, and the error bars indicate standard deviation. **B)** Our FRAP data (Figure 1.19) show that binding events are rapid at the site of damage and can therefore be neglected. The production of active binding sites is mathematically described by [B-WRN] $= m(1 - \exp(-rt))$. The half-lives obtained by fitting the mathematical model are: $t_{\frac{1}{2}WRN} = 1152 \pm 58s$ and $t_{\frac{1}{2}BLM} = 2310 \pm 154s$ which is roughly 20 and 40 minutes.

We created a simple mathematical model for the accumulation of WRN and BLM. Upon induction of damage B^* potential binding sites are created. These potential binding sites are then converted at a constant rate r into active binding sites for WRN and BLM. WRN and BLM bind rapidly to these new binding sites (B) (see figure 1.20 B). Neglecting the second step, the accumulation can therefore be mathematically described as the conversion of potential binding sites B^* to binding sites B:

$$[\text{B-WRN}] = m(1 - \exp(-r \cdot t)). \tag{1.37}$$

Here m is just a factor to scale the saturation level. The rates are $r_{\text{WRN}} = 0.6 \pm 0.03 ms^{-1}$ and $r_{\text{BLM}} = 0.3 \pm 0.02 ms^{-1}$, corresponding to "half lives" of 20 minutes and 40 minutes.

1.4.8 Classification of DNA repair proteins into either scanners or responders

When comparing DNA repair protein data from the literature, we find that purely from their mobility they could be classified into two groups (see Figure 1.21). In one group where the experimental measured diffusion coefficient is only slightly smaller than the theoretically estimated free diffusion coefficient. This slight decrease can be explained by complex formations, polymerization or a crowded environment in the nucleoplasm.

However, a second group of DNA repair proteins was shown to have diffusion coefficients five times smaller than the theoretically estimated diffusion coefficient. A five-fold difference in diffusion coefficient would correspond to a 100-fold increase of mass. Complex formation and polymerization thus cannot explain the difference between the theoretical and experimental diffusion coefficients. Since proteins of roughly the same size (measured in nucleotides) do not show the same decrease in mobility, crowding in the cell does not seem to explain the large difference in mobility.
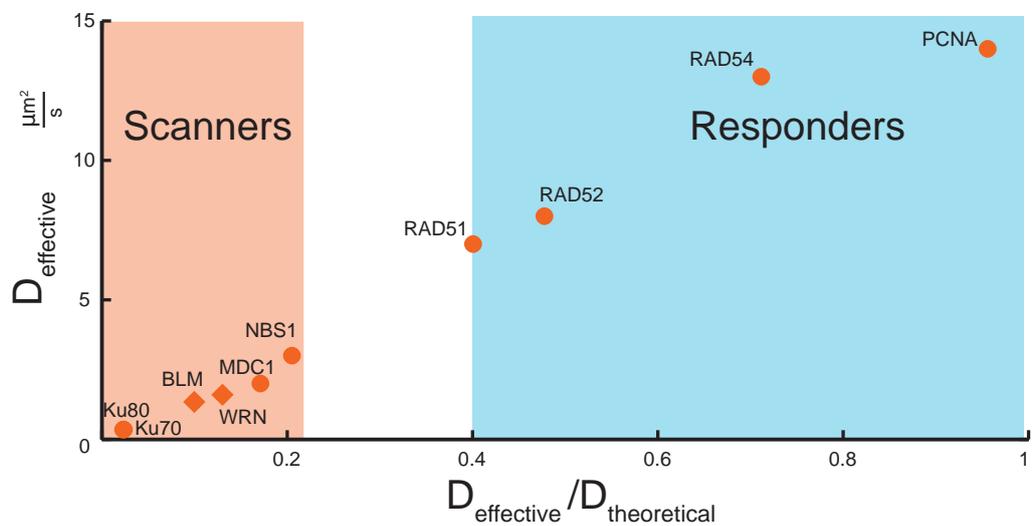
Figure 1.21: **DNA repair proteins can be classified into two major categories: scanners and responders.** Plotted are the theoretical, $D_{theoretical}$, and measured effective, $D_{effective}$, diffusion coefficients for respective DNA repair proteins (for exact values and references, see table in appendix A.1). When proteins do not bind to chromatin or other static cellular components, the ratio between the theoretical diffusion coefficient based on mass and shape ($D_{theoretical}$) and the effective diffusion coefficient determined experimentally ($D_{effective}$) approaches 1.

We find that a reasonable explanation for the large difference in theoretical diffusion coefficients and measured diffusion coefficients is that these DNA repair proteins are mostly bound to chromatin or other static cellular components. We classify DNA repair proteins into two groups, scanners or responders, based on their reduced mobility. We compare the estimated theoretical diffusion coefficient with the experimentally measured effective diffusion in Figure 1.21 and categorize the proteins as follows:

1. $\frac{D_{eff}}{D_{theoretical}} < 0.2$ (Scanners regime, see Figure 1.21) The reduced mobility is due to binding/unbinding with DNA. A ratio of 0.2 or lower would otherwise require a complex formation of ≈ 100 or more proteins. These proteins can potentially scan the DNA for damage

2. $\frac{D_{eff}}{D_{theoretical}} > 0.4$ (Responders regime, see figure 1.21) The reduced mobility could be due to complex formations of 15 or fewer proteins (of same size), or it could be due to imperfect theoretical assumptions (e.g., the shape correction). These proteins accumulate at DNA damage and interact with DNA when other proteins have detected the damage.

1.4.9 Discussion and conclusions

Previous research has noted the difference in diffusion coefficients of DNA repair proteins [77]; however this is the first time they are combined with an expected diffusion coefficient (theoretical diffusion coefficient). The exact regions for when a protein is a potential scanner or responder, are set arbitrarily in our analysis. However, the rationale behind, that is that when a protein scans the DNA for damage its mobility decreases, and that this effect on mobility can be investigated when the theoretical diffusion coefficient is known, is independent on the threshold for the classification.

The decrease of WRN and BLM mobility in nucleoli compared to nucleoplasm is probably due to WRN and BLM participating in ribosomal RNA transcription. The reasons are that ribosomal RNA is transcribed in the nucleoli[78] and WRN that has been show to interact with RNA polymerase [59].

Investigation of the WRN and BLMs dynamics at DNA damage suggest that at least two types of interactions take place. An interesting experiment could be to investigate the dynamics of WRN and BLM in a γH2AX free context, asking whether γH2AX will change the immobile fraction or accumulation time.

Please note that our classification of WRN and BLM as scanners suggests that both proteins should be among the first proteins at the site of DNA damage. However, the accumulation of WRN and BLM at DNA damage saturates after two-three hours. Speculating on this, a possible explanation of these two different observations could be that WRN and BLM have multiple roles. First, WRN and BLM scan the DNA for damage and get recruited to the DNA damage; however, because of the low number of DNA damages, the initial recruitment of WRN and BLM is not detected. Second, WRN and BLM are recruited later in the DNA repair process due to some biochemical reactions, possibly mediated by γH2AX.

Transcription Regulation

2.1 Introduction to Transcription Regulation

Cells respond to environmental stimuli by changing their gene expression profile. Transcription, the first step in the central dogma, is a process in which the RNA polymerase (RNAP) binds to the DNA at a specific sequence called the promoter site. Transcription is a unidirectional process, so from the promoter site the RNAP transcribes the downstream gene. The process of transcription maps the DNA code to the RNA code, which afterwards becomes translated into a protein by the ribosomes.

The logic of regulatory-protein-mediated transcription depends upon whether the regulatory protein is an activator (positive control) or a repressor (negative control). Proteins called transcription factors control expressions of genes. Transcription factors are usually classified as either activators or repressors (see Figure 2.1).

Conditioning the regulation of genes on the activation (or deactivation) of another gene allows allowing for a whole zoology of different conditioned gene expression patterns. These conditioned expression patterns are usually represented by an interaction network where activation is indicated by a normal arrow and repression is represented by a flat-headed arrow (see Figure2.1 **B**). The whole regulatory network is built up by smaller regulatory networks called motifs. The emergence of many complex cellular functions (such as regulation of metabolism, decision making, memory, biological rhythm, and homeostasis) are emerging from motifs that have feedback loops [79]. A positive feedback in a transcriptional motif typically promotes bi- or multistability [80], allowing cells to be in two or more states. Negative feedbacks are widely used to induce stability; however, with a time delay it can result in stable oscillations [81].

In the first chapter we investigated the trancription regulatory effects that RNA polymerase has when two promoters are overlapping. In the second chapter I have investigated the effect of combining the two simplest regulatory motifs, namely a negative auto regulatory (NAR) motif with a positive auto regulatory (PAR) motif.
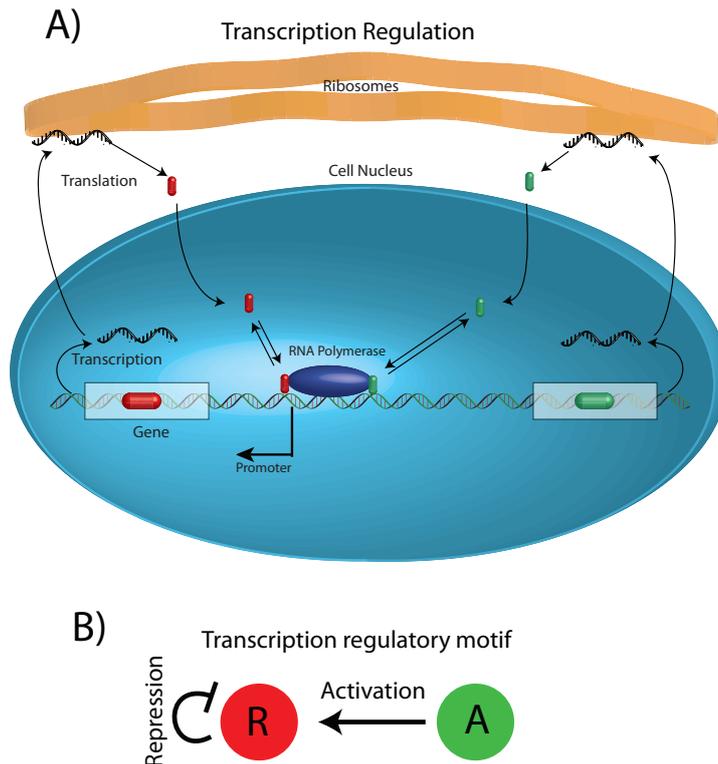
Figure 2.1: **Network representation of transcription regulation: A)** RNA polymerase can transcribe the genes, thereby producing mRNA. The mRNA is then exported to the ribosomes where the mRNA is translated into proteins, the central dogma in biology. If the proteins are transcription factors they can regulate the transcription of other proteins (or themselves). If the transcription is up regulated the protein is called an activator and conversly if the transcription is down regulated, the protein is called a repressor. **B)** A simplified network representation of the biological processes in **A)** where the red protein is a repressor performing self-repression and the green protein is an activator of the red protein.

2.2 Regulatory effects of RNA polymerase for overlapping promoters

As explained in the introduction, transcription factors are usually classified as either: activators or repressors. However some transcription factors are dual functioning regulators, meaning they can act both as activators and repressors. Opposite regulation of two promoters can be achieved when two promoters share a common regulatory region in such a way that one of the two promoters is subjected to positive regulation by the very same protein that represses the activity of the second promoter [82]. In principle, the binding of RNAP to a promoter that overlaps a second promoter can inhibit RNAP binding to the overlapping promoter [83, 84] or interfere with open complex formation at a nearby promoter [85]; RNAP itself can therefore act as a transcriptional regulator for overlapping promoters.

We were interested in finding out if promoters in E.coli are actually overlapping; since the position of the promoter regions in the E. coli genome [86] has recently been mapped this is now possible. Using the promoter positions and the fact that RNAPs occupy 75 basepairs when bound to DNA [87], it is possible to find which promoters are overlapping. Szabolcs Semsey used a slightly more conservative estimate for the overlap (50 basepairs) but still identified 314 promoters that were overlapped by a second promoter. The number of promoters that have overlaps corresponds to 14% of the total number of promoters in the E. coli genome. The orientation of overlapping promoters can be, either head-to-head or tail-to-tail (see Figure(2.2)), but in both cases the overlap would be the same.

From analysing the E.coli transcription unit database we found three different promoter arrangements:

1) Both promoters transcribe a gene and at least one is regulated by a transcription factor.

2) Both promoters transcribe a gene but there is no known transcription regulation by transcription factors.

3) Only one of the promoters transcribes a known gene.

In the last case, since the overlapping promoter does not transcribe a gene, the RNAP seems only to act as a regulator. Further supporting the case for RNAP as a regulator is the fact that a significant proportion of promoters in E. coli are bound by RNAP, which has no transcriptional activity [88].

Because overlapping promoters are common in E.coli we have built a mathematical model to investigate how the level of interference depends on the characteristics of the promoters involved. Additionally, Szabolcs Semsey built a synthetic regulatory region to demonstrate how regulation of one of the promoters can affect the activity of the other.

Traditionally, promoters are classified by strength. The strength is measured by the amount mRNA produced per time. That is, a strong promoter has a higher rate of mRNA production than a weak promoter. The most striking result from this research is that the strength of a promoter does not determine how strongly it will interfere with an overlapping promoter. The promoter characteristic that determines the level of interference, is the amount of time that a RNAP spends bound to the promoter. To capture this property, we introduce a new way of classifying promoters, which we call aggressiveness. This means that promoters can be classified both by their mRNA production (strength) and by their aggressiveness (magnitude of interference).

In addition to this work, a previously related phenomenon namely interference of elongating RNAP, has been modelled [89, 90].

2.2.1   Results and discussion
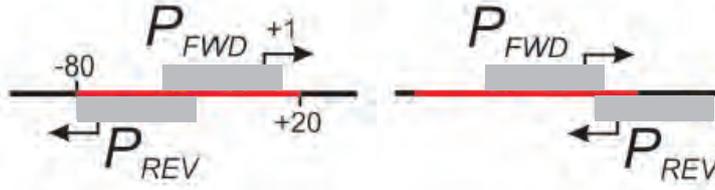
Figure 2.2: **Arrangement of overlapping promoters:** Promoters can be arranged head-to-head (right) or tail-to-tail (left). A promoter occupies 75 bp (from -55 to +20) when bound to the promoter [87]. We consider an overlap between promoters to be from -80 to +20 (indicated by red). Note that this is a more conservative estimate, since this means that a RNA polymerase only occupies 50 bp (from -40 to +10), as indicated by gray boxes. The transcription start of the forward promoter (+1) is used to enumerate the base pair positions. Even with the conservative estimate, Szabolcs Semsey found that 14% of the E.coli promoters overlap.

2.2.2   Mathematical model

If two promoters overlap, a RNAP bound to the promoter of one DNA strand could inhibit the binding of a second RNAP on the opposite DNA strand. We have investigated this process in order to quantify how much interference might be expected by this mutual inhibition. Mathematically the dynamics of the RNAPs' mutual exclusion can be described by a master equation. The system can be in one of three states:

(1) The RNAP is bound to promoter 1, which happens with probability Θ_1

(2) The RNAP is bound to promoter 2, which happens with probability Θ_2

(3) Both promoters are free, which happens with probability $1 - \Theta_1 - \Theta_2$

The system is shown in Figure 2.3.

For this particular system, the master equations governing are as follows:

$$\dot{\Theta}_1 = k_{on}^1 (1 - \Theta_1 - \Theta_2) - k_f^1 \Theta_1, \tag{2.1}$$

$$\dot{\Theta}_2 = k_{on}^2 (1 - \Theta_1 - \Theta_2) - k_f^2 \Theta_2. \tag{2.2}$$

Θ_1 and Θ_2 are probabilities of the RNAP being bound respectively to promoter 1 (P_1) or promoter 2 (P_2). $k_f^{1,2}$ are the firing rates of the promoters and $k_{on}^{1,2}$ is the pseudo on-rate of the RNAP. These equations are the dimensionless equivalent of the biochemical reactions shown in appendix B.2. In our model we assume that RNAP concentration is constant and that RNAP does not dissociate from the DNA when bound. We call the probability of being in state 1 or 2 (Θ_1 or Θ_2) the occupancies.

We measure the interference between the RNAPs as the repression of promoters 1 compared to an equivalent but isolated promoter. For the isolated promoter system there are only two states; either the promoter is bound by RNAP or it is free.

$$\dot{\Theta}_1^0 = k_{on}^1 (1 - \Theta_1^0) - k_f^1 \Theta_1^0.$$
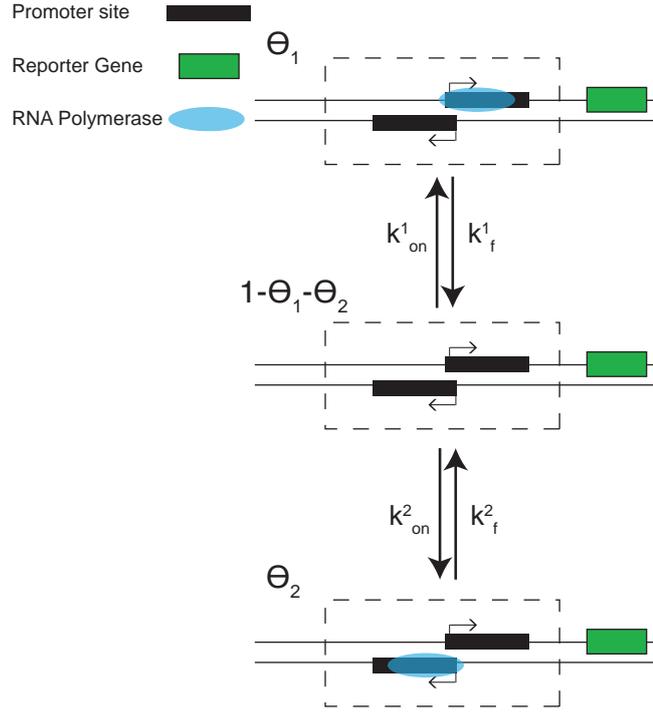
Figure 2.3: **Schematic of the mathematical model:** Spatially the two promoters cannot be occupied by RNAPs at the same time. This creates a situation where one RNAP might repress the other. From a mathematical point of view the simplest model to describe this phenomenon is a three-state Markov chain with transitions between the states as shown in the figure. The probability of being in state one, two and three is given by (Θ_1, Θ_2, $1 - \Theta_1 - \Theta_2$), the rate for RNAP to bind to the promoter is $k_{on}^{1,2}$ and the rate of the RNAP to elongate (or the promoter to fire) is given by $k_f^{1,2}$.

Solving for steady state we get that:

$$\Theta_1^0 = \frac{k_{on}^1}{k_{on}^1 + k_f^1} = \frac{\alpha_1}{1 + \alpha_1}. \tag{2.3}$$

Note that the only parameter that determines the occupancy is the aspect ratio $\alpha_1 = \frac{k_{on}^1}{k_f^1}$. The rate of mRNA production, or the strength of the promoter P_1 is given as:

$$\Omega_1^0 = k_f^1 \Theta_1^0 = \frac{k_f^1 k_{on}^1}{k_{on}^1 + k_f^1}.$$

Solving (2.1) and (2.2) for steady state($\dot{\Theta}_1 = \dot{\Theta}_2 = 0$), the steady state solutions for the overlapping promoters become:

$$\Theta_1 = \frac{\alpha_1}{1 + \alpha_1 + \alpha_2},$$
$$\Theta_2 = \frac{\alpha_2}{1 + \alpha_1 + \alpha_2}.$$

From the occupancies of the overlapping promoter and the isolated promoter (2.3) we can calculate the relative promoter activities of both promoters. Here I only show it for

promoter 1 (P_1):

$$\frac{\Omega_1}{\Omega_1^0} = \frac{k_f \Theta_1}{k_f \Theta_1^0},$$

$$= \frac{\frac{\alpha_1}{1+\alpha_1+\alpha_2}}{\frac{\alpha_1}{1+\alpha_1}},$$

$$= \frac{1+\alpha_1}{1+\alpha_1+\alpha_2}.$$

$$\boxed{\frac{\Omega_1}{\Omega_1^0} = \frac{1+\alpha_1}{1+\alpha_1+\alpha_2}} \tag{2.4}$$

Using (2.3) $\left(\alpha_1 = \frac{\Theta_1^0}{1-\Theta_1^0} \right)$ these equations can be reformulated in terms of the isolated occupancy ($\Theta_{1,2}^0$):

$$\boxed{\frac{\Omega_1}{\Omega_1^0} = \frac{1}{1 + \frac{\Theta_2^0}{(1-\Theta_2^0)}(1-\Theta_1^0)}} \tag{2.5}$$

In Figure 2.4 the interference of promoter 1 is show both as a function of aspect ratios and the occupancies.
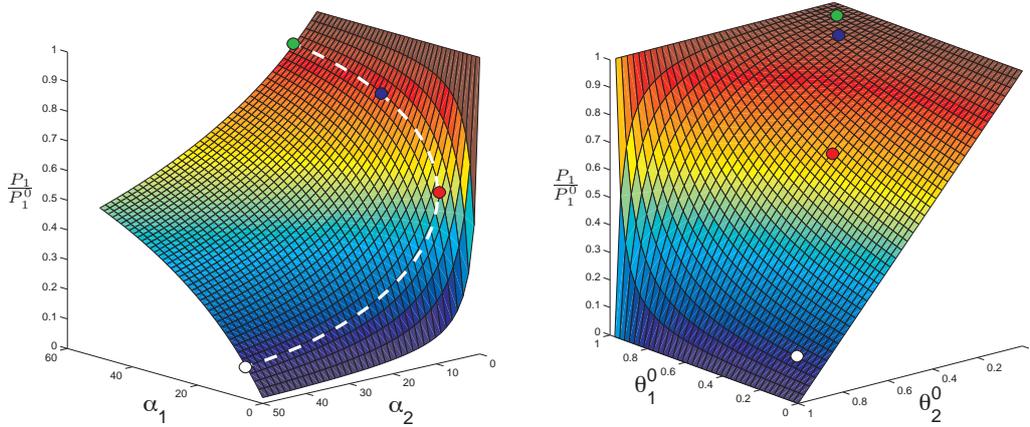
Figure 2.4: **Simulation results for interference of overlapping promoters:** Left panel: Interference of promoter (P_1) activity as a function of its own aspect ration(α_1) vs. overlapping promoter (P_2) aspect ratio (α_2). Right panel: Interference as a function of the promoter's basal occupancy. Interference is measured as the the activity ratio between a non-repressed promoter and the repressed promoter. The four circles shows different examples of promoter pairs: Ranging from almost complete repression (white) to almost no repression (green) with blue and red as intermediates.

Our model predicts that mutual interference of overlapping promoters is independent of the intrinsic strength (Ω^0); it only depends on the aspect ratios. It might be somewhat surprising that a strong promoter and a weak promoter can be equally inhibiting. To account for the capacity of a promoter to inhibit an overlapping promoter we introduce the term called the promoter aggressiveness. The aggressiveness is a measure of the time the RNAP spends at the promoter site.

Our model predicts that the activity of overlapping promoters depends on:

1) The intrinsic activity of the individual promoters.

2) Promoter interference (aggressiveness).

3) The effect of regulatory proteins.

Naturally, this means that the overlapping promoter activities can be regulated in any of these three ways. The simplest case is when there are no regulatory proteins. In this case the promoters are regulated by the RNAP affinity to the promoters. In the case of regulatory proteins, the overlapping promoter complex enables indirect regulation. For example, when a regulatory protein affects only one promoter directly, there will be an indirect effect on the overlapping promoter. Whether the indirect effect activates or represses the overlapping promoter depends on whether the promoter becomes more or less aggressive and not whether it becomes stronger or weaker.

An example could be an activator. The activator can act in two different ways: it can either increase the rate of RNAP binding to the promoter or it can increase the rate of elongation when the RNAP is bound to the DNA. In the first case it would increase aggressiveness of the promoter, hence down-regulate the overlapping promoter. In the second case it would decrease the aggressiveness and therefore effectively up-regulate the overlapping promoter.

Similarly, repressors can decrease the rate of promoter loading by inhibiting RNAP binding or by trapping the RNAP on the promoter. Both would decrease transcription but in the first scenario the aggressiveness is decreased, whereas in the latter scenario the aggressiveness is increased.

We explored these scenarios by simulating how repression or activation changes transcription of overlapping promoters of different nature. Repression of strong promoters was simulated by a ten-fold reduction of the RNAP binding rate (k_{on}^1) or by a ten-fold decrease of firing rate (k_f^1) (see Figure2.5).
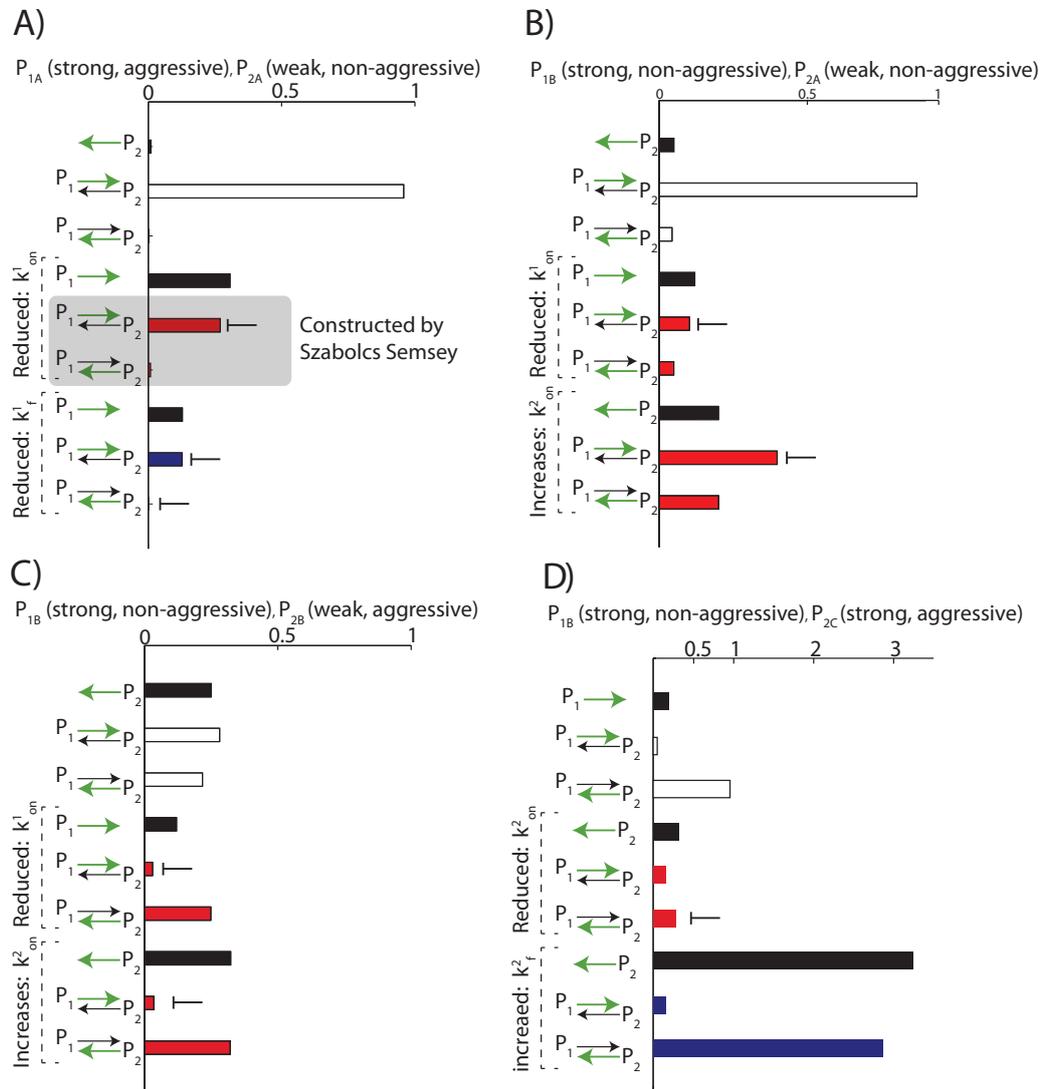
Figure 2.5: **Examples of simulated regulation responses of overlapping promoters:** The bars show the relative activity of promoters labelled with green arrows. The activities are measured as relative to the activity of the strong promoter that is either P_{1A}, P_{1B} or P_{2C}. Black bars show intrinsic promoter activities, which mean no overlap. White bars show promoter activities when the promoters overlap. Red bars indicate regulation in different direction (one promoter is up-regulated, while the other is down regulated): the down regulated promoter is indicated by an repression arrow. Blue indicates regulation in the same direction. The parameters for the unregulated promoters were: $P_{1A} = P_{2C}(10s, 75\%)$, $P_{2A}(200s, 15\%)$, $P_{1B}(10s, 15\%)$, $P_{2B}(200s, 75\%)$, where the first parameter is the inverse firing rate $(\frac{1}{k_f})$ measured in seconds. The second parameter is the occupancy (Θ_0), the fraction of time spent on the promoter. Note that k_f regulation generally changes activities in the same direction, whereas k_{on} regulation changes activities in opposite directions.

In general, changing the RNAP binding rate (k_{on}) has an opposite effect on the activities of the two promoters, whereas both promoters activities are changed in the same direction when the firing rate is altered. Opposite regulation obtained by direct and indirect regulatory effects can result in a transcriptional switch. However, a successful transcriptional switch requires a strong effect of the opposite regulation on both promoters combined with a large difference in activity for the overlapped promoters in the absence of the regulatory protein. Put in another way: without regulation, promoter 1 could be "on" and promoter 2 could be "off", while the situaion is completely opposite when the regulatory protein is bound; that is promoter 1 should now be "off" and promoter 2 should be "on". There are several ways to obtain high activity differences (on/off behaviour), including:

1) Overlapping a strong aggressive promoter (P_{1A}) with a weak non-aggressive promoter (P_{2A})

2) Overlapping a strong non-aggressive promoter (P_{1B}) with a weak non-aggressive promoter (P_{2B})

3) Overlapping a strong non-aggressive promoter (P_{1B}) with a strong aggressive promoter (P_{2C})

In the first case, repressing the on-rate by a regulatory protein directly decreases the activity and aggressiveness of the strong promoter (P_{1A}), thus allowing transcription of the overlapping non-aggressive promoter (see Figure 2.5 A rows 5 and 6).

In the second case, increasing the on-rate (k_{on}) of the weak promoter (P_{2B}) increases its activity and makes the promoter more aggressive, thus repressing the overlapping strong and non-aggressive promoter (see Figure 2.5 B rows 8 and 9). However, in this setup repressing the strong promoter P_{1B} by decreasing the on-rate does not affect the activity of the overlapping weak promoter. Hence, individual regulation can be achieved through indirect regulation.

In the third case, reducing the on-rate of the aggressive promoter (P_{2C}), cause the aggressive promoter to decrease two-three fold, and the overlapping strong non-aggressive promoter to increase roughly three-fold.

These examples show that direct and indirect regulation can affect promoter activities to different extents, and allow practically independent regulation of one of the promoters. In certain cases the regulated promoter remains unchanged while the indirect effect on the overlapping promoter is significant (see Figure2.5B row 7, 8 and 9). This independent regulation could happen, by increasing the RNAP binding rate (k_{on}) for a weak promoter, which is limited by the firing rate. The increase of k_{on} has little effect on the activity of the promoter. However, it increases the time RNAP spends on the DNA, which increases the promoter's aggressiveness, (see Figure 2.5 B). The possibility of independent regulation of overlapping promoters can contribute to the compact genome organization of prokaryotes.

In this model we have neglected the time it takes the RNAP to clear the overlapping region. Since the promoter overlaps can be arranged in two different ways, either head-to-head or tail-to-tail (see Figure 2.2), the time it will take to clear the overlapping region will be different. For a tail-to-tail configuration, this clearance time is effectively zero. A larger discrepancy will be when the promoters are in a head-to-head configuration. It will take roughly two seconds to clear a head-to-head configuration assuming a speed of 50 nucleotides/s [91]. The clearance time will increase the aggressiveness of the pro-

moter, so head-to-head configurations will be more aggressive than our model predicts. Another detail that has not been included in this model is the distinction between binding to the DNA and formation of an open complex. These details could be included in the model, but it would be at the expense of introducing more parameters.

2.2.3 Synthetic regulation

Our simulation suggested that overlapping promoters can provide indirect gene regulation, which can be explored by construction of synthetic genetic circuits. For example, a non-aggressive promoter can be inhibited when overlapped by a strong and aggressive promoter. Furthermore the non-aggressive promoter can be activated (de-repressed) by a repressor, which inhibits RNAP binding to the strong promoter. The potential advantage of this setup is that practically any DNA binding protein can be used for repression of the aggressive promoter, and no direct action is required for the activation of the non-aggressive promoter.
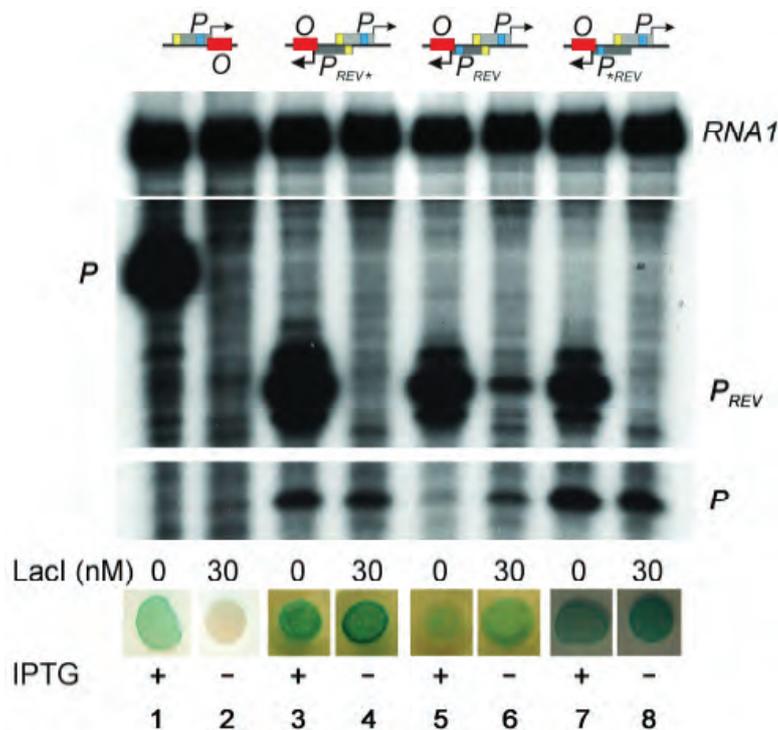
Figure 2.6: **In vitro and in vivo regulation of overlapping promoters:** Schematic drawings of regulatory regions are shown on top. The results of in vitro transcription from the promoter P and P_{REV} in the present and absence of LacI are shown below each drawing. The RNA1 transcript, which is not affected by LacI binding, was used as an internal control between lanes. The regulatory regions were inserted into the E.coli chromosome in such orientation that the promoter P transcribes the $uidA$ reporter gene, that encodes for β-glucuronidase. Expression of the reporter gene in the presence and absence of IPTG is indicated by the blueish-coloured colonies, which result from degradation of X-gluc by the β-glucuronidase enzyme.

To test this, Szabolcs Semsey constructed the synthetic systems shown in Figure 2.6. The promoter sequences with most influence on promoter activity are positions

-8 to -13 and -32 to -37, which are also known as the -10 and -35 hexamers. The -10 and -35 hexamers are indicated respectively by blue and yellow. Szabolcs Semsey overlapped a promoter (P) that has a weak -35 and a consensus -10 element (weak, non-aggressive) with a reverse promoter P_{REV}, that has consensus -10 and -35 elements and the extended -10 element (strong, aggressive). The transcription start site of P_{REV} overlaps with a symmetric *LacI* operator site (O). The position of the operator site was chosen to be neutral for the transcription of P [92]. This synthetic construct corresponds to the simulations in figure 2.5 A.

As a first step, activity of P was assayed qualitatively *in vivo* using a P promoter-*uidA* (encoding for β-glucuronidase) transcriptional fusion in the presence and absence of IPTG (see Figure2.5). LacI is a repressor for the Lac-operon, and since IPTG binds to LacI, it effectively activates the Lac-operon. For the construct with no overlapping promoter, we observed a decrease in P activity in the presence of IPTG (see lanes 1 and 2). For the second construct, where P is overlapped by P_{REV} and P_{REV} is regulated directly by LacI, we show that P is inactivated in the presence of the inducer IPTG (see lanes 5 and 6).

To understand the system more quantitatively, we measured activities of both P and P_{REV} in the presence and absence of LacI *in vitro*. Figure 2.6 shows that P_{REV} inhibits P activity, and that this inhibition depends on having an accessible transcription start site in P_{REV} by binding to the overlapping *lac* operator site (see lanes 5 and 6). We expect a four-fold higher activity of P in the presence of LacI (figure 2.5 A), which is roughly what we get. This demonstrates that the repressor LacI can indeed function an activator of P through the repressing of P_{REV}.

Szabolcs Semsey further found that a single base pair mutation in the -10 sequence P_{REV*} or a two base-pair mutation in the -35 element P_{*REV} of P_{REV} results in loss of repression of P (compare lanes 3 and 4, and lane 7 and 8).

2.2.4   Summary

We have introduced a new term for promoters, namely the aggressiveness. The more aggressive a promoter is, the more capable it is of inhibiting an overlapping promoter. From our experimental results we conclude that given a proper sequence, RNAP binding to one promoter can repress an overlapping promoter P. Inhibition of RNAP binding to P_{REV} by a DNA binding protein (LacI) can activate the transcription of P, resulting in a compact and simple activation-repression switch (which was expected by our model).

2.3 Regulation of oscillations and stability by coupled positive and negative feedback systems

From previous work it is known that a negative autoregulation motif with a time delay can cause protein levels to oscillate [93, 94, 95]. Additionally, the frustrated bistability motif (FBM), a transcriptional motif consisting of a positive autoregulation motif coupled with a repressor, has also been shown to produce oscillations [96, 97].

These two motifs can be combined into a motif we dub the Negative Autoregulated Frustated bistability motif, or for short the NAF motif (see Figure 2.7 A). The NAF motif occurs in biology in both memory formation [98], and cell differentiation [99, 100]. The simplicity and biological relevance of the NAF motif makes it interesting to investigate. Previous theoretical work on the NAF motif has only considered protein levels, thereby course-graining transcription and translation events into a single-step process [98, 99, 100].

In the following sections we include mRNA levels in our model, but course-grain mRNA import and export events with an explicit time delay. Experimentally the NAF motif can be engineered using tetracycline transcriptional regulatory (TET) elements. Most of the interactions between the TET system components have been characterized and the dynamics of various TET-based synthetic networks have been recently simulated [101]. Closely following the biochemical restrictions of the TET system allows our model to have predictions that are easy to test with experiments.

The dynamics originating from a motif are often parameter-dependent. Thus, estimating the dynamics caused by the motif without knowing the exact value of the parameters is often not possible. However, using recent data on half-lives and transcription- and translation-rates from 5000 different proteins [102], allows us to make reasonable guesses about expected dynamics.

Surprisingly, we find that the dynamics of the motif only produces sustained oscillations in a limited parameter regime, where both the activator mRNA and activator protein are short lived. However, the vast majority of transcription factors do not meet these requirements.

2.3.1 The NAF-model

The dynamics of the system are given by the following equations:

$$\frac{dmRNA_A}{dt} = \langle r_1(\tau) \rangle_{(A,R)} - \gamma_{1,A} \cdot mRNA_A$$

$$\frac{dmRNA_R}{dt} = \langle r_1(\tau) \rangle_{(A,R)} - \gamma_{1,R} \cdot mRNA_R$$

$$\frac{dA_{tot}}{dt} = r_2 \cdot mRNA_A - \gamma_{2,A} \cdot A_{tot}$$

$$\frac{dR_{tot}}{dt} = r_2 \cdot mRNA_R - \gamma_{2,R} \cdot R_{tot}$$

Where $mRNA_{A,R}$ are the activator and repressor mRNA levels, and A_{tot}, R_{tot} are the total amounts of activator and repressor. r_2 is the translation rate (assumed equal for activator and repressor). γ is the degradation rate and $\langle r_1(\tau) \rangle_{(A,R)}$ is the transcription rate which is a non-linear function of the repressor R_{tot} and the activator A_{tot}. τ represents the time delay between production of mRNA, its modification and export [103, 104]. We chose not to include a similar delay in the protein equations because protein import is very fast [105] The derivation of the model can be found in appendix C.2.3.

The following assumptions were used for deriving the model:

1. The binding/unbinding of transcription factors to operator sites and dimerization occurs on a timescale much faster than other processes [106] and is therefore assumed to be in quasi-equilibrium.

2. There is no cooperativity in binding to operator sites [107].

3. RNA polymerase (and Ribosome) levels do not become a limiting factor even at high expression rates.

2.3.2 Classifying dynamical behaviour of the NAF model

We distinguish between three types of dynamics: sustained oscillations, damped oscillations and no oscillations. We used a modifying version of the matlab function *findpeaks* to find peaks in the time series of the repressor. To ensure that peaks originating from numerical errors were discarded, we only included peaks with at least a two-fold amplitude. Simulations were run such that the length of each time series corresponded to 50 days. The dynamics were classified as follows:

- **Sustained oscillations:** 10 or more oscillations

- **Damped oscillations:** 9 or fewer oscillations

- **No oscillations** No peaks

There is agreement between this classification and the linear stability of the fixed points (compare Figures 2.9 A, C.). The fixed points for the linear stability analysis were numerically computed using Mathematica.
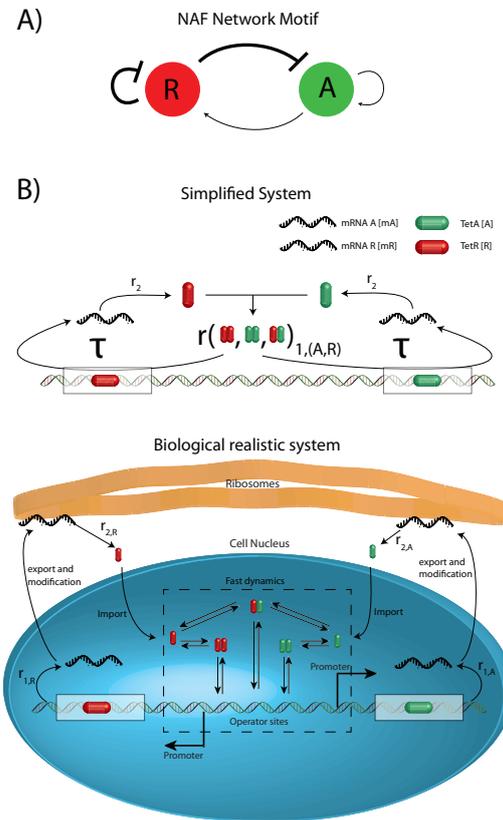
Figure 2.7: **Schematic overview of the NAF model**
A) The NAF motif consist of an autoregulated repressor coupled with an autoregulated activator. **B)** The biologically realistic system involves several steps. Which we simplify the system as follows. Starting from the top left, mRNA at ribosomes are translated into proteins with a rate r_2, the import of proteins is very fast, so it is neglected [105]. Second, the chemical reactions are fast and both dimerization and protein-DNA binding is therefore treated as in an equilibrium [106]. The last step, in the simple model is transcription of mRNA. This step covers transcription of mRNA, as well as modification and export of mRNA from the cell nucleus to the ribosomes. The transcription rate is regulated by dimers (activator-activator, repressor-repressor, activator-repressor) binding to the operator sites and the modification and export of mRNA is simplified by a 30-minute time delay (τ) [103, 104]. The simplified system can be schematically repressed by the NAF motif in **A)**.

2.3.3   Parameters for the NAF model

The applied transcriptional rates, translational rates and half-lives of mRNA and protein are based on a previously reported measurements for about 5000 mammalian genes [102]. Dissociation constants for dimerization and for protein-DNA binding were based on previous estimates for TET elements [101] (see appendix C.1). We assumed a maximum of a 100-fold decrease (or increase) in transcription for saturating levels of repressors (or activators).

2.3.4   Time delay from mRNA export has a limited effect on dynamics

Scanning the four dimensional parameter space of half-lives, we find that there is no effect of a 30-minute time delay within biological realistic parameters (see Figure 2.8). The biologically relevant half-lives range from 0.01 to 1000 hours [102]. For each scan the repressor and repressor mRNA half-lives were fixed and the half-lives of the activator and mRNA activator were varied. The parameter ranges were as follows:

PARAMETER RANGES:	Lower limit [hours]	Upper limit [hours]	Fixed [hours]
Fast mRNA/Protein	0.01	1	0.5
Medium mRNA/Comparable Protein	1	25	12.5
Slow mRNA/Medium Protein	25	100	50
Slow Protein	100	1000	500

Table 2.1: **Half-life parameter ranges:** The repressor and repressor mRNA were kept fixed, while the activator and activator half-lives were scanned.

A previous estimate is that a delay should be roughly twice as long as the time scale for protein degradation before the time delay has an effect [95]. It is thus only surprising that for the very short-lived proteins and mRNAs the time delay does not have any effect for the NAF motif.

This does not mean that an explicit delay has no effect at all. If activation is removed (e.i., the transcription of the activator is set to zero) and the repressor (and mRNA) are short-lived (four minutes), the explicit time delay can cause oscillations (see Figure 2.12 A). The equations are different but the motif is the same as the Hes Oscillator analysed by Jensen et al. [93], where oscillations were also caused by an explicit time delay.

From the large parameter scan (see Figure 2.8) it is clear the NAF motif produces oscillations that only in a very limited parameter range. In the following sections we closely investigate the NAF motif within the following parameter range: repressor and repressor mRNA half lives fixed to 9 hours and activator and activator mRNA half lives ranging from 0.2 to 50 hours. This corresponds to (Medium, Medium, Comparable, Comparable) in Figure 2.8 .

2.3.5   Unstable activators and mRNAs are needed for sustained oscillations

The half-life range for the activator and its mRNA, that produces oscillations is roughly from half an hour to one-and-a-half hours. It is interesting to note that since the half-lives that produce oscillations are bound both from the top and the bottom, activation is needed for oscillations, although it needs to have fast dynamics. In other words a pure
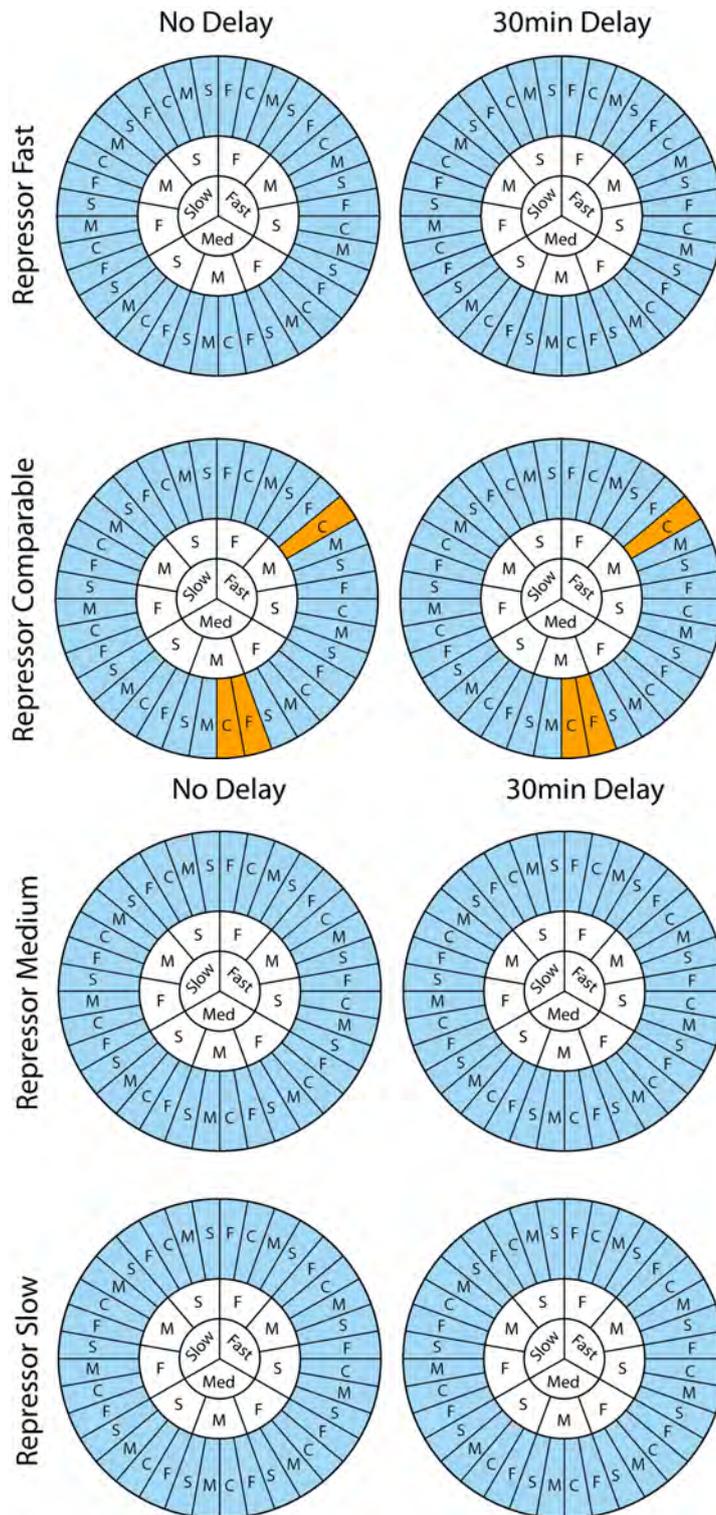
Figure 2.8: **Half-life parameter scans in the interval 0.01 - 1000 hours:**
Scanning the half lives shows that areas with oscillations (orange) and the regimes with
no oscillations (blue) are the same for no delay and for 30min time delays. The innermost
ring is the activator mRNA half-life, the middle ring is the repressor mRNA half-life and
the outermost ring is the activator half-life.

Negative Autoregulation motif with dimer repression would not produce oscillations in a biological realistic parameter regime.

As stated previously we find agreement between the linear stability of the fixed points and the numerical simulation (see Figure 2.9 A, C.) Since the amplitude of the oscillations change continuously, this indicates that the bifurcation is a supercritical Hopf bifurcation.

We find that the system in the non-oscillatory regime is monostable (see Figure C.6). This is interesting from a biological point of view, since it means that the NAF motif with our equations, within biological relevant parameters cannot function as a switch (This requires bistability).

2.3.6   Intermediate repressor stability needed for oscillations

We observe that oscillations can only occur in an intermediate range of repressor and repressor mRNA half-lives (see Figure 2.10A). Both the repressor and mRNA need to have half-lives between 5 and 25 hours. The majority of proteins have longer half-lives than 25 hours [102], which makes it less plausible that the NAF motif, would produce oscillations in nature. We also scanned transcription rates and again found that oscillations only appears within a limited parameter range (see appendix C.2).

2.3.7   Experimental data for regulatory proteins suggest NAF is not oscillatory

Identifying transcription factors in the data set from Schwanhausser et al. [102] and overlaying the transcription factors and their corresponding mRNA half-lives on our parameter scan shows that most transcription factors fall into the non-oscillatory regime (see Figure 2.11). The function of transcription factors is often context-dependent [108]; therefore we did not discriminate between activators and repressors. The lack of transcription factors falling into the oscillatory (or even damped oscillatory regime, and the fact that the non-oscillatory regime is monostable strongly suggest that the NAF motif with our equations predominantly promotes stability.

2.3.8   NAF motif; the combination of the negative autoregulation and frustrated bistability motif

We previously showed that for certain parameter sets, the NAF motif reduces to the NAR motif (Hes oscillator) and shows similar dynamics (see Figure 2.12 A). However when we introduced activators (even unstable ones), the activators cancel the oscillations (see appendix Figure C.3). This means that in the regime where the NAR part of the NAF motif can oscillate on its own the NAF motif cannot; conversely in the regime where the NAF motif can oscillate, the NAR motif cannot.

The NAF motif can be reduced to the NAR motif simply by changing a parameter; however, it is not possible to reduce the NAF motif to the frustrated bistability motif (FBM) in the same way. As an alternative I constructed a version of the NAF motif using the same equations as Krishna et al. used for the FBM[96], but with a self-repression term (see appendix C.2.4). These equations only take protein levels into account. When contrasting the two motifs we find that the parameter space where the NAF motif oscillates is substantially smaller than the parameter space for FBM (see Figure 2.12 B).
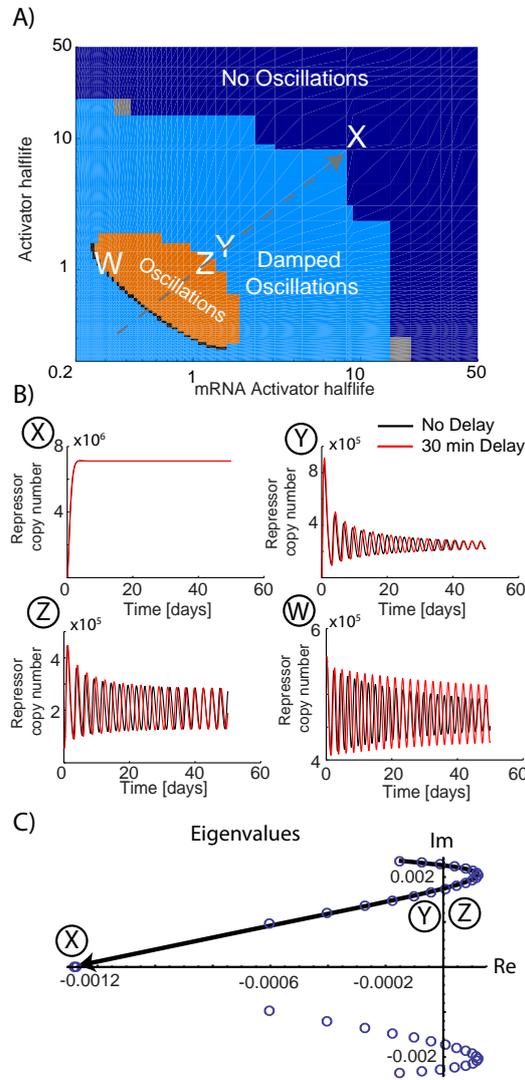
Figure 2.9: **Dynamics of the coupled NAF system:**
A) A parameter scan of the asymmetry where the repressor protein and mRNA half-lives are fixed at nine hours and the activator protein and mRNA half-lives are varied from 12 minutes to 50 hours. The effect of a 30-minute time delay is negliglible, since the delay actually changes the behaviour only in a very limited range of parameters (**W**). **B) (X)** For the symmetrical case, we find no oscillations as expected from Figure C.1. We find that the delay influences the period, increasing it by about 10%. **C)**. There is good agreement between the linearized analysis **C)** and the simulations **A)**.
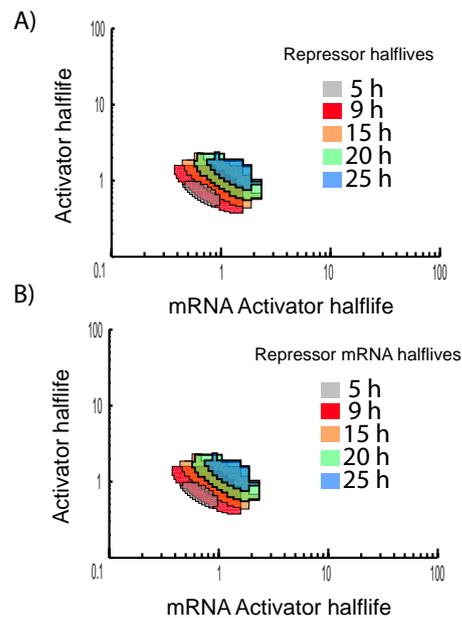
Figure 2.10: **Effect of repressor mRNA and protein half-lives on oscillations**
Scanning for half-lives from two to 35 hours shows that the system oscillates for half-lives above two hours and below 30 hours. We see that the parameters for oscillations are completely symmetrical for protein **A)** and mRNA **B)** half-lives. Note that the parameter range that allows oscillations is bigger for the repressor (5-25 hours) than for the activator (0.5-1-5 hours). These ranges are of biological interest; since the median of protein half-lives is 46 hours, it suggests that the system is not prone to having sustained oscillation.

This further supports that our NAF motif might not be a good motif for producing oscillations.

2.3.9   Comparison with previous studies

Even though the NAF motif is a combination of two motifs that are known to produce oscillations, we show that in a biologically realistic parameter range our NAF motif, promotes stability. Since a motif does not uniquely describe protein interactions, we can only state that the NAF motif governed by our equations promotes stability within biologically relevant parameters.

The NAF motif has been previously explored theoretically [98, 110, 111, 100]. None of the dynamical equations for these NAF motifs are the same, but qualitative comparison should still be possible. Common to all of the papers except that of Suel et al. is that the degradation is passive. Interestingly except for Suel et al. all of the NAF motifs require the repressor to be more stable than the activator to produce oscillations. This is in agreement with our findings. The two studies that explore the effect of half-lives [98, 110] also find that for oscillations to occur, the repressor half-life needs to be in a bounded regime, not too low nor too high. Again this is in agreement with our findings.

Song et al. find that in addition to producing oscillations, their NAF motif is bistable, although only in a very limited regime [98]. On the contrary Hasty et al. only find their NAF motif to be in a steady state when not oscillating [110]. Like Hasty et al. we did not
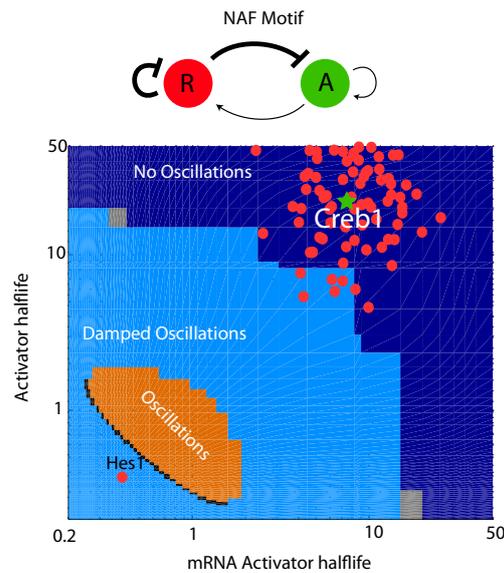
Figure 2.11: **Biological regime**
When plotting the half-lives for transcription factors reported by Schwanhausser et al. [102], we find that almost none fall into the oscillatory regime. Since repressors and activators often are context-dependent we do not discriminate between repressors or activators. Plotting the half-life of Hes1, (a negatively autoregulated repressor having oscillatory behaviour), shows that some transcription factors actually fall into the regime of (damped) oscillations [109]. However, CREB1, an activator involved in a NAF motif falls into the regime of no oscillations (green star) [98]

find any bistable regime (see appendix Figure C.6). However, since the regime found by Song et al. that produces bistability seems rather small, it could simply be that we have not found this regime. A notable difference between the equations behind Song et al. and our NAF motif is that our proteins both interact at the transcriptional level and form heterodimers. That is activators and repressors bind to each other and form dimers. In contrast the proteins described by Song et al. only interact at the transcriptional level [98]. Since our transcription rate is a function of number of repressors and activators bound, the heterodimer formation effectively causes the transcription to depend on the monomer levels. This effectively reduces the non-linearity of the system. If the logic is changed such that repressors are dominant, the non-linearity is recovered and we see that this can produce oscillations (see appendix Figure C.5). Changing the equations such that heterodimers are not formed also increases the previous range of sustained oscillations from 0.5 hours to 1.5 hours to roughly 0.5 to 3 hours (see appendix Figure C.4). However, we still did not find a bistable region in the parameter space we checked.

2.3.10   Conclusions

We find that the NAF motif with our equations has the ability to oscillate, however within biological relevant parameter ranges this probably does not happen. Contrasting our study with previous theoretical NAF motif studies indicates that a general requirement for all the NAF motifs to produce oscillations is that the activator (and activator mRNA) needs to be more unstable than the repressor. Additionally we found that the
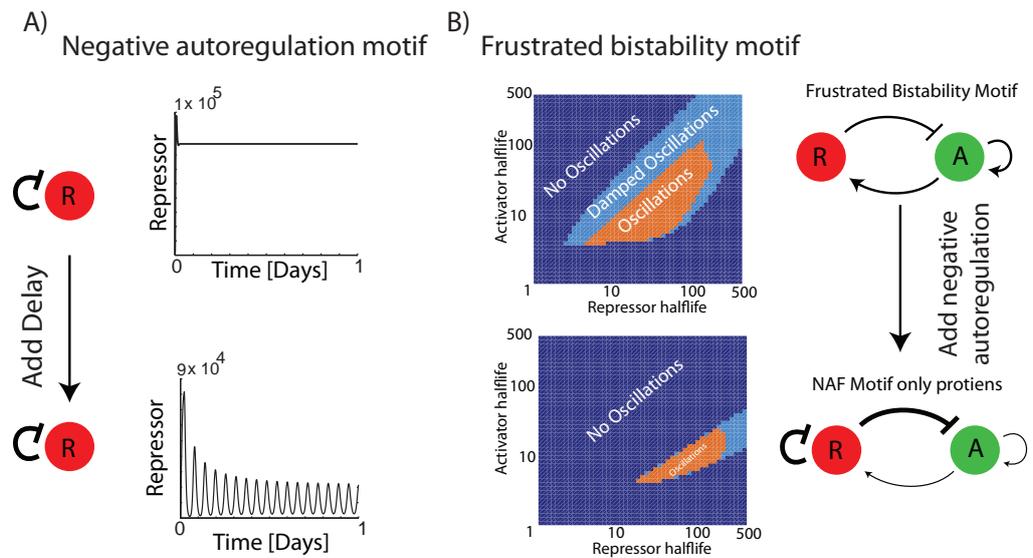
Figure 2.12: **Motif Comparison**
A) The Negative Autoregulation motif (NAR), can be recovered by setting activator transcription to zero $r_{2,A} = 0$. We find that when half-lives of repressor and repressor mRNA are short lived a 30 minute explicit timedelay can produce oscillations. Parameters changed $r_{2,A} = 0$, $\beta_3 = 0.16\text{min}^{-1}$, $\beta_4 = 160\text{min}^{-1}$. This is qualitatively similar to dynamics of the Hes oscillator[93]. **B)** We modified the previous frustrated bistability motif (FBM) equations to allow for autorepression, thereby creating a protein level NAF motif. When contrasting the original FBM with the protein level NAF we find the regime with oscillations is smaller for the NAF motif. See parameters and equations in supplementary C.2.4

repressor needs to been in an intermediate range of stability. Although experimentally constructing the NAF motif using TET elements should produce interesting dynamics, it would offer limited insight into other NAF motifs found in biology. The reason is that the dynamical regime is far away from the biologically relevant parameters.

Communication Networks

3.1 The Expert Game

There is increasing interest in the formation of communication and social networks. This is partly driven by access to large digitalized datasets [112, 113, 114]. Interesting conclusions have been derived from these large datasets, such as that the limit of emails sent per person saturates slightly above 100 emails per day [112]. How these communication networks emerge is however inherently difficult to infer from these datasets. The two primary reasons are that first, usually only a subset of communication channels are monitored (e.i., email communication is monitored but face-to-face or phone communication is not), and second that the network is never monitored from the very beginning, (e.i., a network always exists prior to the first measurement). In this study we investigate the initial formation of a communication network, starting from a completely blank sheet. As a complementary approach to the big data sets, we constructed a tool called "The Expert Game". "The Expert Game" is a simplistic game where N players send standardized email to each other. The game is a repeated game in our case it was played roughly 25 times per session. Using this tool we were able not just to monitor the emergence of a communication network, but also to quantify the information flow and generation of social capital within the network.

3.1.1 Rules of The Expert Game

"The Expert Game" is a game played by *N* players, each of whom is assigned a task and an expertise. Tasks and expertises are matched such that each task is uniquely matched with an expertise. The goal of the game is then for each player to find and get help from their expert. This can only be achieved through sending standardized electronic messages. In each round of the game, every player can at most send one message. The four types of standardized messages are as follows:

- (I) Inquiry: A request for help that contains the sender's task and expertise.

- (C) Confirmation: Confirms that the sender's expertise matches the receiver's task.

- (R) Referral: The sender tells the receiver the name of the expert for the receiver's task

- (N) Negation: The sender does not know who the expert for the receiver's task is.

The last three (C,R and N) are replies to an inquiry and can only be sent in reply to such. Lies are not possible, which makes the replies mutually exclusive. A game consists of on average six rounds, but the number of rounds is chosen randomly for each game, which makes it impossible for the players to predict the number of rounds in a given game. A

monetary reward (roughly three euros) is only given to the receiver of a confirmation. The round limitation mimics real life time constraints, and forces players to prioritize between communication partners and types of messages. An example of a game with four players is shown in Figure 3.1.
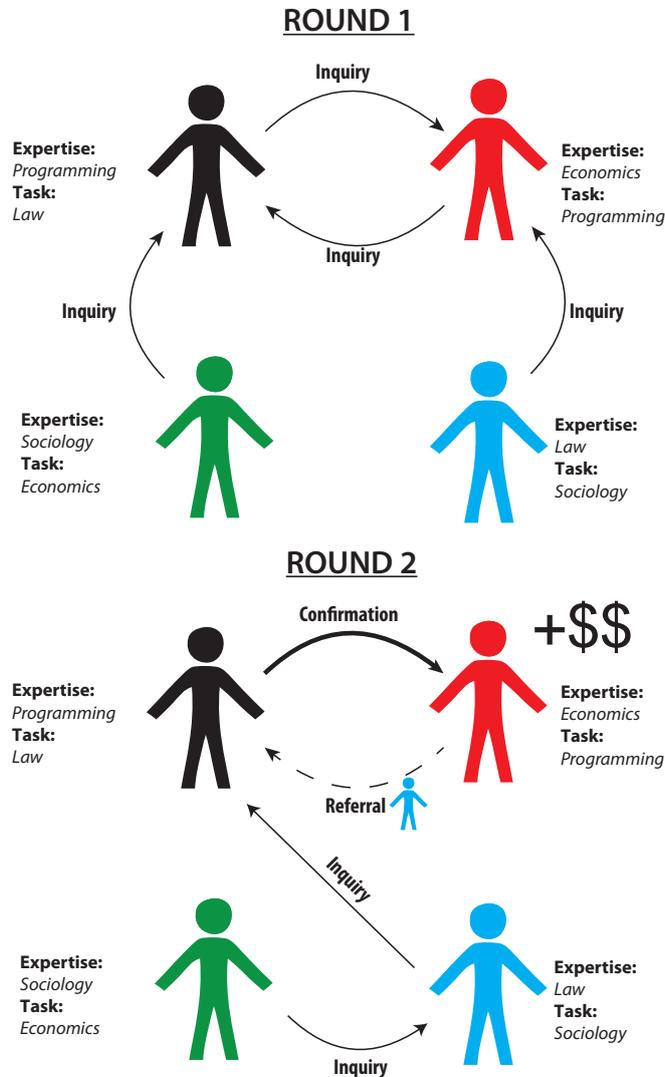
Figure 3.1: **Example of "The Expert Game".** All message sending is synchronized so that during each round players can send at most one message.

Round 1: In the first round only inquiry messages can be sent. After round 1 the black player now knows that he is the expert of the red player. The red player knows that the black player's expert is the blue player.

Round 2: The black player sends a confirmation to the red player, the red player thereby earns money. At the same time the red player sends a referral to the black player, which means that the black player now knows that the blue player is the expert he is looking for. Additionally, the blue player knows his expert is the green player and the black player knows that the green player is seeking the red player. If the game ended after two rounds, only the red player would have earned money.

3.1.2   Expert game: Experimental setup

We used the facilities at Copenhagen University's Centre for Experimental Economics. They provided players randomly chosen from a database and computer facilities where the computer screens were shielded such that players could not look at each other's screens. We conducted two sessions. one session where players' identities were kept between games (ID) and one where players were given new randomly chosen names between games (NO-ID), which thereby prevented the identification of players between games. A player's identity was a randomly chosen common last name, thereby eliminating possible gender bias. Each session lasted four hours which covered introduction, playing, answering a survey and payment. In each session roughly 25 games were played, each lasting on average six rounds. The possibility of perturbing the system (having a NO-ID session) is an advantage of the game approach as opposed to the big data approach.

3.1.3   Cooperative behaviour increases when communication partners can be identified

"The Expert Game" imitates real life communication is several aspects. First, the time constraint forces prioritization of communication partners [112]. Second, information has been shown to spread at most over two links, meaning a person at most knows what his "friends of friends" are working with [115]. Referrals in "The Expert Game" allow players to have knowledge about the expertise of their "friends of friends". Third, there is no immediate benefits for helping others with their tasks [116].

When contrasting our two experiments we find that overall players are more efficient in finding their expert in the ID session (see Figure 3.2 C). The reason for the increased efficiency is the increased rate of referrals sent (see Figure 3.2 A). Since referrals have no immediate benefit for the sender, we interpret this increase in referrals as an increase in cooperative behaviour.

From simulations we know that the increased efficiency due to cooperation increases with system size, and 16 players is actually close to the minimal system size where the effect can be observed (see appendix D.1).

Previous work on cooperative behaviour in networks includes experimental and theoretical research on cooperative dilemmas such as prisoner's dilemma played on networks [117, 118, 119, 120, 121]. An interesting result from these studies is that cooperation is only promoted in networks that are dynamic, meaning that links can be rewired [117, 119]. We find similar results where the NO-ID session corresponds to a well-mixed population and the ID session allows for a dynamical network to form, thereby promoting cooperation.

There are two main differences between previous research on cooperation in networks and "The Expert Game". First, the games played in previous studies have digital choices, a player chooses to either cooperate or defect. "The Expert Game", on the other hand allows for nuanced choices that have a closer resemblance to real life. Second, the network formation in previous studies is restricted by the rules. In those studies, players are initialized on a network and after each game asked whether they want to rewire links [117, 119]. Conversely, in "The Expert Game" each player can interact with all other players; a network is only observed afterwards as a consequence of the players'

interactions, which means that communication partners are never forced on players.
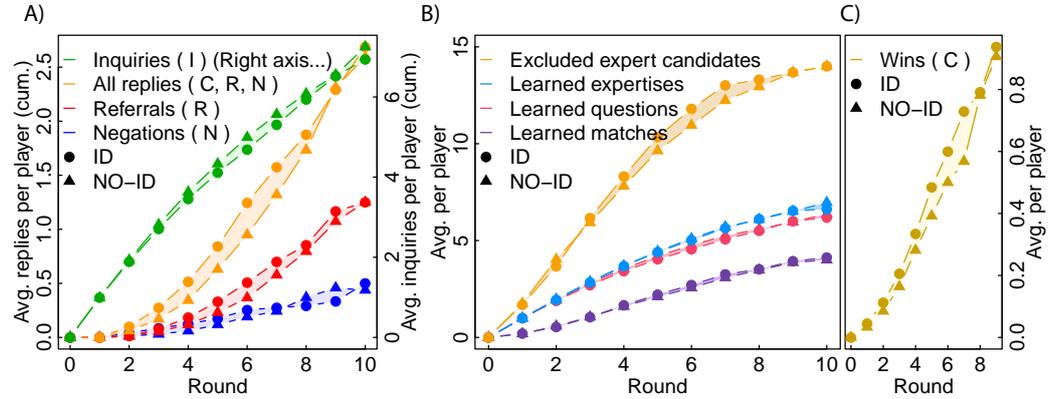
Figure 3.2: **More cooperation when identification is allowed:**
A) Cumulative message rate as a function of rounds. When identities are kept between games a higher reply rate is observed. **C** An increase in efficiency (confirmations received per round) is seen when identities are kept between games. This is due to the increase of knowledge **B**. The increase in players that can be excluded as not being the expert increases efficiency. This increase is due to the higher reply rate

3.1.4 Emergence of a communication network

In the ID session, the identities are kept between games. This allows for preferential treatment of players due to previous interactions. Analysing the data we find that indeed some communication links are heavily favoured over others (see Figure 3.3 A,D). This is quantified by conditional probabilities; what is the probability that player A sends a non-redundant message to player B given that player A can send a non-redundant message. Because of the time constraint (e.i., limited number of messages sent), preferring some links, causes other links to be neglected more than at random. This polarization of communication partners creates a network (see Figure3.3 D). When analysing the information flow across these links, we find that favoured links do not just have more messages, but even the information per message increases (see Figure 3.3 B). That more information is transferred across strong links is in agreement with experimental work by N.E. Friedkin [122]. This increase of information is due to a higher fraction of messages being referrals (see Figure 3.3 B). When contrasting with a random network, we find that favoured links are reciprocal (see Figure 3.3 C).

Since there is no immediate benefit for a player to send referrals we interpret these reciprocal high-information links as a display of trust between two players. A referral is sent because the player trusts that the recipient will reciprocate later by sending a referral when that is possible.

That cooperation can emerge because of repeated interactions, and reciprocal trust is in agreement with previous research of repeated prisoner's dilemma [123].

The polarization of preferences can be quantified as an entropy, where the entropy for each player's preference distribution is calculated and averaged. A uniform distribution of preference between players would mean no polarization and a high entropy $-\log_2(\frac{1}{15}) \approx 3.9$ bit. Note that plotting a histogram of preferences where all players have a uniform distribution would result in a peaked histogram, with peak at $\frac{1}{15}$. On

the other hand, highly polarized players would have distributions that are far from a uniform distribution, which results in a lower entropy. Using the average entropy of player preference as a measure of polarization and splitting the data into three sections, we show that the polarization increase over time (see Figure 3.3 E). This means that players find favoured communication partners. This is in agreement with survey data, where players answered that they "had friends" after 5-10 games. A more preferred
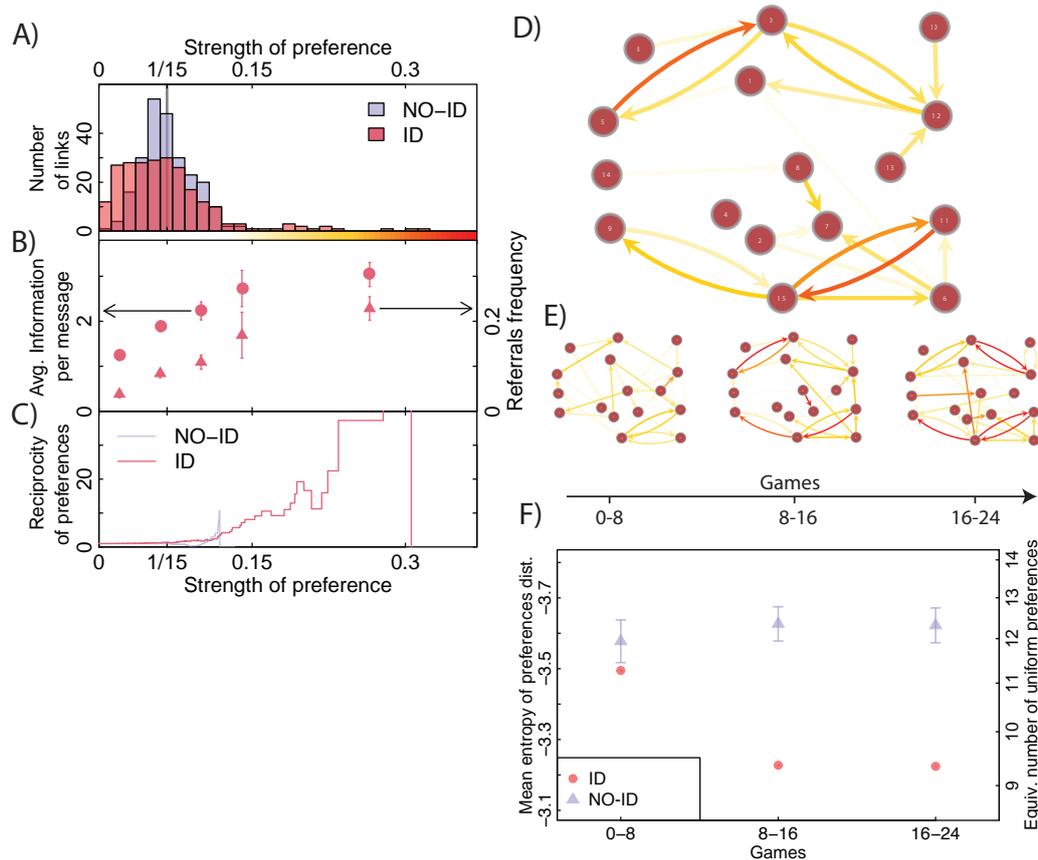
Figure 3.3: **Emergence of communication network. (A)** Distributions of preference for all links corresponding to experiments ID and NO-ID as labelled in plot. Spread in the NO-ID distribution is due to sample size, for infinite data the NO-ID distribution would be a sharp peak at $\frac{1}{15}$ (indicated by vertical gray line). **(B)** Average information per message (circles) and referral frequency (triangles) as a function of preference. Information is measured in bits: "Is player A my expert (Yes/No)" **(C)** Average link reciprocity for all links with preference larger than a threshold (horizontal axis). **(D)** Overall trust network, corresponding to the preference distribution. **(E)** Preference or trust network emergence as a function of games played. **(F)** Quantifying the network emergence using average entropy of the players' preference distributions, entropy measured in bits.

player gets more information; this player should generate more confirmations ergo more wins. However the randomness of the game and the size of our data does not allow us to make that claim (see appendix D.2).

3.1.5   Mathematical model for the Expert Game

In order to generate predictions we constructed a mathematical model. The model is an agent-based model, which for each agent computes which messages type to send and to whom. Our most likely model was selected among several models that all had the same overall equation:

$$P_{i \to j}(t) = f(\mathcal{T}_{i \to j}, p_t) \tag{3.1}$$

Where $P_{i \to j}(t)$ is the probability that agent i sends message type t to agent j. p_t is the priority of message type t and $\mathcal{T}_{i \to j}$ is agent i's trust in agent j. The best model is selected based on the likelihood that it reproduces the ID data. Among the models was one where the message priorities were based on the information content and agents were trusted based on the amount of information they provided. A second model was based on the reply probability, where each agent was trusted based on the probability that they would send a reply. The best model turned out to be a simple model where trust in a agent was a simple weighted sum of incoming messages. Weights and message priority were based on data fitting. For all of the models the best combination of message priority and player trust was the two values multiplied. The best model is therefore given as:

$$P_{i \to j}(t) \propto (\mathcal{T}_0 + \mathcal{T}_{i \to j}) \cdot p_t \ , \tag{3.2}$$

$$\mathcal{T}_{i \to j} = 4.5 \cdot C + 5.8 \cdot R + 1.5 \cdot N + I + 4.8 \ , \tag{3.3}$$

Where C is the number of confirmations i has received from j, R is number of referrals, N is negations and I is inquires. We find that these value qualitatively corresponds to survey data where players where asked to rank messages types according to how much they values them as a sign of "friendship" (see Figure 3.4 D). In addition to these parameters, all models had in common that the priority of inquiries was dropped by roughly a factor of 2 after an agent had sent an inquiry to his expert. This effect was also seen in the data (see appendix D.3).

   Since the model mimics an "average" player, it does not capture the extremes (tails) of the preference distribution (see Figure 3.4 A). However, the model clearly captures most aspects of the data. It shows the same increase of information across preferred communication links, and the same qualitatively reciprocal behaviour, found in the data (compare Figure 3.4 B, C with Figure 3.3 B, C).

3.1.6   Model predictions: Anti-exploitation and social capital

It is interesting to investigate what would happen if an agent who employs a very egoistic (defection) strategy is introduced into a game where the rest of players play according to our "average" player. We introduce an egoist agent who only sends inquires into a session where the rest of the agents play as the average players. We find that at first the egoist wins more than the "average" player, but as the games progress the average players pick up on the egoist and lowers their trust in him, thereby sending fewer messages his way. This means that the trust network that we found players building functions an anti-exploitation mechanism. Introducing more egoists into the system shows that the system is actually bistable, when there are few egoists in the system, on the long run it is not beneficial to be egoistic. However, when there are several egoistic players it becomes beneficial to become egoistic as well (see Figure 3.5 D).
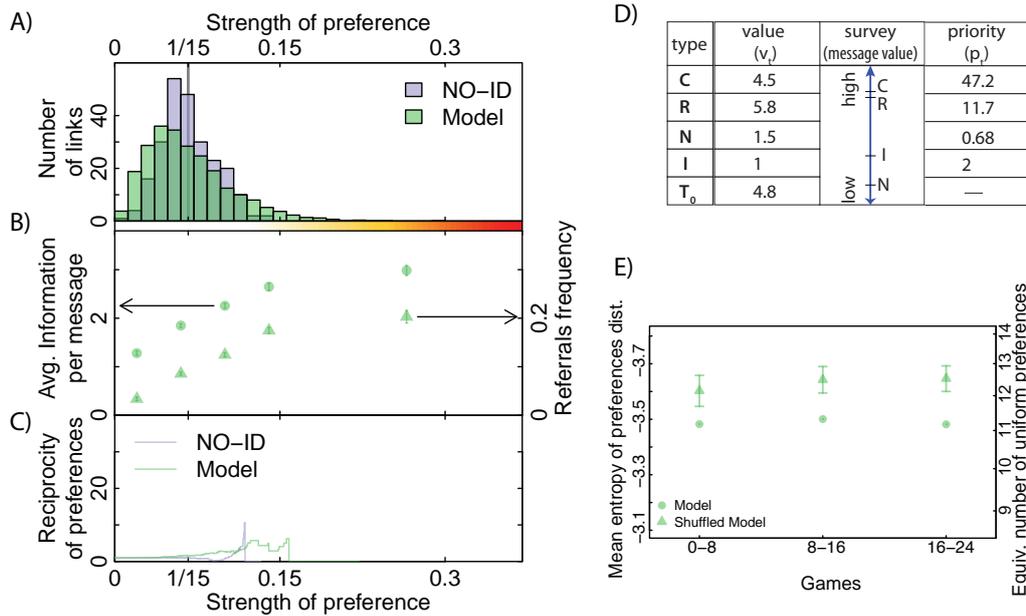
Figure 3.4: **The best model's reproduction of data A)**, Distributions of preference for all links corresponding to experiments NO-ID and model. The model is clearly broader than the NO-ID data. When comparing with fig. 3.3 A the model does not capture the extreme behaviours, very high preference and neglection of the ID session. **B)** In comparison with Figure 3.3 B the model captures the increase of information across preferred communication links. **C)**, The model also captures the reciprocity of players,it is more reciprocal than the NO-ID session, but it does not have the extreme preferences as compared with Figure 3.3 C. **D)** Qualitatively our model parameters are similar to survey data (compare first and second column). **E)** When we compare the entropy of the model, with the entropy of the model where agent identities are shuffled between games we find that the entropy indeed does decrease. However, the model is faster at establishing communication partners that in our ID session (see Figure 3.3 F). The model data is an average over 10 simulations.

An interesting observation is that monetary gain is not just due to the strategy a player plays, but also to how trusted the player is. In our models preference is simply normalized trust. Therefore more trust means higher preference, which means receiving more information, which again increases the probability of getting a confirmation.

The term social capital (SC) has many definitions [124], but the original definition that covers some of the other definitions can be rephrased as follows: SC is the economic benefits a person gains due to preferential treatment [116]. Roughly speaking this means that the number of ones friends and acquaintances and how "useful" they are have a monetary value. In "The Expert Game" a player can build up social capital by playing a non-egoistic strategy, which increases other players' trust in him. The complexity of social capital makes it notoriously difficult to measure, and often experiments opt to measure indicators of SC instead of actually SC [124]. Our mathematical model allows us to measure the social capital in "The Expert Game" as the difference in capital gain between two agents playing the same strategy (parameter set) but differing in how trusted they are. In practice this is done by simulating 15 agents following the "average"
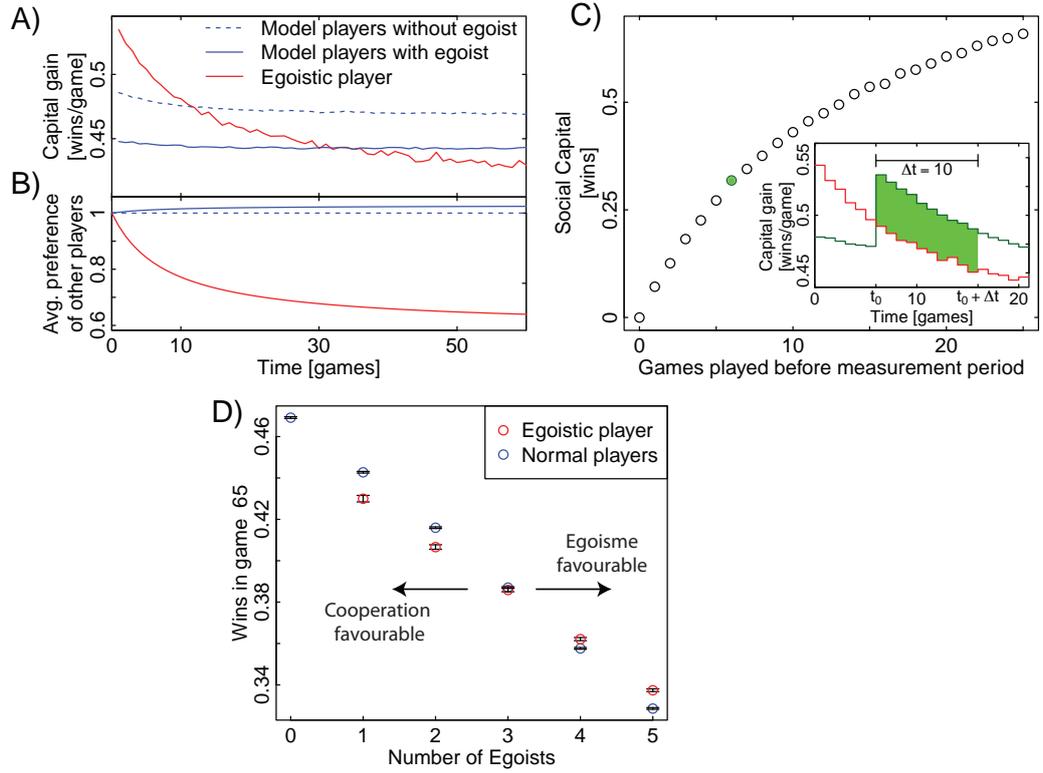
Figure 3.5: Model predictions. **(A)** Comparing the wins of an egoist (red line) with average players (blue line) we find that after roughly 30 games it becomes disadvantageous to have an egoistic strategy. Comparing games where no egoist is present (blue dashed line) with the wins of average player in the present of an egoist (blue line) shows that the egoist has an effect of the wins of the other players. **(B)** Others players preference of the egoist declines as time progresses. Meaning that players catch on the to egoist being egoistic, and thereby sending less messages his way. **(D)** In the long run it is not beneficial to play the complete egoistic strategy, unless there is atleast 3 egoists. Note that this is only beneficial in relative wins, but not in absolute wins. **(C)** Social capital difference between an egoist and an average player increases as time progresses

strategy and one additional agent who switches from an "average" strategy to the egoistic strategy after t_0 games. During the time period τ we then compare this agents capital gain with that of an agent who immediately employs the egoistic strategy ($t_0 = 0$). We can thereby measure the social capital that the switching player has built during the first t_0 games. The difference in social capital (ΔSC) between two agents is:

$$\Delta \mathcal{SC}_\tau(t_0) = \sum_{t=t_0}^{t_0+\tau} C(t, \mathcal{T}_{t_0,egoist}) - C(t, \mathcal{T}_{t_0,normal}) . \tag{3.4}$$

Here, $C(t, \mathcal{T})$ is the capital gained by an agent in game t, with $\mathcal{T}$ representing the agent's social situation in terms of trust, with the subscripts denoting the time leading up to the measurement t_0 and the strategy the agent used during that time. Figure 3.5 C shows how this social capital difference $\Delta \mathcal{SC}_{\tau=10}$ between an egoist and a normal player during the first 25 games.

3.1.7   Conclusions

Using "The Expert Game" it was possible for us to quantify the emergence of a communication network. We were able to quantify the flow of information in this network and to measure the relationship between trust and information flow. Constructing a mathematical models that mimics the average player, we could predict that players could not be exploited by a very egoistic player. Lastly using our mathematical model we were able to quantify the construction of social capital. Speculating further, we propose that our results could be used to circumvent the difficulty of measuring social capital directly in real life. We propose that this could be done in a setting where communication and information can be measured. Social capital of an individual could be quantified using the response rate to questions and associated information gain. Further work could be done to address the adverse effects of social capital and trust network. This could be done by modelling the disadvantage a new player would have when he is introduced to an already existing network. Experimentally I see three further studies being of interest. One is simply having triplets of our current 16-player sessions to solidify or dismiss our current conclusions. A second (and more interesting) direction would be to scale the experiment to 64 players, since we predict that the effects of collaborative behaviour would be more significant with more players. Lastly, it would be interesting to experimentally introduce an egoist into a system and to see how the other players would behave towards him.

Appendix DNA repair proteins

A.1 Derivation of Full RD model:

I want to stress that these derivations follow Sprague et al. [65] and I have simply used these equations for my work. The reason these derivations are featured in my thesis is simply to make it easier for the reader to find them.

First off is changing variables from F and B to $u = F_{eq} - F$ and $v = B_{eq} - B$. Where $\frac{dF}{dt} = \frac{dF}{du}\frac{du}{dt} = -\frac{du}{dt}$. Using this, the fact that F_{eq} is constant in space $\nabla^2 F_{eq} = 0$ and that $k_{on}^* F - k_{off} B_{eq} = 0$. We get that

$$\frac{\partial u}{\partial t} = D\nabla^2 u - k_{on}^* u + k_{off} v \tag{A.1}$$

$$\frac{dv}{dt} = k_{on}^* u - k_{off} v \tag{A.2}$$

Using the Laplace transform $\bar{u} = \bar{u}(s,r) = \int_0^\infty e^{-st} u(t,r) dt$ and integration by part $\int_a^b u(x)v'(x)dx = [u(x)v(x)]_a^b - \int_a^b u'(x)v(x)dx$: Solving the second equation for v:

$$\bar{v} = \frac{k_{on}^* \bar{u} + v(r,0)}{s + k_{off}} \tag{A.3}$$

Insert into the first equation.

$$s\bar{u} = D_f \nabla^2 \bar{u} - k_{on}^* \bar{u} + k_{off} \frac{k_{on}^* \bar{u} + v(r,0)}{s + k_{off}} + u(r,0) \tag{A.4}$$

Isolating $\bar{u}$ on the left hand side and using that $u(r,0) = F_{eq}$ and $v(r,0) = B_{eq}$ for $r \le r_c$ where r_c is the radius of the bleaching spot.

$$s\left(1 + \frac{k_{on}^*}{s + k_{off}}\right)\bar{u} - D_f \nabla^2 \bar{u} = k_{off} \frac{v(r,0)}{s + k_{off}} + u(r,0) \tag{A.5}$$

$$= k_{off} \frac{B_{eq}}{s + k_{off}} + F_{eq} = \frac{k_{on}^* F_{eq}}{s + k_{off}} + F_{eq} \tag{A.6}$$

Cleaning this up we get:

$$-q^2 \bar{u} + \nabla^2 \bar{u} = -\frac{F_{eq}}{D_f}\left(1 + \frac{k_{on}^*}{s + k_{off}}\right) = -V \tag{A.7}$$

$$\tag{A.8}$$

Where $q^2 = \frac{s}{D_f}\left(1 + \frac{k_{on}^*}{s + k_{off}}\right)$ and $V = \frac{F_{eq}}{D_f}\left(1 + \frac{k_{on}^*}{s + k_{off}}\right)$ for $r \le r_c$ and zero elsewhere. The solutions are:

$$\bar{u} = \begin{cases} \frac{V}{q^2} - \alpha_1 I_0(qr) & r \le r_c \\ \alpha_2 K_0(qr) & r > r_c \end{cases} \tag{A.9}$$

We require the solution is continous and differentiable across $r = r_c$. Using the identities for modified bessel functions. $I_0' = I_1$ and $K_0' = -K_1$. We can write the two requirements as:

$$\alpha_2 K_0(qr_c) = \frac{V}{q^2} - \alpha_1 I_0(qr_c) \tag{A.10}$$

$$-\alpha_2 K_1(qr_c) = -\alpha_1 I_1(qr_c) \tag{A.11}$$

Since we are only interested in the intensity inside the bleaching spot, we only solve for α_1. Using the identity $I_1(x)K_0(x) + K_1(x)I_0(x) = \frac{1}{x}$.

$$\alpha_1 = \frac{V}{q^2} qr_c K_1(qr_c) \tag{A.12}$$

Since the intensity is both from bound and free proteins, $frap(t,r) = F + B = 1 - u - v$. The Laplace transform is then:

$$\mathcal{L}(frap(t,r)) = \frac{1}{s} - \bar{u} - \bar{v} \tag{A.13}$$

Since we are measuring the average across the bleaching spot we calculate this. Remember that only u varies in space, so the average of $\bar{u}$ is given as:

$$Avg(\bar{u}) = \frac{1}{\pi r_c^2} \int_0^{2\pi} d\theta \int_0^{r_c} \left(\frac{V}{q^2} - \alpha_1 I_0(qr) \right) r \, dr \tag{A.14}$$

$$= \frac{V}{q^2} - \frac{2\alpha_1}{qr_c} I_0(qr_c) \tag{A.15}$$

Using the average of $\bar{u}$, inserting $\bar{v}$ and α_1. Having that $\frac{V}{q^2} = \frac{F_{eq}}{s}$ the Laplace transform of the intensity inside the bleaching spot becomes.

$$\mathcal{L}(frap(t,r)) = \frac{1}{s} - 2I_1(qr_c)K_1(qr_c) \left(1 + \frac{k_{on}^*}{s + k_{off}} \right) \frac{F_{eq}}{s} - \frac{C_{eq}}{s + k_{off}} \tag{A.16}$$

There is no known closed form for the inverse Laplace transform for this function. So following Sprague et al. [65] we use the *invlap.m* matlab function to numerically calculate the inverse and get $frap(t)$.

A.2 Effective diffusion model:

Using $\tau_{eff} = \frac{r_c^2}{D_{eff}} = \frac{r_c^2}{D_f} \left(1 + \frac{k_{on}^*}{k_{off}} \right)$ as the correct time scale and following [65], equation A.16 can be reduced to:

$$\mathcal{L}(frap(t,r)) = \frac{1}{s'} - \frac{1}{s'} \left(1 - 2I_1(s')K_1(s') \right) \tag{A.17}$$

Which has the solution:

$$Frap(t) = \left[I_0 \left(\frac{r_c^2}{2tD_{eff}} \right) + I_1 \left(\frac{r_c^2}{2tD_{eff}} \right) \right] \exp \left(-\frac{r_c^2}{2tD_{eff}} \right) \tag{A.18}$$

Where D_{eff} is the effective diffusion coefficient and r_c is the radius of the bleaching spot.

A.3   Reaction model:

When free protein diffusion is fast compared to binding/unbinding events. This means that the free proteins are in equilibrium F_{eq}. The equation governing the dynamics is then:

$$\frac{dB}{dt} = k^*_{on}F_{eq} - k_{off}B \tag{A.19}$$

This is a first-order linear equation with the solution:

$$B(t) = \frac{k^*_{on}F_{eq}}{k_{off}} + K\exp(-k_{off}t) \tag{A.20}$$

Using that $k^*_{on}F_{eq} = k_{off}B_{eq}$ and that there is ideally no intensity when bleached, $B(t = 0) = 0$ we get.

$$B(t) = C_{eq} - C_{eq}\exp(-k_{off}t) = C_{eq}\left(1 - \exp(-k_{off}t)\right) \tag{A.21}$$

The total intensity $Frap(t)$ is then:

$$Frap(t) = F_{eq} + C_{eq}\left(1 - \exp(-k_{off}t)\right) = 1 - C_{eq}\exp(-k_{off}t) \tag{A.22}$$

Where in the last step we use that $F_{eq} + C_{eq} = 1$ since the intensity is normalized.

A.4   BLM datasets
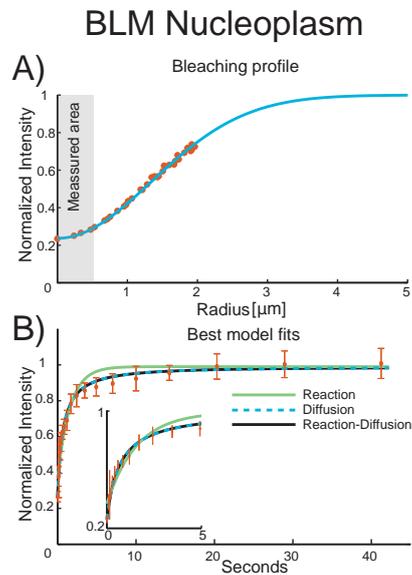
Figure A.1: Analyses of the BLM diffusion dynamics in the nucleoplasm. A) Initial Bleaching profile $I_0(r)$ fitting. B) Fitting of FRAP, the best model is the diffusion model with a effective diffusion coefficient of 1.34 $\frac{\mu m^2}{s}$.
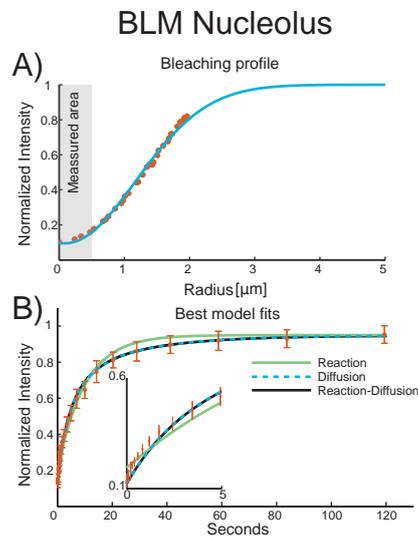
Figure A.2: Analyses of the BLM diffusion dynamics in nucleoli, error bars represent standard deviation. A)The initial bleaching profile $I_0(r)$ fitted to a modified Gaussian profile. B) Fitting recovery of fluorescence after photobleaching using the 3 models: Reaction, Diffusion and Reaction-Diffusion. From the AIC model selection we get that the diffusion model is the superior model. The effective diffusion coefficient for BLM is 0.13 $\frac{\mu m^2}{s}$.
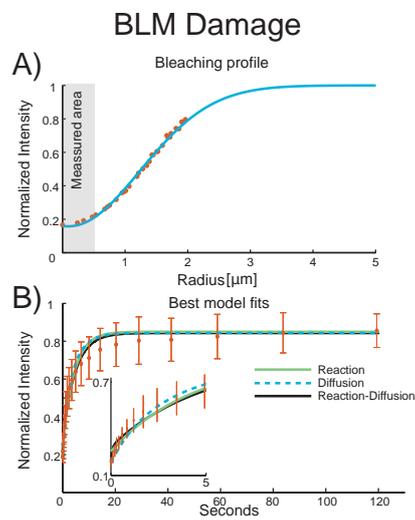
Figure A.3: Analyses of the BLM diffusion dynamics at the site of damage. Data is an average over 5 cells, error bars represent standard deviation A) The initial bleaching profile is fitted to a Gaussian profile. B) None of the models provide a good fit. However we find that both BLM have a fraction of 15% bound very strongly to the damage.

A.5 No recruitment for low laser power
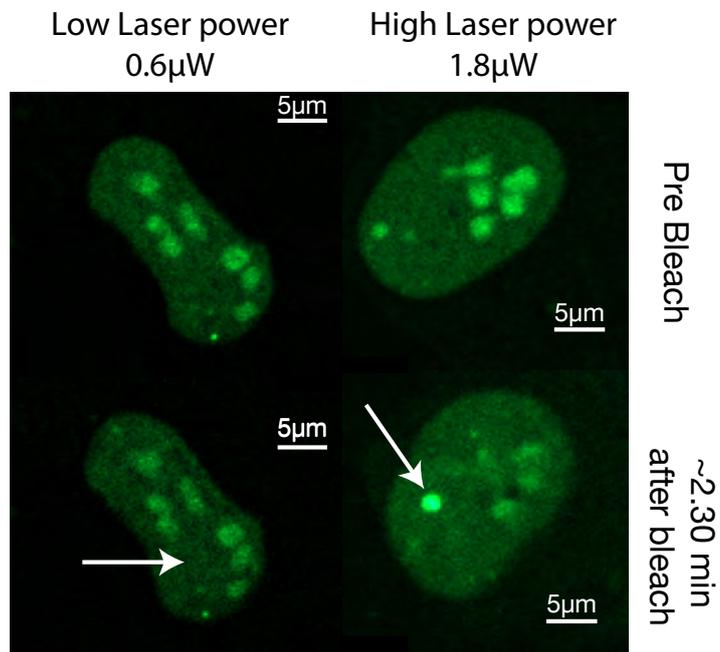
Figure A.4: . No double stranded breaks were created with the low laser power(0.6μW). As a control to see if the FRAP experiment induced double stranded breaks we bleached cells and monitored them for roughly 2 and a half minute, but no recruitment was seen

	Measured Diff Coeff. D_{eff}	Theoretical Diff. Coeff. D_{theo}	References
RAD 54	14	14.6	Hamster ovary cell [125]
RAD 52	8	14.6	Hamster ovary cell [125]
RAD 51	7	17.5	Hamster ovary cell [125]
NBS1	3	14.6	Human U2OS [126]
MDC1	2	11.7	Human U2OS [126]
Ku70	0.35	15.2	HeLa cells [127]
Ku86	0.35	14.7	HeLa cells [127]

Table A.1: Theoretical and measured diffusion coefficients

A.6   Table for diffusion coefficients

Appendix for regulatory effects of RNA polymerase

First of all the total promoter complex concentration is the sum for the free promoters, the promoters where the first promoter is bound by an RNA polymerase and the promoters where the second promoter is bound by RNA polymerase:

$$[P]_{tot} = [P] + [P_1R] + [P_2R]$$

$$1 = \frac{[P]}{[P]_{tot}} + \frac{[P_1R]}{[P]_{tot}} + \frac{[P_2R]}{[P]_{tot}}$$

Note that the fractions is equivalent to the probability of being in that state e.g. $\Theta_2 = \frac{[P_2R]}{[P]_{tot}}$.

Second the biochemical differential equation governing the dynamics are given as:

$$[\dot{P_1R}] = \tilde{k}_{on}^1 [P][RNAP] - k_f^1 [P_1R]$$

$$[\dot{P_2R}] = \tilde{k}_{on}^2 [P][RNAP] - k_f^2 [P_2R]$$

We now define the pseudo on-rate as $k_{on}^1 = \tilde{k}_{on}^1 [RNAP]$, which is a dimensionless quantity. We here assume that for all intent and purposes the RNA polymerase concentration is constant. Since the RNA polymerases greatly outnumber number of promoter sites this is a fair assumption. Now if the concentrations are measured as concentrations relative to the total concentration, the equations become dimensionless and we recover the master equation.

$$\frac{[\dot{P_1R}]}{[P]_{tot}} = k_{on}^1 \left(1 - \frac{[P_1R]}{[P]_{tot}} - \frac{[P_2R]}{[P]_{tot}}\right) - k_f^1 \frac{[P_1R]}{[P]_{tot}} \tag{B.1}$$

$$\frac{[\dot{P_2R}]}{[P]_{tot}} = k_{on}^2 \left(1 - \frac{[P_1R]}{[P]_{tot}} - \frac{[P_2R]}{[P]_{tot}}\right) - k_f^2 \frac{[P_2R]}{[P]_{tot}} \tag{B.2}$$

$$\dot{\Theta}_1 = k_{on}^1(1 - \Theta_1 - \Theta_2) - k_f^1 \Theta_1$$

$$\dot{\Theta}_2 = k_{on}^2(1 - \Theta_1 - \Theta_2) - k_f^2 \Theta_2$$

Appendix for NAF motif

C.1 Tuning the transcription rates for oscillation

The production rates for repressor and activator mRNAs (transcription rates) can be tuned by mutations in the promoter sequences. We explore the effects of decreasing the maximal transcription rate, keeping the ratios between the maximal production rates of activator and repressor mRNAs. The top right corner in Figure C.2 (X,Y,Z) corresponds to the parameters used for points (X,Y,Z) in Figure 2.9. Starting from either damped oscillations (Y) or sustained oscillations (Z) we show that only a certain range of transcription rates can produce damped/sustained oscillations. Note that starting from (X) (the completely symmetric system), inducing an asymmetry in transcriptional rates does not produce oscillations. Which was the case for an asymmetry in half-lives (Figure 2.9).

	Minimum	Median	Maximum
mRNA Degradation $\left[\frac{mRNA}{min}\right]$ [102]	$3.7 \cdot 10^{-4}$	$1.3 \cdot 10^{-3}$	$2.3 \cdot 10^{-2}$
Protein Degradation $\left[\frac{Protein}{min}\right]$ [102]	$5.7 \cdot 10^{-5}$	$2.4 \cdot 10^{-4}$	$2.3 \cdot 10^{-2}$
mRNA half life $\left[\frac{1}{hour}\right]$ [102]	0.5	9	31
Protein half life $\left[\frac{1}{hour}\right]$ [102]	0.5	46	200
Translation $\left[\frac{Protein}{mRNA\ min}\right]$ [102]	$1.6 \cdot 10^{-2}$	2.3	16
Transcription $\left[\frac{mRNA}{min}\right]$ [102]	$1.6 \cdot 10^{-3}$	0.03	1.6
Transcription rate when activated $\left[\frac{mRNA}{min}\right]$ [102]	-	1.6	-
Repressed Transcription rate $\left[\frac{mRNA}{min}\right]$	-	$1.6 \cdot 10^{-4}$	-
Unregulated Transcription rate $\left[\frac{mRNA}{min}\right]$ [128]	-	$1.6 \cdot 10^{-2}$	-
K_D^{Dimer} [nM] [101]	-	10	-
K_D^{DNA} [nM] [101]	-	0.18	-

Table C.1: Parameters for the model: Parameters are taken from Schwanhausser et al. 2011 [102] and transcription rates are chosen so they mapped to the mRNA and protein abundances measured in [102] .
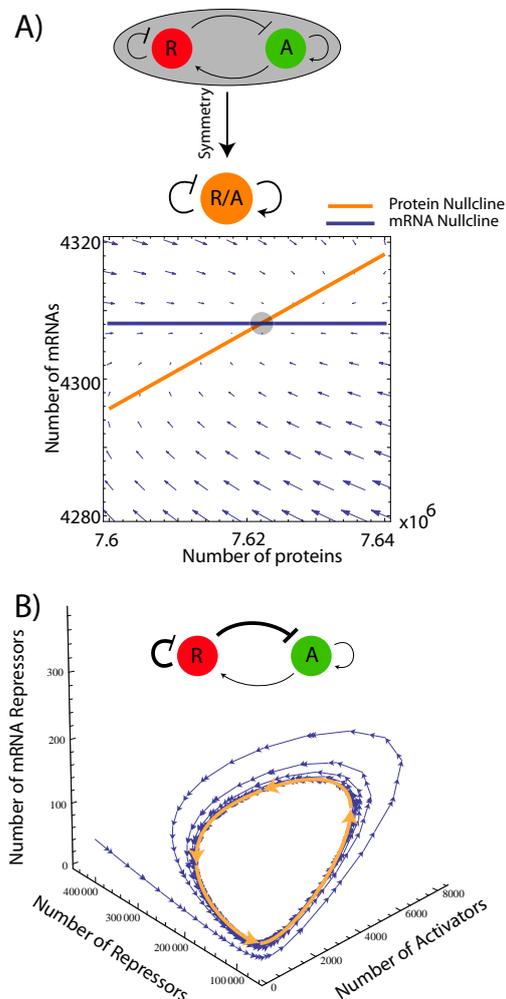
Figure C.1: **Dynamics of the coupled NAF system both symmetrical and asymmetric degradations:**
A) The Tet two-dimensional symmetric NAF motif, where the repressor and activator are collapsed into one protein, cannot oscillate. The system will always settle at a stable node. This two-dimensional system will have the qualitative same behaviour as a four-dimensional system where the activator/repressor parameters are identical. **B)** When the mRNA and protein half-life of the activator is reduced by a factor 10 (to 0.9 hours) the system undergoes a Hopf bifurcation and a stable limit cycle occurs. We show the transient behaviour which settles at the limit cycle which is highlighted in orange.
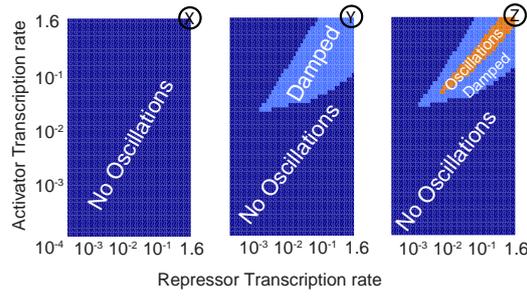
Figure C.2: **Asymmetry in transcription**
The top right corner in Figure C.2 (X,Y,Z) corresponds to the parameters used for points (X,Y,Z) in Figure 2.9. **X** Asymmetry in transcription can not produce oscillations. **Y** Changing the transcription asymmetry can not change damped oscillations to sustained oscillations. **Z** Tuning of the transcription rates is needed to produce oscillations
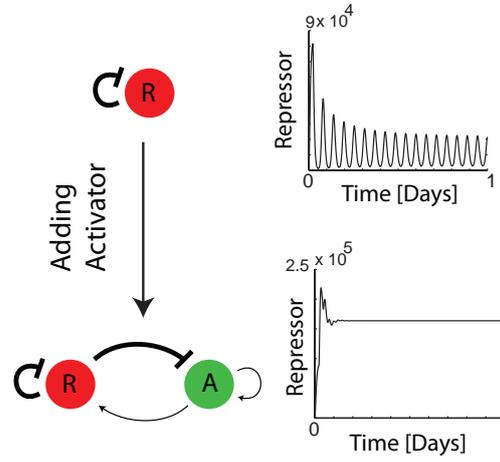
Figure C.3: **Adding Activator**
Adding an activator to the Hes Oscillator quench the oscillations. All half-lives are set to 4 minutes

C.2   Effect of hetereodimers

C.2.1   Transcription rate $\langle r_1(\tau) \rangle_{(A,R)}$:

$$\langle r_1(\tau) \rangle_{(A,R)} = \frac{N}{ZK_{D*}^{DNA}} (\beta_1^* AA_{tot}^* + \beta_2^* AR_{tot}^* + \beta_3^* RR_{tot}^*) + N\beta_4^*$$

here AA_{tot}^* is the total number of dimer-activators, RR_{tot}^* is the total of dimer-repressors, AR_{tot}^* is the total of hetero-dimers, K_{D*}^{DNA} is the dimensionless dissociation for the protein-DNA binding. β_1 is activated transcription rate, β_2 is heterodimer transcription rate, β_3 is repression transcription rate and β_4 is unregulated(basal) transcription. N is the number of regulator sites. Z is the partition function $Z = 1 + \frac{AA_{tot}^*}{K_{D*}^{DNA}} + \frac{RR_{tot}^*}{K_{D*}^{DNA}} + \frac{AR_{tot}^*}{K_{D*}^{DNA}}$.
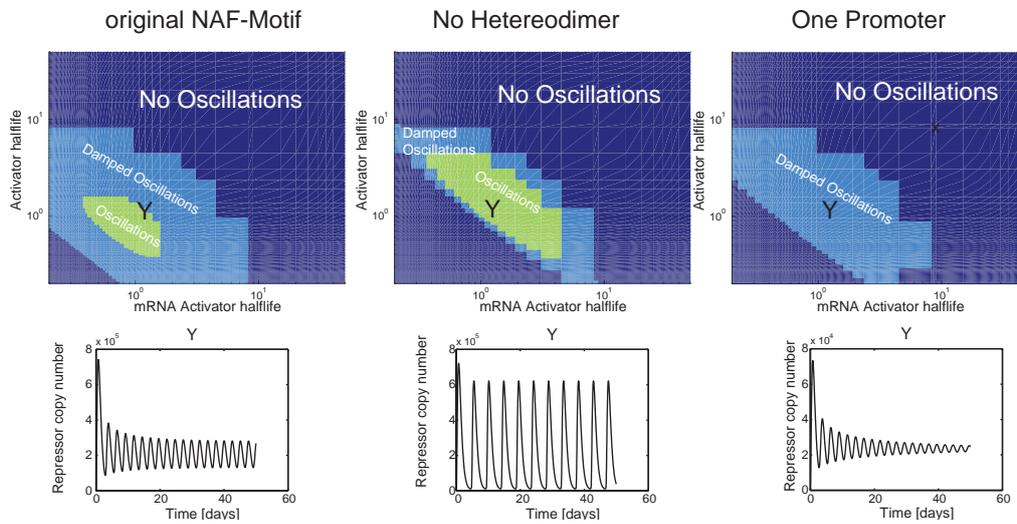
Figure C.4: **Hetereodimer formation and number of promoters:**
When the equations are changed such that hetereodimer formation is not possible the area of oscillations increases. Oppositely if the number of promoter sites is reduced to 1 then this quenches the area of oscillations.
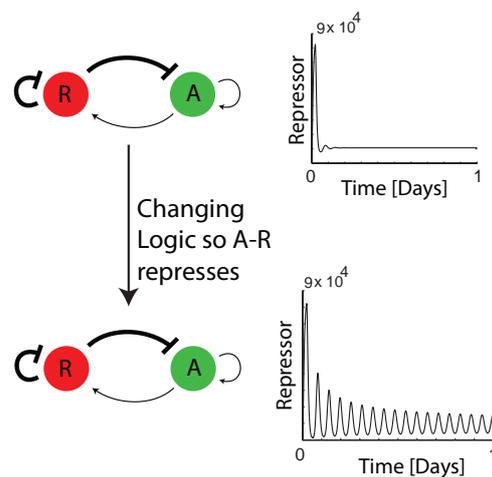
Figure C.5: **Changing Motif Logics**
By changing the logics of the transcription regulation, such that the repressor becomes dominant. That is the hetereodimer acts as a repressor can facilitate oscillations. All half-lives are set to 4 minutes

The transcription rate where hetereodimer are not allowed is very similar.

$$\langle r_1(\tau) \rangle_{(A,R)} = \frac{N}{ZK_{D*}^{DNA}} \left(\beta_1^* AA_{tot}^* + \beta_3^* RR_{tot}^* \right) + N\beta_4^*$$

However the AA_{tot}^* and RR_{tot}^* can be found analytically:

$$AA_{tot}^* = \frac{-K_{D*}^{Dimer} + \sqrt{K_{D*}^{Dimer\,2} + 16A_{tot}^2}}{8}$$

Where A_{tot} is the total amount of activators, the equation for repressors is equivalent. Note that when hetereodimers are not allowed the partition function Z, becomes $Z = 1 + \frac{AA_{tot}^*}{K_{D*}^{DNA}} + \frac{RR_{tot}^*}{K_{D*}^{DNA}}$.

C.2.2   Mono stable non-oscillatory regime

We find two fixed points of the system, analyzing the stability of these we find that the system is monostable in the non-oscillatory regime, see figure C.6.
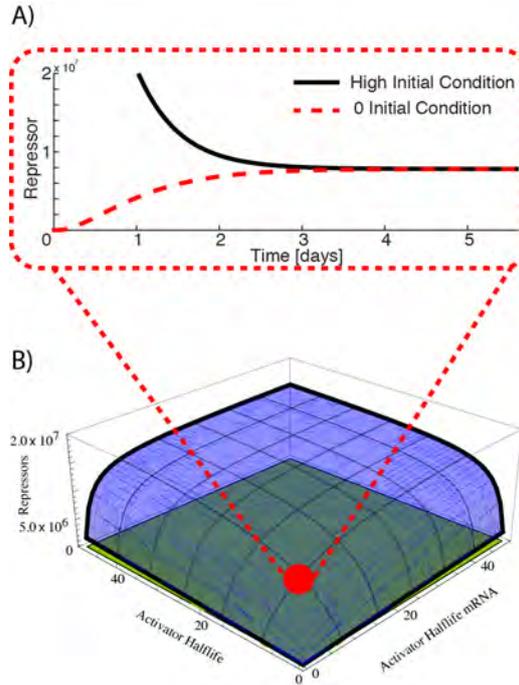
Figure C.6: **The no-oscalilation regime is monostable:**
We plot the fixed points for the repressor within the same regime as scanned in figure 2.9. **B)** The system has two fix points, the yellow and the blue. **A)** Testing for stability, we see that only the blue solution is stable.

C.2.3   Deriving the transcription rate $\langle r_1(\tau)\rangle_{(A,R)}$:

Dimerization

We have the following equation for dimerization of TetR and TetA:

$$\frac{d\,[RR]}{dt} = k_{on}^1\,[R]\,[R] - k_{off}^1\,[RR]$$

$$\frac{d\,[AR]}{dt} = k_{on}^2\,[A]\,[R] - k_{off}^2\,[AR]$$

$$\frac{d\,[AA]}{dt} = k_{on}^3\,[A]\,[A] - k_{off}^3\,[AA]$$

Due to biochemical identities we have that: $k_{on} = k_{on}^1 = k_{on}^2 = k_{on}^3$ and $k_{off} = k_{off}^1 = k_{off}^2 = k_{off}^3$. Solving for steady state we get that:

$$[R]^2 = K_D^{Dimer}\,[RR] \tag{C.1}$$

$$[A]^2 = K_D^{Dimer}\,[AA] \tag{C.2}$$

$$[A]\,[R] = K_D^{Dimer}\,[AR] \tag{C.3}$$

In addition we have that the total activator and repressor concentration is conserved on short timescales:

$$[A]_{tot} = [A] + 2\,[AA]_{tot} + [AR]_{tot} = [A] + 2\,[AA] + 2\,[AAO] + [AR] + [ARO]$$

$$[R]_{tot} = [R] + 2\,[RR]_{tot} + [AR]_{tot} = [R] + 2\,[RR] + 2\,[RRO] + [AR] + [ARO]$$

Where $[AAO]$ is the amount of activators (or repressors) bound to the operator sites. Since the number of operator sites is small (roughly 7), we assume that the majority of dimers are free. $[TF]_{tot} \approx [TF]$.

We then get:

$$[A]_{tot} = [A] + 2\,[AA]_{tot} + [AR]_{tot}$$

$$[R]_{tot} = [R] + 2\,[RR]_{tot} + [AR]_{tot}$$

and

$$[R]^2 = K_D^{Dimer}\,[RR]_{tot} \tag{C.4}$$

$$[A]^2 = K_D^{Dimer}\,[AA]_{tot} \tag{C.5}$$

$$[A]\,[R] = K_D^{Dimer}\,[AR]_{tot} \tag{C.6}$$

We are interested is an algebraic relation which gives us the free dimer concentrations when we consider the total concentrations:

$$[A] + 2\frac{[A]^2}{K_D^{Dimer}} + \frac{[A]\,[R]}{K_D^{Dimer}} - [A]_{tot} = 0$$

$$[R] + 2\frac{[R]^2}{K_D^{Dimer}} + \frac{[A]\,[R]}{K_D^{Dimer}} - [R]_{tot} = 0$$

Solving the first equation with respect to $[A]$, we obtain:

$$[A] = \frac{1}{4}\left(\sqrt{8\cdot[A]_{tot}\cdot K_D^{Dimer} + K_D^{Dimer\,2} + 2K_D^{Dimer}\,[R] + [R]^2} - K_D^{Dimer} - [R]\right)$$

Inserting into the second equation:

$$[R] + 2\frac{[R]^2}{K_D^{Dimer}} - [R]_{tot} +$$

$$\frac{1}{4}\left(\sqrt{8 \cdot [A]_{tot} \cdot K_D^{Dimer} + K_D^{Dimer^2} + 2K_D^{Dimer}[R] + [R]^2} - K_D^{Dimer} - [R]\right)\frac{[R]}{K_D^{Dimer}} = 0$$

The solution to this equation is so cumbersome that it is easier to solve numerical for each time step using the Newton method with tolerance of 10^{-3}.

For the later equations we denote the solutions for R and A as A^* and R^*, and the dimers as:

$$AA_{tot}^* = \frac{A^{*2}}{K_D^{Dimer}}$$

$$RR_{tot}^* = \frac{R^{*2}}{K_D^{Dimer}}$$

$$AR_{tot}^* = \frac{A^* R^*}{K_D^{Dimer}}$$

The concentration equations can be changed to equations for protein numbers by changing the dissociation so they are dimensionless (by multiplying with the nucleus volume and Avogado's number). $K_D^{Dimer} \cdot V_{nucleus} \cdot C_A = K_{D*}^{Dimer}$. For mammalian cells the nuclear volume is $1650 \cdot 10^{-18}$ m^3

DNA binding

In addition to the dimerization the dimers bind to the operator sites on the DNA.

The total concentration for a single operator site is the sum of free and bound operator site.

$$[O]_{tot} = [O]_{free} + [RRO] + [ARO] + [AAO]$$

The differential equation governing a single operator site binding is:

$$\frac{d\,[AAO]}{dt} = k_{on}^{DNA}\,[AA]\,[O] - k_{off}^{DNA}\,[AAO]$$

$$\frac{d\,[ARO]}{dt} = k_{on}^{DNA}\,[AR]\,[O] - k_{off}^{DNA}\,[ARO]$$

$$\frac{d\,[RRO]}{dt} = k_{on}^{DNA}\,[RR]\,[O] - k_{off}^{DNA}\,[RRO]$$

Note, that the biochemical "symmetries" mean that every dimer interacts in the same way with the DNA and the on rates and the off rates are therefore equal for dimers. The total transcription factor concentration is both the concentration of transcription factor bound to the operator site and free transcription factor.

$$[RR]_{tot} = [RR] + [RRO]$$

$$[AR]_{tot} = [AR] + [ARO]$$

$$[AA]_{tot} = [AA] + [AAO]$$

However we previously assumed that the bound fraction is small compared to the free fraction. Which means that the free concentration is almost the total concentration, $[TF]_{tot} \approx [TF]$. Using this assumption, solving the DNA binding for steady state we find that:

$$[AAO] = \frac{[AA]_{tot}}{K_D^{DNA}} [O] = \frac{AA_{tot}}{K_{D*}^{DNA}} [O]$$

$$[ARO] = \frac{[AR]_{tot}}{K_D^{DNA}} [O] = \frac{AR_{tot}}{K_{D*}^{DNA}} [O]$$

$$[RRO] = \frac{[RR]_{tot}}{K_D^{DNA}} [O] = \frac{RR_{tot}}{K_{D*}^{DNA}} [O]$$

Where the dissociation constants are made dimensionless
Using the steady states and conservation of operator sites we can calculate the probability for a single promoter to be free, bound by activator dimers AA, activator-repressor dimers AR, or repressor dimers RR.

$$[O]_{tot} = [O] + \left(\frac{AA_{tot}}{K_{D*}^{DNA}} + \frac{RR_{tot}}{K_{D*}^{DNA}} + \frac{AR_{tot}}{K_{D*}^{DNA}} \right) [O]$$

$$P(free) = \frac{[O]}{[O]_{tot}} = \frac{1}{1 + \frac{AA_{tot}^*}{K_{D*}^{DNA}} + \frac{RR_{tot}^*}{K_{D*}^{DNA}} + \frac{AR_{tot}^*}{K_{D*}^{DNA}}} = \frac{1}{Z}$$

$$P(AA) = \frac{[AAO]}{[O]_{tot}} = \frac{AA_{tot}^*}{K_{D*}^{DNA}} \frac{1}{1 + \frac{AA_{tot}^*}{K_{D*}^{DNA}} + \frac{RR_{tot}^*}{K_{D*}^{DNA}} + \frac{AR_{tot}^*}{K_{D*}^{DNA}}} = \frac{AA_{tot}^*}{K_{D*}^{DNA}} \frac{1}{Z}$$

$$P(AR) = \frac{[ARO]}{[O]_{tot}} = \frac{AR_{tot}^*}{K_{D*}^{DNA}} \frac{1}{1 + \frac{AA_{tot}^*}{K_{D*}^{DNA}} + \frac{RR_{tot}^*}{K_{D*}^{DNA}} + \frac{AR_{tot}^*}{K_{D*}^{DNA}}} = \frac{AR_{tot}^*}{K_{D*}^{DNA}} \frac{1}{Z}$$

$$P(RR) = \frac{[RRO]}{[O]_{tot}} = \frac{RR_{tot}^*}{K_{D*}^{DNA}} \frac{1}{1 + \frac{AA_{tot}^*}{K_{D*}^{DNA}} + \frac{RR_{tot}^*}{K_{D*}^{DNA}} + \frac{AR_{tot}^*}{K_{D*}^{DNA}}} = \frac{RR_{tot}^*}{K_{D*}^{DNA}} \frac{1}{Z}$$

Transcriptional probability
From the probabilities for the states of the operator sites we can calculate the probabilities of transcription, which for a single operator site gives the rate of transcription as:

$$r_1 = \beta_1 P(AA) + \beta_2 P(AR) + \beta_3 P(RR) + \beta_4 P(free)$$

We assume there is a linear relation between occupancy and transcription probability, so for N operator sites the average rate is given as:

$$\langle r_1 \rangle = \beta_1 \langle n_{AA} \rangle + \beta_2 \langle n_{AR} \rangle + \beta_3 \langle n_{RR} \rangle + \beta_4 \langle free \rangle$$

We have that $\langle n_{AA} \rangle = N \cdot P(AA)$ where again N is the number of operator sites and P(AA) is the probability the activation dimer has bound. Since $P(free) = 1 - P(AA) - P(AR) - P(RR)$ we get that

$$\langle r_1 \rangle = (\beta_1 - \beta_4) \langle n_{AA} \rangle + (\beta_2 - \beta_4) \langle n_{AR} \rangle + (\beta_3 - \beta_4) \langle n_{RR} \rangle + \beta_4 N$$

$$\langle r_1 \rangle = N\beta_1^* P(AA) + N\beta_2^* P(AR) + N\beta_3^* P(RR) + N\beta_4 \qquad \text{(C.7)}$$

This can be written using the total number of dimer proteins AA_{tot}^*, AR_{tot}^* and RR_{tot}^*.

$$\langle r_1 \rangle = \frac{N}{ZK_{D*}^{DNA}} \left(\beta_1^* AA_{tot}^* + \beta_2^* AR_{tot}^* + \beta_3^* RR_{tot}^* \right) + N\beta_4^*$$

Where AA_{tot}^*, AR_{tot}^* and RR_{tot}^* are numerically calculated from the total concentration of activator and repressor A_{tot} and R_{tot}

$$\dot{A}_{tot} = r_2 mRNA_A - \gamma_2 A_{tot} \qquad \text{(C.8)}$$

$$\dot{R}_{tot} = r_2 mRNA_R - \gamma_2 R_{tot} \qquad \text{(C.9)}$$

$$\dot{mRNA}_A = \langle r_1 \rangle - \gamma_1 mRNA_A \qquad \text{(C.10)}$$

$$\dot{mRNA}_R = \langle r_1 \rangle - \gamma_1 mRNA_R \qquad \text{(C.11)}$$

C.2.4 Frustrated bistability motif and simplified NAF equations:

I constructed simplified NAF equations modifying the frustrated bistability equations:

$$\frac{dA}{dt} = \alpha \frac{1}{1 + \frac{R}{K}^2} \frac{b + \frac{A}{K}^2}{1 + \frac{A}{K}^2} - \gamma_1 A$$

$$\frac{dR}{dt} = \alpha \overbrace{\frac{1}{1 + \frac{R}{K}^2}}^{\text{NAR}} \frac{b + \frac{A}{K}^2}{1 + \frac{A}{K}^2} - \gamma_2 R$$

Here A is the activator, R is the repressor, b=0.0016 is leakages of the promoter, α=2.3 is translation rate, K=55 is the ratio of dimerization dissociation constant to DNA dissociation constant, and γ is the degradation rates. The NAR term (indicated by the curly bracket) is set equal to 1 for the frustrated bistability motif.

Appendix for The Expert Game
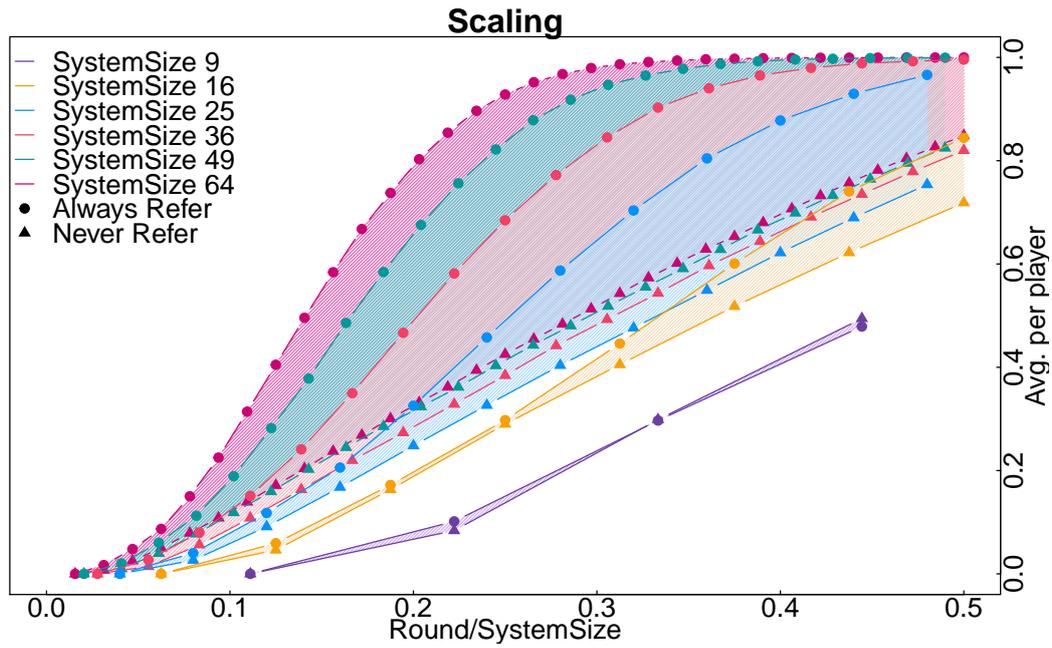
Figure D.1: Effect of cooperation is more pronounces in larger systems
We contrast two strategies. Never send referrals versus always send referrals. It is clear that the large the system size becomes the more important referrals become. Note that we have a system size of 16, where the effect exists but is small.
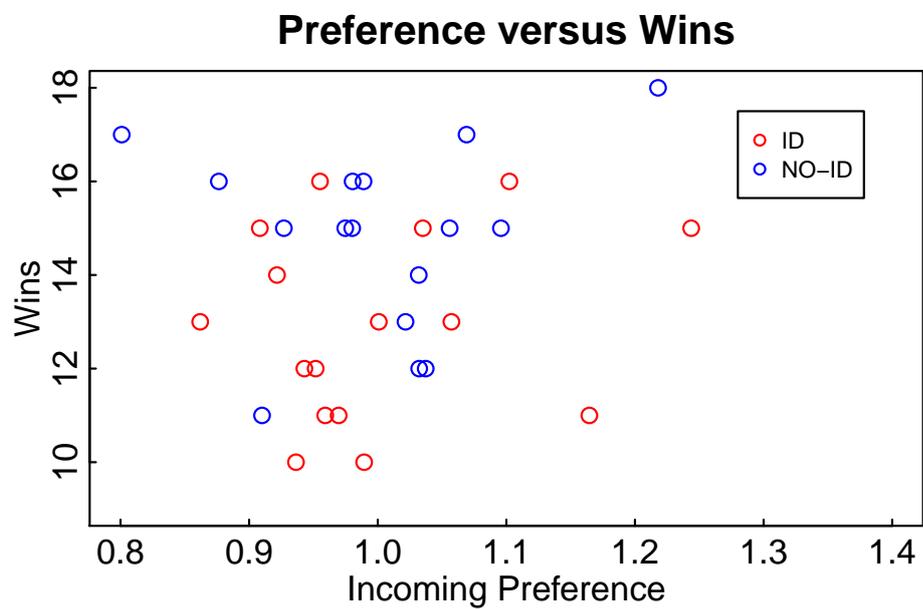
Figure D.2: **Preference versus wins**
We do not see a correlation between the preference of players and wins. Also the spread in incoming preference is comparable between the NO-ID and ID game. This is probably due to the time constraint in both games. Meaning that the players who are strongly preferred spend too much time helping.
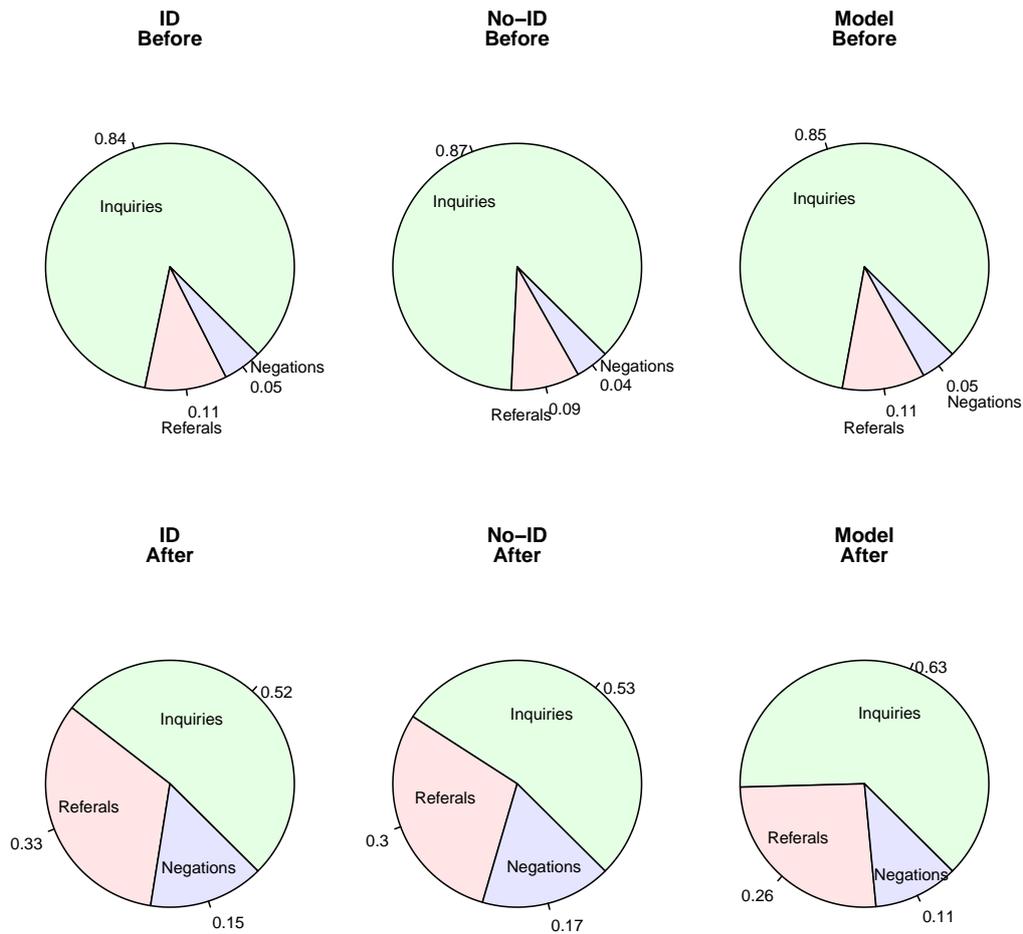
Figure D.3: **Message distribution before and after an inquiry was send to the expert**
For both sessions there is a clear tendency to send more inquiries before the expert is found. This is not a surprise since the goal of the game is to get a confirmation from their respective experts.

Bibliography

[1]   J. Campisi. "Senescent Cells, Tumor Suppression, and Organismal Aging: Good Citizens, Bad Neighbors". *Cell* (2005).

[2]   J.-H. Chen et al. "DNA damage, cellular senescence and organismal ageing: causal or correlative?" *Nucleic Acids Res.* (2007).

[3]   J. Campisi and F. d'Adda di Fagagna. "Cellular senescence: when bad things happen to good cells". *Nat Rev Mol Cell Biol* (2007).

[4]   L. Hayflick. "The limited in vitro lifetime of human diploid cell strains". *Experimental Cell Research* (1965).

[5]   U Herbig et al. "Cellular Senescence in Aging Primates". *Science* (2006).

[6]   W. Xue et al. "Senescence and tumour clearance is triggered by p53 restoration in murine liver carcinomas". *Nature* (2007).

[7]   B. Yankner and T Lu. "The aging brain". *Annu Rev pathol. Mech. Dis.* (2008).

[8]   V. Liu et al. "Mutations in mitochondrial DNA accumulate differentially in three different human tissues during ageing". *Nucleic Acids Res.* (1998).

[9]   J. Hoeijmakers. "DNA Damage, Aging, and Cancer". *New England Journal of Medicine* (2009).

[10] T. A. Rando. "Stem cells, ageing and the quest for immortality". *Nature* (2006).

[11] N. E. Sharpless and R. A. DePinho. "Telomeres, stem cells, senescence, and cancer". *J. Clin. Invest.* (2004).

[12] E Hiyama and K Hiyama. "Telomere and telomerase in stem cells". *British journal of cancer* (2007).

[13] H. B. Sieburg et al. "Predicting clonal self-renewal and extinction of hematopoietic stem cells". *Proc Natl Acad Sci USA* (2011).

[14] G Keller and R Snodgrass. "Life span of multipotential hematopoietic stem cells in vivo." *J. Exp. Med.* (1990).

[15] F. Arai and T. Suda. *Quiescent stem cells in the niche*. StemBook. Cambridge (MA): Harvard Stem Cell Institute, 2008.

[16] L. Li and H. Clevers. "Coexistence of quiescent and active adult stem cells in mammals." *Science* (2010).

[17] J. Kajstura et al. "Myocyte Turnover in the Aging Human Heart". *Circulation Research* (2010).

[18] N. A. Bishop et al. "Neural mechanisms of ageing and cognitive decline." *Nature* (2010).

[19] D. B. Lombard et al. "DNA Repair, Genome Stability, and Aging". *Cell* (2005).

[20] B. Schumacher et al. "Age to survive: DNA damage and aging." *Trends Genet.* (2008).

[21] P Hasty. "Aging and Genome Maintenance: Lessons from the Mouse?" *Science* (2003).

[22] R Busuttil et al. "Genomic Instability, Aging, and Cellular Senescence". *Ann. N. Y. Acad. Sci.* (2004).

[23] M. Dean et al. "Tumour stem cells and drug resistance." *Nat. Rev. Cancer* (2005).

[24] G Evan. "A Matter of Life and Cell Death". *Science* (1998).

[25] M. Pollack et al. "The role of apoptosis in the normal aging brain, skeletal muscle, and heart." *Ann. N. Y. Acad. Sci.* (2002).

[26] T Lindahl. "Instability and decay of the primary structure of DNA". *Nature* (1993).

[27] M. A. Blasco. "Telomeres and human disease: ageing, cancer and beyond". *Nat Rev Genet* (2005).

[28] M Lorenz et al. "BJ fibroblasts display high antioxidant capacity and slow telomere shortening independent of hTERT transfection." *Free Radical Biology and Medicine* (2001).

[29] K. E. Huffman et al. "Telomere Shortening Is Proportional to the Size of the G-rich Telomeric 3'-Overhang". *J. Biol. Chem.* (2000).

[30] R. M. Cawthon et al. "Association between telomere length in blood and mortality in people aged 60 years or older." *Lancet* (2003).

[31] M Collado et al. "Cellular Senescence in Cancer and Aging". *Cell* (2007).

[32] K. M. Bendtsen et al. "Fragile DNA repair mechanism reduces ageing in multicellular model." *PLoS ONE* (2012).

[33] M Collado. "Senescence in tumours: evidence from mice and humans". *Nat. Rev. Cancer* (2010).

[34] J. Campisi et al. "Cellular senescence, cancer and aging: the telomere connection". *Exp. Gerontol.* (2001).

[35] J Campisi. "Aging, tumor suppression and cancer: high wire-act!" *Mech. Ageing Dev.* (2005).

[36] F Rodier and J Campisi. "Four faces of cellular senescence". *The Journal of Cell Biology* (2011).

[37] M. A. Blasco. "Telomere length, stem cells and aging". *Nature chemical biology* (2007).

[38] Y.-S. Cong et al. "Human Telomerase and Its Regulation". *Microbiol. Mol. Biol. Rev.* (2002).

[39] J. Pellettieri and A. Sánchez Alvarado. "Cell turnover and adult tissue homeostasis: from humans to planarians." *Annu. Rev. Genet.* (2007).

[40] N. E. Sharpless and R. A. DePinho. "How stem cells age and why this makes us grow old". *Nat Rev Mol Cell Biol* (2007).

[41] S. Holbek et al. "Moderate stem-cell telomere shortening rate postpones cancer onset in a stochastic model". *Physical Review E* (2013).

[42] R. J. Albertini et al. "In vivo somatic mutations in humans: measurement and analysis". *Annu. Rev. Genet.* (1990).

[43] R. B. Cervantes et al. "Embryonic stem cells and somatic cells differ in mutation frequency and type". *Proc. Natl. Acad. Sci. U.S.A.* (2002).

[44] C. Kandoth et al. "Mutational landscape and significance across 12 major cancer types." *Nature* (2013).

[45] B. Alberts. *Molecular Biology of the Cell*. Reference Edition. Garland Pub, 2008. ISBN: 9780815341116.

[46] D. Schottenfeld. "Principles and applications of cancer prevention". In: *Cancer Epidemiology and prevention*. Ed. by D. Schottenfeld and J. F. Fraumeni. Oxford University Press, 1996.

[47] E González-Suárez et al. "Telomerase-deficient mice with short telomeres are resistant to skin tumorigenesis." *Nat Genet* (2000).

[48] H. W. Lee et al. "Essential role of mouse telomerase in highly proliferative organs." *Nature* (1998).

[49] E. González-Suárez et al. "Cooperation between p53 Mutation and High Telomerase Transgenic Expression in Spontaneous Cancer Development". *Mol. Cell. Biol.* (2002).

[50] K. L. Rudolph et al. "Longevity, stress response, and cancer in aging telomerase-deficient mice." *Cell* (1999).

[51] E Herrera. "Disease states associated with telomerase deficiency appear earlier in mice with short telomeres". *The EMBO Journal* (1999).

[52] B. Bernardes de Jesus et al. "Telomerase gene therapy in adult and old mice delays aging and increases longevity without increasing cancer". *EMBO Mol Med* (2012).

[53] B. M. Weon and J. H. Je. "Trends in scale and shape of survival curves." *Sci Rep* (2012).

[54] M. L. Rossi et al. "Roles of Werner syndrome protein in protection of genome integrity." *DNA Repair* (2010).

[55] L Aravind et al. "Conserved domains in DNA repair proteins and evolution of repair systems." *Nucleic Acids Res.* (1999).

[56] K Futami et al. "Role of Werner syndrome gene product helicase in carcinogenesis and in resistance to genotoxins by cancer cells." *Cancer Science* (2008).

[57] P.-O. Mari et al. "Dynamic assembly of end-joining complexes requires interaction between Ku70/80 and XRCC4". *Proc. Natl. Acad. Sci. U.S.A.* (2006).

[58] A. B. Houtsmuller et al. "Action of DNA repair endonuclease ERCC1/XPF in living cells." *Science* (1999).

[59] A. S. Balajee et al. "The Werner Syndrome Protein Is Involved in RNA Polymerase II Transcription". *Mol. Biol. Cell* (1999).

[60] P. M. Grierson et al. "BLM helicase facilitates RNA polymerase I-mediated ribosomal RNA transcription." *Human Molecular Genetics* (2012).

[61] K. A. Bernstein et al. "The RecQ DNA Helicases in DNA Repair". *Annu. Rev. Genet.* (2010).

[62] E. Reits and J. J. Neefjes. "From fixed to FRAP: measuring protein mobility and activity in living cells". *Nat Cell Biol* (2001).

[63] R. D. Phair and S. A. Gorski. "Measurement of Dynamic Protein Binding to Chromatin In Vivo, Using Photobleaching Microscopy". *Methods in enzymology* (2003).

[64] K. M. Bendtsen et al. "Dynamics of the DNA repair proteins WRN and BLM in the nucleoplasm and nucleoli". *Eur Biophys J* (2014).

[65] B. L. Sprague et al. "Analysis of Binding Reactions by Fluorescence Recovery after Photobleaching". *Biophysical Journal* (2004).

[66] K. P. Burnham. "Multimodel Inference: Understanding AIC and BIC in Model Selection". *Sociological Methods & Research* (2004).

[67] A.-N. Spiess and N. Neumeyer. "An evaluation of R2 as an inadequate measure for nonlinear models in pharmacological and biochemical research: a Monte Carlo approach". *BMC Pharmacol* (2010).

[68] K. P. Burnham and D. R. Anderson. *Model Selection and Multi-Model Inference*. A Practical Information-Theoretic Approach. Springer, 2002. ISBN: 9780387953649.

[69] F. Mueller et al. "Evidence for a Common Mode of Transcription Factor Interaction with Chromatin as Revealed by Improved Quantitative Fluorescence Recovery after Photobleaching". *Biophysical Journal* (2008).

[70] H. P. Erickson. "Size and shape of protein molecules at the nanometer level determined by sedimentation, gel filtration, and electron microscopy." *Biol Proced Online* (2009).

[71] J Braga and J. McNally. "A Reaction-Diffusion Model to Study RNA Motion by Quantitative Fluorescence Recovery after Photobleaching". *Biophysical Journal* (2007).

[72] S. A. Compton et al. "The Werner syndrome protein binds replication fork and holliday junction DNAs as an oligomer." *J. Biol. Chem.* (2008).

[73] V Srivastava et al. "BLM helicase stimulates the ATPase and chromatin-remodeling activities of RAD54". *Journal of Cell Science* (2009).

[74] M Huranová et al. "The differential interaction of snRNPs with pre-mRNA reveals splicing kinetics in living cells." *The Journal of Cell Biology* (2010).

[75] W.-H. Cheng et al. "Werner syndrome protein associates with gH2AX in a manner that depends upon Nbs1". *FEBS Letters* (2005).

[76] E. P. Rogakou. "DNA Double-stranded Breaks Induce Histone H2AX Phosphorylation on Serine 139". *Journal of Biological Chemistry* (1998).

[77] M. E. Van Royen et al. "Nuclear proteins: finding and binding target sites in chromatin". *Chromosome Research* (2011).

[78] D Chen and S Huang. "Nucleolar components involved in ribosome biogenesis cycle between the nucleolus and nucleoplasm in interphase cells." *The Journal of Cell Biology* (2001).

[79] K Sneppen et al. "Simplified Models of Biological Networks". *Annual review of biophysics* (2010).

[80] A. Y. Mitrophanov and E. A. Groisman. "Positive feedback in cellular control systems". *Bioessays* (2008).

[81] S Pigolotti et al. "Oscillation patterns in negative feedback loops". *Proc. Natl. Acad. Sci. U.S.A.* (2007).

[82] J Pérez-Martín et al. "Promoters responsive to DNA bending: a common theme in prokaryotic gene expression." *Microbiol. Rev.* (1994).

[83] K. E. Shearwin et al. "Transcriptional interference–a crash course." *Trends Genet.* (2005).

[84] P Wang et al. "Demonstration that the TyrR protein and RNA polymerase complex formed at the divergent P3 promoter inhibits binding of RNA polymerase to the major promoter, P1, of the aroP gene of Escherichia coli." *J. Bacteriol.* (1998).

[85] P. A. Hershberger et al. "Interference by PR-bound RNA polymerase with PRM function in vitro. Modulation by the bacteriophage lambda cI protein." *J. Biol. Chem.* (1993).

[86] B.-K. Cho et al. "The transcription unit architecture of the Escherichia coli genome." *Nat. Biotechnol.* (2009).

[87] P Schickor et al. "Topography of intermediates in transcription initiation of E.coli." *The EMBO Journal* (1990).

[88] N. B. Reppas et al. "The Transition between Transcriptional Initiation and Elongation in E. coli Is Highly Variable and Often Rate Limiting". *Mol. Cell* (2006).

[89] A. C. Palmer et al. "Potent Transcriptional Interference by Pausing of RNA Polymerases over a Downstream Promoter". *Mol. Cell* (2009).

[90] K. Sneppen et al. "A Mathematical Model for Transcriptional Interference by RNA Polymerase Traffic in Escherichia coli". *Journal of Molecular Biology* (2005).

[91] U Vogel and K. F. Jensen. "The RNA chain elongation rate in Escherichia coli depends on the growth rate." *J. Bacteriol.* (1994).

[92] S Ryu et al. "GalR-mediated repression and activation of hybrid lacUV5 promoter: differential contacts with RNA polymerase." *Gene* (1998).

[93] M. H. Jensen et al. "Sustained oscillations and time delays in gene expression of protein Hes1." *FEBS Letters* (2003).

[94] D Bratsun. "Delay-induced stochastic oscillations in gene regulation". *Proc. Natl. Acad. Sci. U.S.A.* (2005).

[95] B. Novák and J. J. Tyson. "Design principles of biochemical oscillators". *Nat Rev Mol Cell Biol* (2008).

[96] S Krishna et al. "Frustrated bistability as a means to engineer oscillations in biological systems". *Phys Biol* (2009).

[97] R. Guantes and J. F. Poyatos. "Dynamical Principles of Two-Component Genetic Oscillators ". *PLoS Comput. Biol.* (2006).

[98] H Song et al. "Dynamics of a Minimal Model of Interlocked Positive and Negative Feedback Loops of Transcriptional Regulation by cAMP-Response Element Binding Proteins". *Biophysical Journal* (2007).

[99] J Raspopovic et al. "Digit patterning is controlled by a Bmp-Sox9-Wnt Turing network modulated by morphogen gradients". *Science* (2014).

[100] G. M. Süel et al. "An excitable gene regulatory circuit induces transient cellular differentiation." *Nature* (2006).

[101] V. Sotiropoulos and Y. N. Kaznessis. "Synthetic tetracycline-inducible regulatory networks: computer-aided design of dynamic phenotypes". *BMC Syst Biol* (2007).

[102] B Schwanhäusser et al. "Global quantification of mammalian gene expression control". *Nature* (2011).

[103] B. Lewin. *Genes 8*. Prentice Hall, 2004.

[104] A. Mor et al. "Dynamics of single mRNP nucleocytoplasmic transport and export through the nuclear pore in living cells". *Nat Cell Biol* (2010).

[105] B. L. Timney et al. "Simple kinetic relationships and nonspecific competition govern nuclear import rates in vivo." *The Journal of Cell Biology* (2006).

[106] U. Alon. *An Introduction to Systems Biology*. Design Principles of Biological Circuits. CRC Press, 2006. ISBN: 1584886420.

[107] R. Loew et al. "Improved Tet-responsive promoters with minimized background expression". *BMC Biotechnol* (2010).

[108] S. Semsey and S. Adhya. "Regulatory genes". In: *Brenner's encyclopedia of genetics, second edition*. Academic Press, 2013.

[109] H Hirata et al. "Oscillatory Expression of the bHLH Factor Hes1 Regulated by a Negative Feedback Loop". *Science* (2002).

[110] J. Hasty et al. "Synthetic Gene Network for Entraining and Amplifying Cellular Oscillations". *Physical review letters* (2002).

[111] J. Stricker et al. "A fast, robust and tunable synthetic gene oscillator". *Nature* (2008).

[112] J. O. Haerter et al. "Communication Dynamics in Finite Capacity Social Networks". *Physical review letters* (2012).

[113] J. P. Onnela et al. "Structure and tie strengths in mobile communication networks". *Proc. Natl. Acad. Sci. U.S.A.* (2007).

[114] G Palla et al. "Quantifying social group evolution". *Nature* (2007).

[115] N. E. Friedkin. "Horizons of observability and limits of informal control in organizations". *Social Forces* (1983).

[116] A Portes. "Social Capital: Its Origins and Applications in Modern Sociology". *Annu. Rev. Sociol.* (1998).

[117] D. G. Rand et al. "Dynamic social networks promote cooperation in experiments with humans". *Proc. Natl. Acad. Sci. U.S.A.* (2011).

[118] J. Grujić et al. "Social Experiments in the Mesoscale: Humans Playing a Spatial Prisoner's Dilemma". *PLoS ONE* (2010).

[119] K. Fehl et al. "Co-evolution of behaviour and social network structure promotes human cooperation". *Ecology Letters* (2011).

[120] C. Gracia-Lazaro et al. "Heterogeneous networks do not promote cooperation when humans play a Prisoner's Dilemma". *Proc. Natl. Acad. Sci. U.S.A.* (2012).

[121] F. Fu et al. "Reputation-based partner choice promotes cooperation in social networks." *Phys Rev E Stat Nonlin Soft Matter Phys* (2008).

[122] N. E. Friedkin. "Information flow through strong and weak ties in intraorganizational social networks". *Social networks* (1982).

[123] D. G. Rand and M. A. Nowak. "Human cooperation". *Trends in Cognitive Sciences* (2013).

[124] B. Daniel. *Social Capital Modeling in Virtual Communities: Bayesian Belief Network Approaches*. Bayesian Belief Network Approaches. IGI Global, 2009. ISBN: 1605666645.

[125] J Essers. "Nuclear dynamics of RAD52 group homologous recombination proteins in response to DNA damage". *The EMBO Journal* (2002).

[126] C. Lukas et al. "Mdc1 couples DNA double-strand break recognition by Nbs1 with its H2AX-dependent chromatin retention." *The EMBO Journal* (2004).

[127] W Rodgers et al. "Transient association of Ku with nuclear substrates characterized using fluorescence photobleaching." *The Journal of Immunology* (2002).

[128] M Gossen and H Bujard. "Tight control of gene expression in mammalian cells by tetracycline-responsive promoters." *Proc. Natl. Acad. Sci. U.S.A.* (1992).