

Inaugural - Dissertation  
submitted to the  
Combined Faculties for the Natural Sciences and for Mathematics  
of the Ruperto-Carola University of Heidelberg, Germany  
for the degree of  
Doctor of Natural Sciences

presented by

**Tim Heinemann, M.S.**

born in: Frankfurt (Oder)

Date of oral examination: March 2015



# **Molecular Motives for the Robust Regulation of DNA Repair**

Referees: **Prof. Dr. Thomas Höfer**  
Theoretical Systems Biology  
German Cancer Research Center, Heidelberg

**Prof. Dr. Ursula Kummer**  
Modellierung Biologischer Prozesse  
Ruperto-Carola University, Heidelberg



## Contents

<b>Zusammenfassung</b>	i
<b>Summary</b>	iii
<b>1. Introduction</b>	1
1.1. DNA damage and its consequences . . . . .	1
1.1.1. Variety of DNA damage repair mechanisms . . . . .	1
1.2. Mammalian nucleotide excision repair . . . . .	1
1.3. NER as model system for chromatin associated processes . . . . .	2
<b>2. Mathematical Modelling of Nucleotide Excision Repair (NER)</b>	3
2.1. Fluorescent time-laps imaging of the NER process . . . . .	3
2.1.1. Locally inflicted UV-induced DNA damage . . . . .	3
2.1.2. Repair factor accumulation and dissociation occur on different time scales . . . . .	4
2.1.3. Direct measurement of DNA resynthesis . . . . .	8
2.1.4. Repair rate follows first order rate kinetic . . . . .	10
2.2. Kinetic model of NER . . . . .	12
2.2.1. Slow first-order reaction kinetic due to many fast interacting components	12
2.2.2. Model structure and parametrization . . . . .	13
2.3. Maximum likelihood approach together with PLE analysis identifies realistic model of NER . . . . .	19
2.3.1. A maximum likelihood approach for efficient model fitting . . . . .	19
2.3.2. NER model fits accumulation, FLIP and repair synthesis measurements	20
2.3.3. Identifiability analysis . . . . .	21
<b>3. Control analysis of the DNA repair rate</b>	29
3.1. Kinetic NER model predicts collective rate control . . . . .	29
3.1.1. Response coefficients . . . . .	29
3.1.2. Exploiting natural variability in protein expression to quantify rate control	32

3.2. Variability analysis of the repair rate . . . . .	39
3.2.1. Variable NER factor expression and inflicted lesions account for the distribution of repair rates . . . . .	39
<b>4. Co-expression of NER repair factors</b>	<b>45</b>
4.1. Nuclear expression of NER factors is strongly correlated . . . . .	46
4.1.1. Cross-correlation affects repair response . . . . .	46
4.1.2. Microscopy analysis of co-staining experiments . . . . .	47
4.1.3. Flow cytometry . . . . .	50
4.1.4. Cell cycle independent cross-correlation of NER factor expression . .	50
4.1.5. Measured average response agrees with model prediction . . . . .	53
<b>5. Discussion</b>	<b>55</b>
<b>6. Bibliography</b>	<b>63</b>
<b>A. Appendix</b>	<b>71</b>
A.1. Parameter estimation . . . . .	71
A.2. Cross-correlation . . . . .	73

## **Zusammenfassung**



## **Summary**



## **1. Introduction**

### **1.1. DNA damage and its consequences**

The impeccable preservation of DNA is surely one of the most essential requirements for cellular viability. On a single day DNA accumulates on average  $\sim 10^5$  damages per cell, which can be grouped into three different categories [1]. First, DNA constantly undergoes spontaneous unspecific reactions (mostly hydrolysis) that cause deamination and create abasic sites in an aqueous solution [2, 3]. Second, the most common type of DNA damage originates from the cellular metabolism, which produces various types of reactive oxygen and nitrogen species, lipid peroxidation products, endogenous alkylating agents, estrogen and cholesterol metabolites, and reactive carbonyl species. The subsequent molecular distortions range from several kinds of single-strand breaks to various oxidative base and sugar products [4, 5]. And third, DNA damages that are caused by external physical or chemical agents like sunlight, which induces approximately 30,000 pyrimidine dimers per hour in each exposed keratinocyte [6].

The consequences for cellular viability that arise from the diversity of DNA lesions are essentially twofold:

either mutagenic or cytotoxic effects

- Transcription relies on an intact template (Martijn 2014) - DNA damage cannot be replaced but has to be repaired (Martijn 2014) - influence on ageing and cancer (Martijn 2014) - repair mechanisms need to be robust... - genome instability - a hallmark of cancer (Bartek 2007)

#### **1.1.1. Variety of DNA damage repair mechanisms**

- DNA damage response -The repair systems include the direct reversal of damage, nucleotide excision repair (NER), base excision repair (BER), DSB repair, and interstrand crosslink repair (Hoijsmakers 2001)

### **1.2. Mammalian nucleotide excision repair**

- Global genome and transcription coupled NER

## **Key components of the NER pathway**

**NER as a model system for chromatin-associated processes** comparison between CPDs and 6-4PP

### **1.3. NER as model system for chromatin associated processes**

## **2. Mathematical Modelling of Nucleotide Excision Repair (NER)**

An essential advantage for the research on Nucleotide Excision Repair (NER) is the ability to inflict UV-induced DNA damages within a locally confined area in the cell nucleus [7]. This technique reformed the investigation of chromatin associated processes whose *in vivo* analysis was, until then, focusing on enzyme mobility and exchange dynamics at steady state performed with photobleaching experiments [8, 9]. In contrast, the enclosed distribution of single stranded DNA damages causes the local activation of a multi-protein repair machinery which in turn allows to study the *de novo* assembly and dissociation kinetics of single repair factors. Exploiting this standard experimental set-up we acquired a comprehensive data set comprising the accumulation and dissociation time series of seven individual repair enzymes. In addition, we were able to directly quantify newly synthesized DNA measuring the incorporation of a fluorescently modified nucleotide analogue.

Based on this data we developed a mathematical model that is able to explain the connection between the dynamic exchange of individual repair factors at the chromatin on one hand and the overall time scale of the repair process on the other. The exact parametrization of the model was adapted iteratively in correspondence with a profile likelihood estimation leading to identifiable parameters and hence an increased predictive power of the model. As a consequence our model can reliably predict the behaviour of the experimentally not accessible incised DNA intermediate.

The experiments reported in this chapter were planed and conducted by Martijn Luijsterburg (accumulation and FLIP experiments) and Paul Verbruggen (EdU incorporation).

### **2.1. Fluorescent time-laps imaging of the NER process**

#### **2.1.1. Locally inflicted UV-induced DNA damage**

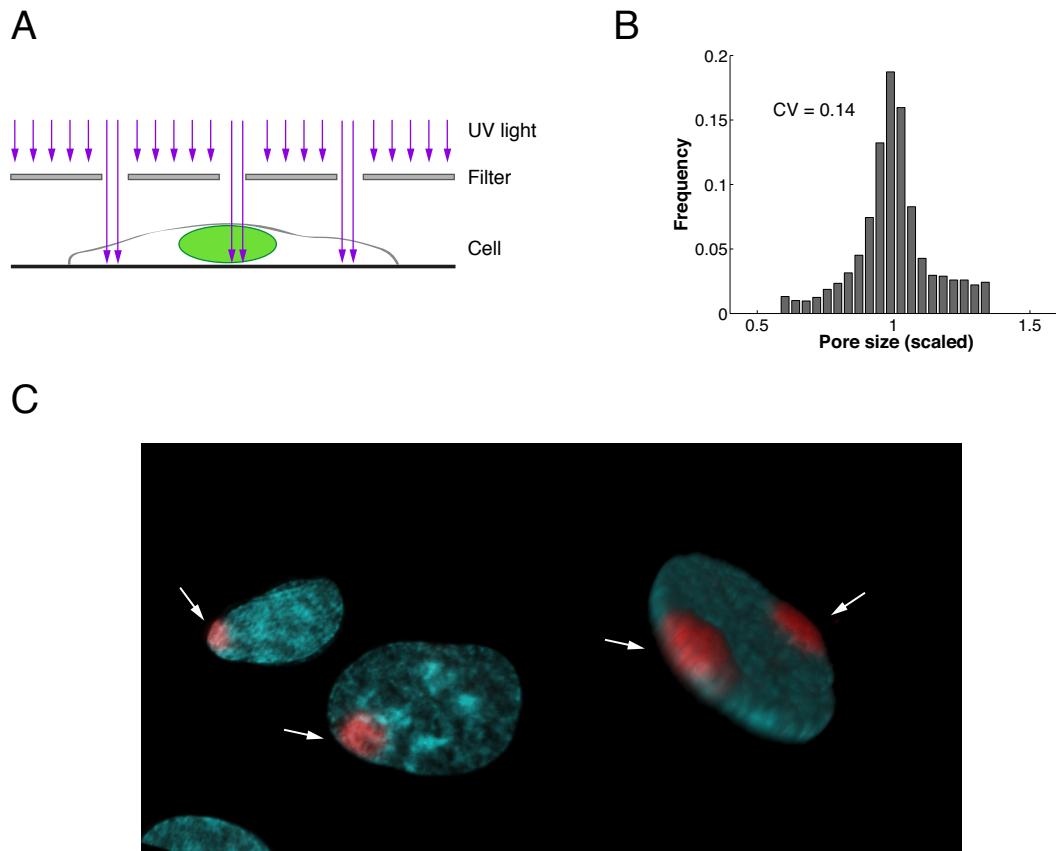
To observe the dynamic behaviour of NER factors upon local infliction of DNA damage we applied the experimental set-up developed by Moné *et al.* (2001) [7]. Therein, NER competent cells are covered with a polycarbonate UV filter containing pores with a given

diameter (cf. Figure 2.1A). For our purposes cells were grown on uncoated 24 mm coverslips before overlaid with a mask including pores with 5  $\mu\text{m}$  diameter [10]. Figure 2.1B illustrates the realistic deviation of pore diameters due to manufacturing inaccuracies. The actual size distribution has a very small CV of 0.14 indicating a negligible effect on the filter's transmissivity. Filter-covered cells were immediately irradiated with a dose of 100  $\text{J/m}^2$  of UV-C with a fluency of 3.85  $\text{W/m}^2$ . Due to the specific pore density ( $4 \times 10^5$  pores/ $\text{cm}^2$ ) we could observe nuclei containing either one damage spot or no spot at all. Very rarely two spots per cell were present (cf. Figure 2.1C, white arrows indicate the damage spot). With this technique we were able to measure NER in damaged nuclei as well as undamaged control nuclei in the same experiment.

### 2.1.2. Repair factor accumulation and dissociation occur on different time scales

The ability to locally inflict DNA damage with a discrete dose of UV-C allows to study the accumulation and exchange behaviour of fluorescently tagged repair proteins under different experimental conditions. In the following we describe the comprehensive dataset acquired by Luijsterburg *et al.* (2010) [6], which represents the basis of our quantitative analysis of the NER process. In total, the kinetics of seven repair factors were measured: i) XPC, the lesion recognition factor; ii) TFIIH, the helicase responsible for DNA unwinding; iii) XPA and RPA, which bind and thereby protect single stranded DNA against cleavage iv) the exonucleases XPF/ERCC1 and XPG performing the incision of the damaged DNA strand and v) PCNA, which loads the DNA polymerase and hence indirectly provides insight into the DNA repair-synthesis kinetic.

Each repair factor is tagged with a green fluorescent protein (GFP) (or its 'enhanced' derivative EGFP) and expressed at physiological levels within the cell nucleus. Before the cells are UV-irradiated at time  $t = 0$  the repair proteins are homogeneously distributed. Immediately afterwards the repair factors accumulate at the damaged chromatin sites, which leads to a higher visible fluorescence intensity (cf. Figure 2.2A). For quantification of the fluorescence intensity image analysis was done by using ImageJ software (NIH Bethesda, MD). Accumulated repair-factor concentrations were determined by multiplying the nuclear

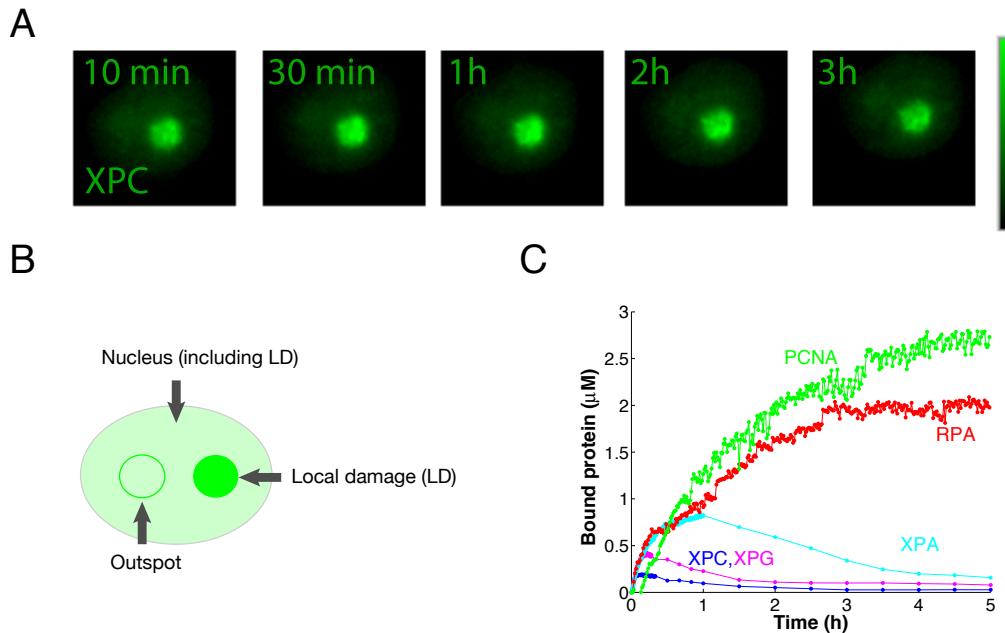


**Figure 2.1.: Local irradiation leads to spatially confined DNA damage.** A) Schematic illustration of the experimental setup. UV-light (purple arrows) is transmitted through a UV filter containing pores with 5  $\mu\text{m}$  diameter. Local irradiation of the chromatin occurs, if a pore is located above a cell nucleus. (reprinted Figure from [11]) B) Distribution of the pore sizes within the UV irradiation mask ( $n=5008$ ) - ask paul how he did it C) Partially UV-irradiated nuclei of cultivated mammalian cells were made visible with a blue-fluorescent DNA stain. Red fluorescence depicts incorporation of labelled nucleotides due to NER.

reference concentrations (cf. Table 2.1) with the fraction of bound proteins at damaged DNA:

$$\text{Bound fraction} = (I_{\text{LD}} - I_{\text{outspot}})A_{\text{LD}} / (I_{\text{nucleus}} - I_{\text{background}})A_{\text{nucleus}} \quad (2.1)$$

where  $I_{\text{LD}}$ ,  $I_{\text{outspot}}$ ,  $I_{\text{nucleus}}$  and  $I_{\text{background}}$  represent the average fluorescence intensities within the locally damaged spot, an equally sized area in the non-damaged nucleus, the whole nucleus and the background, respectively (cf. Figure 2.2B).  $A_{\text{LD}}$  and  $A_{\text{nucleus}}$  depict the damaged area and the size of the nucleus. Finally, the concentrations of accumulated protein are calculated assuming a damaged nuclear volume of 0.03 pL.



**Figure 2.2.: Fluorescently labelled NER factors accumulate at locally irradiated nucleus.** A) XPC-eGFP accumulation stably expressed in XP-C cells at different time points after local irradiation with UV-C ( $100 \text{ J m}^{-2}$  through 5-μm-diameter pores). B) Scheme of a locally UV-irradiated nucleus. Depicted are all regions relevant for signal quantification. C) Time courses of XPC-eGFP ( $n = 12$ ), XPG-eGFP ( $n = 5$ ), eGFP-XPA ( $n = 7$ ), eGFP-PCNA ( $n = 5$ ) and RPA-eGFP ( $n = 5$ ) showing their accumulation at the local damage (LD) spot. For consistency only cell nuclei with one single damage spot were used.

During the timespan of DNA repair (cf. Section 1.2) NER factors accumulate towards and then gradually decrease from their plateau levels at different rates (cf. Figure 2.2C). For

Protein	Concentration	Bound fraction
XPC	0.140 $\mu\text{M}$	13%
TFIILH	0.360 $\mu\text{M}$	10%
XPG	0.440 $\mu\text{M}$	9%
XPA	1.110 $\mu\text{M}$	7%
XPF/ERCC1	0.170 $\mu\text{M}$	7%
RPA	1.110 $\mu\text{M}$	15%
PCNA	1.110 $\mu\text{M}$	20%

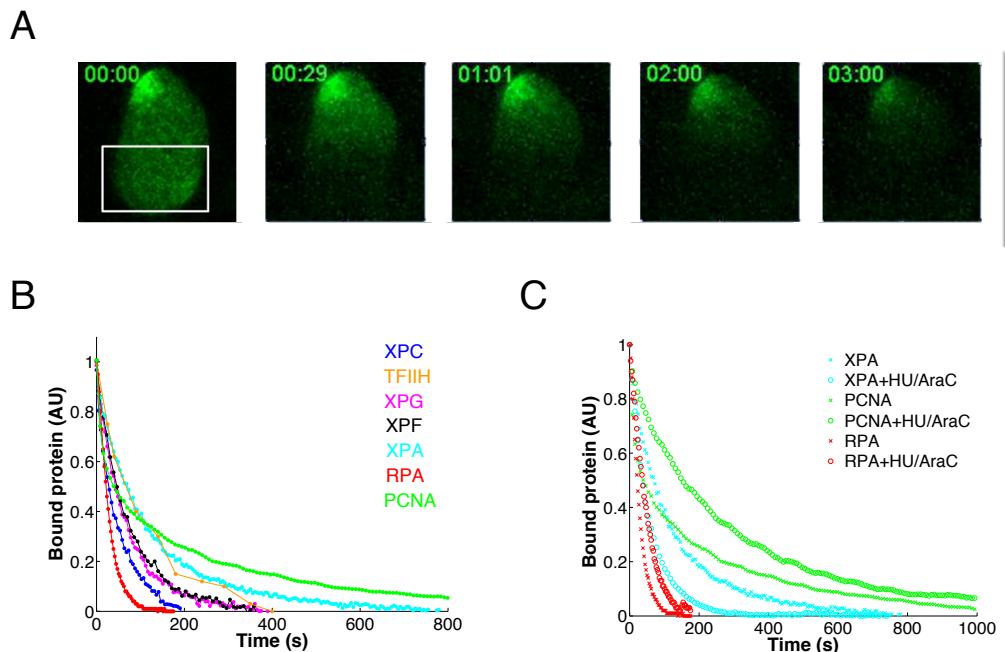
**Table 2.1.: Nuclear concentrations of NER factors (in  $\mu\text{M}$ )** All nuclear quantities are based on published data or on previous estimates [9, 12, 13]. The nuclear concentrations are calculated assuming a nuclear volume of 0.3  $\mu\text{L}$ . The bound fractions are determined by Eqn. 2.1 [6].

example the half-life  $t_{1/2}$  for XPC- and XPG-EGFP is  $\sim 1$  h whereas  $\sim 2.5$  h for XPA-EGFP [6]. Moreover, PCNA and RPA stay present in the damage spot even after lesion removal. These result show that NER factors engage for hours in the repair process with temporal changes in the molecular composition.

To characterize the interaction between repair proteins and DNA intermediates dwell times were determined by fluorescence loss in photobleaching (FLIP) experiments [6]. Thereby, a large part of the nucleus, away from the local damage spot, is continuously bleached at 100 % laser power (cf. 2.3A, white rectangle). At the same time fluorescent proteins are probed at low laser power elsewhere within the nucleus. Repair proteins dissociating from the local damage spot have a high probability to be bleached before rebinding due to their large diffusivity. Accordingly, for the dwell time of the repair proteins binding instead of diffusion seems to be rate limiting [6].

Still, compared to the long timespan (in the order of hours) where repair protein levels are still elevated at site of local damage, all NER factors dissociate very quickly from damaged DNA with half-lives of 20 s (RPA), 25 s (XPC), 50 s (TFIILH, XPG, ERCC1/XPF) and 80 s (RPA) (cf. Figure 2.3B). For PCNA the dissociation is strongly biphasic with half-lives of 10 s and 225 s respectively. To analyse, whether the dwell time of slowly accumulating NER factors changes throughout the repair process resynthesis was stalled by addition of the drugs hydroxyurea (HU) and cytosine- $\beta$ -arabinofuranoside. While NER factors, reaching their plateau level earlier, were not affected [6], omitting the repair progression at this late stage mainly slowed down the dissociation of PCNA and RPA (cf. Figure 2.3C). In contrast, XPA's half-life decreased by twofold indicating its higher affinity to repair synthesis intermediates.

This shows that the dwell times of NER factors change as repair progresses and suggests that their affinity towards damaged chromatin is defined by the state of the DNA substrate.



**Figure 2.3.: Rapid dissociation of NER factors from damaged DNA.** A) FLIP experiment in XP2OS cells stably expressing eGFP-XPA after 2 h of local irradiation. The nucleus is continuously bleached in an undamaged region (white rectangle). Fluorescence loss is recorded at the sites of local UV-irradiation B) Amount of bound NER-factors (XPC-eGFP, TFIID-eGFP, XPG-eGFP, XPF-GFP, eGFP-XPA, RPA-eGFP and eGFP-PCNA) monitored over time at LD. C) Quantification of FLIP experiments in the absence or presence of HU and AraC for stably expressed eGFP-XPA, RPA-eGFP and eGFP-PCNA.

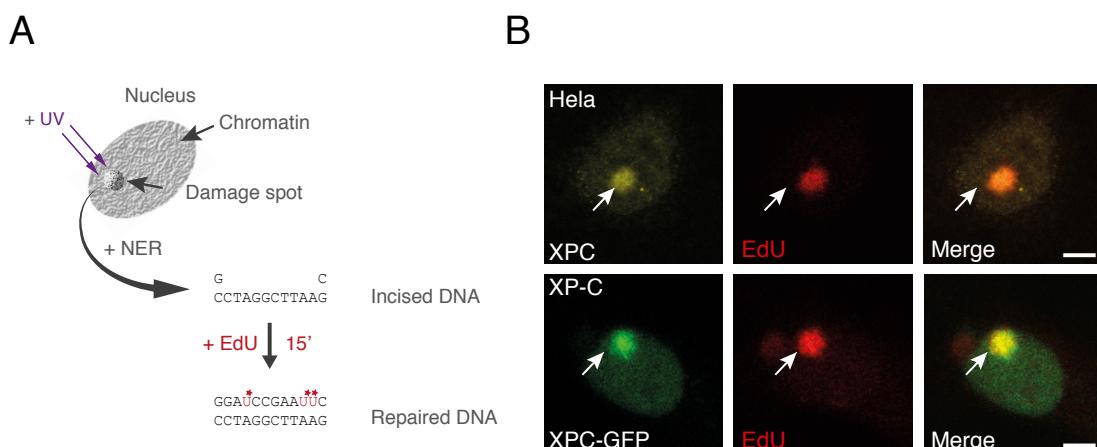
### 2.1.3. Direct measurement of DNA resynthesis

In order to expand our quantitative inspection of the NER pathway in intact mammalian cells we established a protocol for the direct measurement of the repair synthesis process. DNA resynthesis reflects the kinetic of the post-incision repair process, complementing the measurement of DNA lesion removal, which captures the systems behaviour of the pre-incision steps (cf. section 1.2). To visualize newly incorporated DNA shortly before and after local UV irradiation cells were incubated in microscopy medium supplemented

## 2. Mathematical Modelling of Nucleotide Excision Repair (NER)

with 10  $\mu\text{M}$  5-ethynyl-2'-deoxyuridine (EdU). Due to its excessive presence in solution the DNA polymerase integrates EdU (a thymidine analogue) into the new DNA strand instead of the endogenous thymine (cf. Figure 2.4A). After incubation for the desired time cells were fixed stopping EdU incorporation and subsequently permeabilized. EdU is then tagged with the fluorescent azide (AlexaFluor-555, Life Technologies) forming a covalent bond by click chemistry [14]. Analogous to the quantification of NER factor dynamics EdU intensities were captured with a laser scanning microscope (Zeiss).

Accordingly, incorporated EdU is exclusively present at the locally confined area of damaged DNA, which coincides with the localization of immunostained XPC (cf. Figure 2.4B, upper row). The same result occurs for XP-C cells with stably transfected XPC-eGFP (cf. Figure 2.4B, lower row). Replicating cells could be excluded easily from the analysis due to their prevalent EdU incorporation distributed over the entire nucleus.



**Figure 2.4.: EdU is incorporated at sites of local DNA damage.** A) Schematic illustration depicting the experimental procedure of EdU incorporation at sites of local damage. B) Endogenous XPC in HeLa cells (upper panel) or stably expressed XPC-eGFP (lower panel) accumulate in the LD spot and co-localize with incorporated EdU.

#### **2.1.4. Repair rate follows first order rate kinetic**

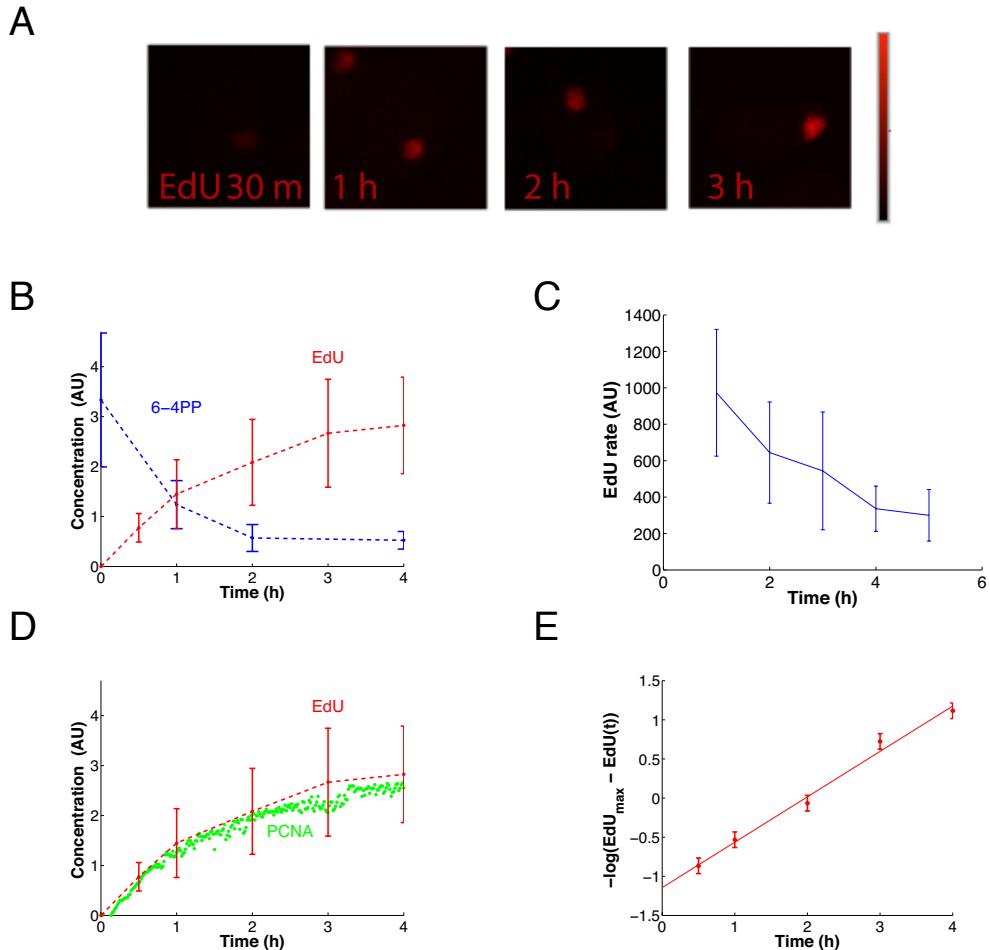
In contrast to the real-time measurements for accumulation and dissociation of the NER factors the incorporation of EdU cannot be followed continuously. As mentioned in section 2.1.3 cells have to be fixed and permeabilized before the newly incorporated bases can be fluorescently labelled. Therefore, only the accumulated EdU incorporated in the time interval between UV-irradiation and fixation can be followed. By repeating this procedure for growing time intervals we acquired successively the repair kinetic for newly repaired DNA (cf. Figure 2.5A and B). The signal at each time point represents the EdU signal averaged over multiple cells.

We found that EdU incorporation essentially stops after 4 h, which coincides with the removal of 6-4PP (cf. [6] and Figure 2.5B red and blue curve, respectively). This agrees with the observation that NER is not primarily engaged in the removal of cyclobutane pyrimidine dimers (CPD) which are repaired on a much longer time scale. To test whether the availability of EdU is rate limiting we measured the incorporation of EdU in discrete equidistant time intervals after UV-irradiation (cf. Figure 2.5C). As it turns out the amount of EdU incorporation per time interval is indeed continuously declining and hence the rate of repair synthesis. Moreover the EdU kinetic follows the trajectory of PCNA accumulation as measured by Luijsterburg *et al.* [6] (cf. Figure 2.5D). PCNA is thought to act as processivity factor for the DNA polymerase and remains bound to the DNA [6, 15, 16]. Taken together, these data establish EdU incorporation as a direct and quantitative measure of DNA repair synthesis in locally damaged cells.

The DNA repair time series characterized by incorporated EdU is fitted by a mono-exponential kinetic (cf. Figure 2.5E):

$$EdU(t) = EdU_{\max}(1 - e^{\lambda t}), \quad (2.2)$$

where  $EdU(t)$  and  $EdU_{\max}$  depict the amount of incorporated EdU and its value at saturation, respectively. The time constant is  $\lambda=0.58 (\pm 0.07) \text{ h}^{-1}$ . This result indicates that, despite its molecular complexity, 6-4PP removal by NER is a slow first-order reaction with a half-time of 1.2 hours.



**Figure 2.5.: EdU incorporation reflects DNA repair synthesis quantitatively.** A) EdU signal at various time points after local UV-C irradiation. B) Repair DNA synthesis (EdU, red curve) coincides with damage removal (6-4PP, blue curve) measured previously by Luijsterburg *et al.*(2010) [6]. The EdU trajectory represents mean  $\pm$  SD (derived from three independent experiments) of DNA repair in  $n=150$  locally damaged cells per time point. C) Irradiated XP-C XPC-eGFP cells were treated with EdU for different time intervals post-irradiation. Intervals were chosen as follows: 15 minutes pre-irradiation to 60 minutes post-irradiation, 45-120 minutes, 105-180 minutes, 165-240 minutes and 225-300 minutes. Line plot of experimental means derived from distributions associated to a particular time interval. D) DNA repair synthesis follows PCNA accumulation as measured previously [6]. E) Mono-exponential fit of the EdU kinetic according to Eqn. 2.2.

## 2.2. Kinetic model of NER

### 2.2.1. Slow first-order reaction kinetic due to many fast interacting components

To investigate the mechanistic connection between fast NER factor exchange at the DNA template and the overall slow repair time it appears practical to apply an analytical approach. During her Ph.D. thesis Gesa Terstiege supervised by Prof. Thomas Höfer performed this analysis [10, 11], considering the complex formation with a simplified model including  $N$  repair factors. They examined different scenarios distinguishing random and sequential protein assembly; reversible and irreversible protein binding or mixed mechanisms specific for each DNA repair intermediate (cf. Figure 2.6A, [10]). Under appropriate assumptions, the mean repair time of such a process can be generally expressed with:

$$\tau = \underbrace{\frac{1}{k} \sum_{i=0}^{N-1} A_i \left(\frac{l}{k}\right)^i}_{\text{first assembly}} + \underbrace{\frac{1}{\rho} \sum_{i=0}^N B_i \left(\frac{l}{k}\right)^i}_{\text{reassembly and reaction}} \quad (2.3)$$

where  $k$  denotes the pseudo-first order association rate constant of NER factors,  $l$  the dissociation rate constant of NER factors and  $\rho$  the rate of the repair reaction.  $A_i$  and  $B_i$  differ between random assembly

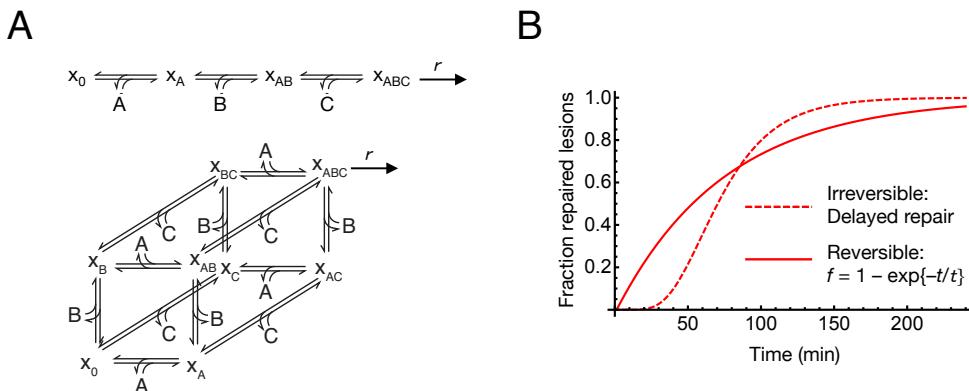
$$A_i^{\text{rand}} = \sum_{j=1}^{N-i} \frac{1}{i+j} \frac{\binom{N}{j-1}}{\binom{N}{i+j}}, \quad B_i^{\text{rand}} = \binom{N}{i}, \quad (2.4)$$

and sequential assembly

$$A_i^{\text{seq}} = N - i, \quad B_i^{\text{seq}} = 1. \quad (2.5)$$

Sequential and random assembly occur on a similar time scale for a number of repair components smaller than 10 considering reversible protein binding ( $l = k = 1$ ) [11], which coincides with the number of core NER factors [6]. The theoretical component number for equal repair times is even higher for increasing reaction rates ( $\rho \gg k, l$ ). Given that all repair factor dwell times are in the order of  $\sim 1$  min ( $k = l = 1 \text{ min}^{-1}$ ) the model was simulated for  $N = 9$  components. Remarkably, the resulting trajectory followed a single-exponential kinetic with a half-time of  $\sim 1$  hour in case of balanced reversibility. In contrast, for an

irreversible process ( $l/k = 0$ ) the repair kinetic resembles a sigmoid time course (cf. Figure 2.6B). These results suggest that the rapidly exchanging NER factors naturally generate a mono-exponential repair kinetic whereas the slow overall time span can be explained by the stochastic distribution of repair events.

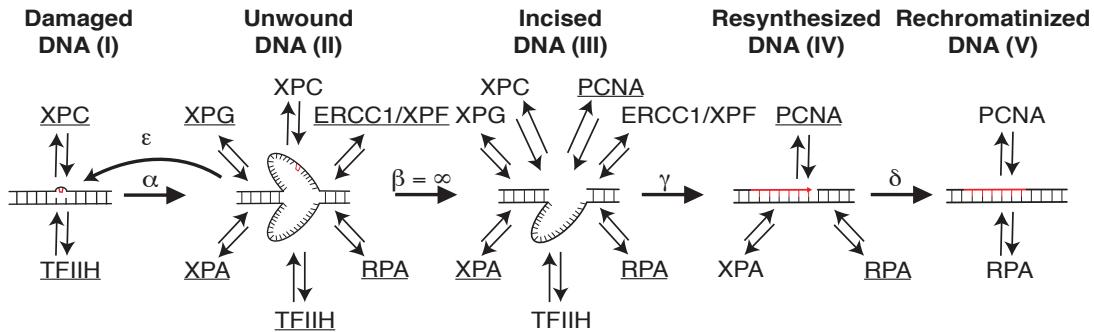


**Figure 2.6.: Simplified analytical model of the repair complex formation** A) Sequential (above) and random (below) assembly scheme of the repair factors A, B and C to the DNA template  $x_i$ .  $x_0$  denotes the empty lesion (e.g. damaged DNA). When the complex is fully assembled ( $x_{ABC}$ ) it performs a catalytic repair step with rate  $r$ . B) Simulated repair time courses for random reversible (solid line) and irreversible (dashed line) repair factor assembly. For reversible binding ( $k = l = 1 \text{ min}^{-1}, N = 9$ ) the trajectory fits a mono-exponential repair kinetic with time constant  $\tau$  whereas for irreversible binding the fit has sigmoidal shape ( $l = 0, k = 0.037 \text{ min}^{-1}, N = 9$ , chosen to get the same time constant).

### 2.2.2. Model structure and parametrization

To examine the relation between rapid repair factor exchange and the slow first-order reaction kinetic we extended this analysis on the basis of a realistic NER model. Conceptually, it follows a model worked out by Luijsterburg *et al.* (2010) [6]. In this model NER factors bind transiently to DNA repair intermediates thereby forming catalytic complexes that, if complete, perform the next repair step. These are usually irreversible reactions embodying the sequential characteristics of this pathway (cf. Figure **tbm**).

The nature of these distinguishable repair intermediates is widely investigated in biochemical but also in *in vivo* studies [17–20]. Accordingly, the evolved model distinguishes five DNA repair intermediates: i) damaged DNA, unwound DNA, incised DNA, resynthesised DNA



**Figure 2.7.: Schematic model description of the DNA repair mechanism.** The model distinguishes five individual repair intermediates: Damaged (I), unwound (II), incised (III), resynthesised (IV) and rechromatinised DNA (V). As indicated, specific tuples of repair proteins bind reversibly to the intermediates. Catalytic reactions, denoted by Greek letters, occur when the catalytic reaction complex (underlined proteins) is assembled.

and rechromatinised DNA. The molecular state of these intermediates defines the binding affinity for a specific set of repair proteins (cf. Figure 2.7). Table 2.2 summarizes all repair intermediates and lists the corresponding repair factors that show affinity for a particular repair intermediate. Repair factors that catalyse the enzymatic reaction if assembled are indicated. As stated before, FLIP measurements indicate that diffusion is not rate limiting for protein binding and thus not included in the model (cf. Section 2.1.2 and [21, 22]).

A prominent role within the NER pathway belongs to the lesion recognition factor XPC. Theoretical results suggest that the detection of damaged DNA sites by a single element instead of multiple simultaneously is crucial for the efficient initiation of the repair process [23]. Beside this exception the model allows random, non-cooperative binding for all NER factors.

The model structure introduced above and shown in Figure 2.7 was translated into an ordinary differential equation system assuming mass-action kinetics for all protein-DNA interaction processes. Each equation describes the concentration of a single DNA state  $y_\pi^R$  (cf. Eqn. 2.6) associated with a specific repair intermediate  $R = I, II, \dots, V$  (damaged (I), unwound (II), incised (III), resynthesised (IV), rechromatinised (V)). The index  $\pi$  represents a binary vector where each position displays the presence or absence of a repair protein  $p$  ( $p \in C, T, G, A, F, R, P$ , where XPC (C), TFIID (T), XPG (G), XPA (A), ERCC1/XPF (F), RPA (R) and PCNA (P)).  $\pi(p) = 1$  if protein  $p$  is bound and  $\pi(p) = 0$  if not. The length

of  $\pi(p)$  is defined according to the model structure (cf. Table 2.2) with:  $\pi_{R=I} = \{\text{C,T}\}$ ;  $\pi_{R=II} = \{\text{C,T,G,A,F,R}\}$ ;  $\pi_{R=III} = \{\text{C,T,G,A,F,R,P}\}$ ;  $\pi_{R=IV} = \{\text{A,R,P}\}$ ;  $\pi_{R=V} = \{\text{R,P}\}$ . The time course of this model is then described by:

$$\frac{d}{dt} y_\pi^R = \sum_{p \text{ in } R} \eta \left( (-1)^{\pi(p)} l_p^R y_\pi^R |_{\pi(p)=1} + (-1)^{1+\pi(p)} k_p^R C_p(t) y_\pi^R |_{\pi(p)=0} \right) + E(y_\pi^R), \quad (2.6)$$

where  $\eta$  represents a cooperativity ensuring the inclusion of all cooperative binding events:

$$\eta = \begin{cases} = 0 & \text{if } R = \text{I} \wedge p = \text{C} \wedge \text{T} = 1, \\ & R = \text{I} \wedge p = \text{T} \wedge \text{C} = 0, \\ = 1 & \text{else.} \end{cases}$$

The kinetics describing protein exchange at the repair intermediates are characterized by the binding  $k_p^R$  and dissociation constants  $l_p^R$ , respectively. The free protein concentrations are denoted by  $C_p(t)$  representing seven additional differential equations for  $p \in \{\text{C,T,G,A,F,R,P}\}$ :

$$\frac{d}{dt} C_p = \sum_{R=1}^V \sum_{\pi} \xi \left( \delta_{\pi(p)1} l_p^R y_\pi^R - \delta_{\pi(p)0} k_p^R C_p y_\pi^R \right) \quad (2.7)$$

Analogous to  $\eta$ ,  $\xi$  governs the sequential binding of XPC and TFIIH by:

$$\xi = \begin{cases} = 0 & \text{if } R = \text{I} \wedge \text{C} = 0 \wedge \text{T} = 1, \\ & R = \text{I} \wedge p = \text{C} \wedge \text{C} = \text{T} = 1, \\ & R = \text{I} \wedge p = \text{T} \wedge \text{C} = \text{T} = 0, \\ = 1 & \text{else.} \end{cases}$$

*2. Mathematical Modelling of Nucleotide Excision Repair (NER)*

<b>Repair intermediate</b>	<b>Binding proteins</b>	<b>Catalyzed process Required proteins</b>	<b>Remarks</b>	<b>Ref.</b>
(I) Damaged DNA	XPC,TFIIH (3 states)	Unwinding (reaction $\alpha$ ) XPC and TFIIH	Initiation by binding of XPC and subsequent recruitment of TFIIH.	[17] [24] [25] [21] [26]
(II) Unwound DNA	XPC,TFIIH, XPG, XPA, ERCC1/XPF, RPA (64 states)	Dual incision (reaction $\beta$ ) TFIIH,XPG, XPA, RPA and ERCC1/XPF	If the DNA becomes devoid of any protein, it will re-anneal (reaction $\varepsilon$ ). Dual incision requires the endonucleases XPG and ERCC1/XPF and is stimulated by TFIIH, XPA, RPA and possibly XPC.	[17] [27] [28] [29] [30]
(III) Incised DNA	XPC,TFIIH, XPG, XPA, ERCC1/XPF, RPA, PCNA (128 states)	Repair-synthesis (reaction $\gamma$ ) XPA, RPA and PCNA	PCNA binds to the free 3'-OH group generated by the ERCC1/XPF incision. DNA polymerase is also required (not measured).	[17] [29]
(IV) Resynthesized DNA	XPA, RPA, PCNA (8 states)	Rechromatinization (reaction $\delta$ ) RPA and PCNA	Accumulation data imply that XPA binds to repaired DNA while the pre-incision proteins do not (Figure 2.2C).	[31] [32] [6]
(V) Rechromatinized DNA	RPA, PCNA (4 states)		RPA and PCNA associate with rep. intermediate I, as levels of bound EGFP- PCNA and EGFP-RPA are high up to at least 4h after UV irradiation while other repair proteins are no longer bound.	[24] [6]

**Table 2.2.: Model assumptions.** Adapted from Terstiege *et al.* (2010) [11]

## 2. Mathematical Modelling of Nucleotide Excision Repair (NER)

---

Finally, if an enzymatic complex has fully assembled at the DNA template it catalyses the next repair step, which is represented in the model with the term  $E(y_\pi^R)$ . After damage recognition by XPC damaged DNA is unwound by the helicase TFIIH with the unwinding activity  $\alpha$ , whereas XPC acts as a stabilizing/proof reading factor in parallel. Accordingly  $E(y_\pi^R)$  translates into the following catalytic reactions for damaged DNA ( $R = \text{I}$ ):

$$E(y_{00}^{\text{I}}) = \varepsilon y_{000000}^{\text{II}} \quad \text{and} \quad E(y_{11}^{\text{I}}) = -\alpha y_{11}^{\text{I}},$$

If all proteins fall off the DNA template due to false damage detection the DNA will re-anneal with the rate  $\epsilon$ . Otherwise a complex formed by TFIIH, XPG, XPA, XPF, XPA and RPA will eventually promote the excision of the lesion DNA strand leading to the following catalytic reactions for unwound DNA ( $R = \text{II}$ ):

$$\begin{aligned} E(y_{000000}^{\text{II}}) &= -\varepsilon y_{000000}^{\text{II}}, & E(y_{110000}^{\text{II}}) &= \alpha y_{11}^{\text{I}}, \\ \text{and} & & E(y_{011111}^{\text{II}}) &= -\beta y_{011111}^{\text{II}}. \end{aligned}$$

Once the lesion strand is excised with incision rate  $\beta$  the remaining repair steps are irreversible. Incised DNA is resynthesised with the rate  $\gamma$  by the resynthesis complex XPA-RPA-PCNA. XPA is assumed to assemble at post incision repair intermediates as suggested by experiments with inhibited incision that showed accelerated dissociation FLIP kinetics for XPA [6]. This result is supported by a chromatin immunoprecipitation (ChIP) experiment with antibodies against XRCC1-Lig III showing the co-precipitation of XPA and RPA but not XPC and TFIIH [33]. Evidence for the importance of PCNA and RPA for the resynthesis reaction was shown by Shivji *et al.* (1995) [32]. This results into the following catalytic reactions for incised DNA ( $R = \text{III}$ ):

$$E(y_{111111}^{\text{III}}) = \beta y_{011111}^{\text{II}} \quad \text{and} \quad E(y_{0001011}^{\text{III}}) = -\gamma y_{0001011}^{\text{III}}.$$

In correspondence to the previously described accumulation measurements (cf. Figure 2.2) only RPA and PCNA stay bound during chromatin remodelling, the last modelled repair intermediate. This leads to the following enzymatic reactions for resynthesised DNA ( $R = \text{IV}$ )

and rechromatinised DNA ( $R = \text{IV}$ ):

$$E(y_{111}^{\text{IV}}) = \gamma y_{0001011}^{\text{III}}, \quad E(y_{011}^{\text{IV}}) = -\delta y_{011}^{\text{IV}},$$

and  $E(y_{11}^{\text{V}}) = \delta y_{011}^{\text{IV}}$ .

Since all NER factors can bind independently to the DNA template for each repair intermediate there are a total of  $2^N$  repair states where  $N$  is the number of repair proteins assembling to the particular repair intermediate. Due to the sequential assembly of TFIIH after lesion detection of XPC the number of states for damaged DNA is reduced to  $2^2 - 1 = 3$  states. For the remaining repair intermediates we derive  $2^6 = 64$  for unwound DNA;  $2^7 = 128$  states for incised DNA;  $2^3 = 8$  states for resynthesised DNA and  $2^2 1 = 4$  states for rechromatinised DNA. This results in a total of 214 states including seven differential equations for the free NER-factor protein concentrations.

Summing over all repair states and the respective intermediates associated to one repair factor we can simulate its accumulation kinetic. As initial conditions serve the measured free protein concentrations denoted in Table 2.1 [6, 11] and the initial amount of inflicted damages whose concentration was estimated with  $y_{00}^1 = 3.33 \mu\text{M}$ . The remaining states start at zero. To reproduce the FLIP kinetic for a specific repair factor all corresponding dissociation constants were set to zero at the time of maximal accumulation when the FLIP experiment started. Accordingly, FLIP kinetics were acquired after 600 s for XPC and ERCCC1/XPF, 900 s for XPG and TFIIH and 2000 s for XPA and 7200 s for PCNA.

### 2.3. Maximum likelihood approach together with PLE analysis identifies realistic model of NER

#### 2.3.1. A maximum likelihood approach for efficient model fitting

To find a realistic parametrisation for the temporal development of the repair states  $y_\pi^R$  (cf. Eqn. 2.6) the model was mapped to  $m$  observables via the function  $f_z$ :

$$z(t_i, \theta) = f_z(t_i, y(t_i, \theta), \theta). \quad (2.8)$$

The observables  $z$  are parametrised by  $\theta$  and resemble experimentally derived quantities at time  $t_i$ .  $\theta$  depicts binding, dissociation and catalytic constants adding up to a total of 45 model parameters. Each observable  $z_k(t_i, \theta)$  corresponds to the measured data  $z_k^\dagger(t_i)$  with intrinsic noise  $\epsilon_{ki}$ . Its origin can be traced back to a mixture of measurement noise combined with the naturally occurring biological variability. Assuming additive, normally distributed noise leads to  $z_k(t_i)^\dagger = z_k(t_i, \theta) + \epsilon_{ki}$  with  $\epsilon_{ki} \sim N(0, \sigma_{ki}^2)$ . To calibrate the measured data  $z_k(t_i)^\dagger$  with the model observables  $z_k(t_i, \theta) + \epsilon_{ki}$  we applied a maximum likelihood approach as distance measure which translates for normally distributed noise into:

$$L(z^\dagger | \theta) = \prod_{k=1}^m \prod_{i=1}^{d_k} \frac{1}{\sqrt{2\pi\sigma_{ki}^2}} \exp \left( -\frac{1}{2\sigma_{ki}^2} (z_{ki}^\dagger - z_k(t_i, \theta))^2 \right) \quad (2.9)$$

Here,  $d_k$  denotes the number of distinct experimental data sets  $z^\dagger$  for each observable  $k \sim 1 \dots m$ , measured at time points  $t_i$  with  $i \sim 1 \dots d_k$ .  $\sigma_{ki}^2$  are the variance components of the measurement noise of each data point. Instead of maximizing the likelihood it is equivalent and numerically more efficient to minimize its negative logarithm  $-2\log(L(z^\dagger | \theta))$  instead. In the following we will refer to it as  $X^2$ . For the minimization of  $X^2$  the choice for  $\theta$  is controlled by  $\sigma_{ki}^2$  (cf. Eqn. 2.9). As shown by Raue and colleagues for the reliable estimation of the model parameters  $\theta$  the simultaneous approximation of  $\sigma_{ki}^2$  together with the model dynamics leads to a statistically more accurate assessment of the model parameters than using noise estimations from preprocessed data [34]. Accordingly  $\sigma_{ki}^2$  can be considered as parametrised function

$$\sigma_k(t_i, \theta) = f_{\sigma_k}(t_i, z(t_i, \theta), \theta). \quad (2.10)$$

wherein additional parameters are introduced representing the type and magnitude of the modelled noise. Analogous we applied an additive error model for each observable with the parametrised function  $\sigma_k(t_i, \theta) = s_a$  and  $\epsilon_{ki} \sim N(0, \sigma_{ki}^2)$  where  $s_a$  is included in  $\theta$ .

For reasons of fitting-speed efficiency we implemented our model into the online-available D2D software environment [34] optimized for MATLAB (2011a, The Mathworks Inc., Natick, MA). The integrated fitting procedure applies the trust region algorithm LSQNONLIN, which is pre-implemented in MATLAB. The algorithm requires the derivatives of the objective function with respect to the parameters (cf. Eqn. 2.9). The inner derivatives  $dy(t, /theta)/d\theta$ , also called sensitivities, provide gradient information about the parameter landscape and thus are crucial to guide the optimization algorithm to the nearest optimum. The sensitivities can be passed in form of sensitivity equations

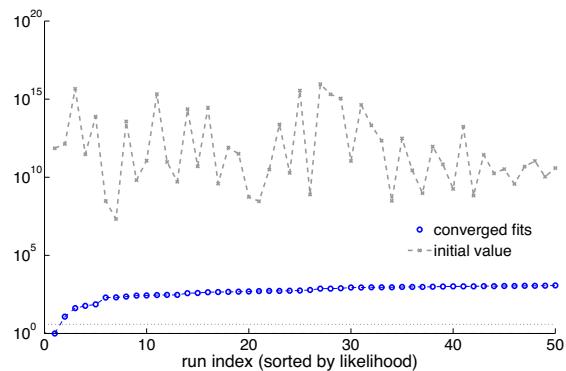
$$\frac{d}{dt} \frac{dy(t, \theta)}{d\theta} = \frac{\partial f_y}{\partial y} \frac{dy(t, \theta)}{d\theta} + \frac{\partial f_y}{\partial \theta}, \quad (2.11)$$

which represent additional ODEs that are solved in parallel to the original ODE system (cf. Eqn. 2.6) [35]. Applying sensitivity equations instead of a simple finite difference approximation proved to be numerically more accurate and computationally also faster [34]. Both, model and sensitivity ODEs were solved with the CVODES algorithm written in ANSI standard C [36].

To avoid terminating the optimization procedure in a local minimum we used a 'multi-start' approach by drawing the initial parameter vector using Latin hypercube sampling (LHS) [37]. In contrast to a random sampling approach LHS provides a better coverage of the sampling space by maximizing the distance between successive parameter drawings [34].

### 2.3.2. NER model fits accumulation, FLIP and repair synthesis measurements

To derive faithful fitting results we reiterated the optimization process for 250 times. For each iteration the starting parameters were redrawn by Latin hypercube sampling. Despite the size of the model, with respect to data points and the number of parameters, the majority of fits terminated relatively close to the global  $X^2$  minimum (cf. Figure 2.8). Nevertheless



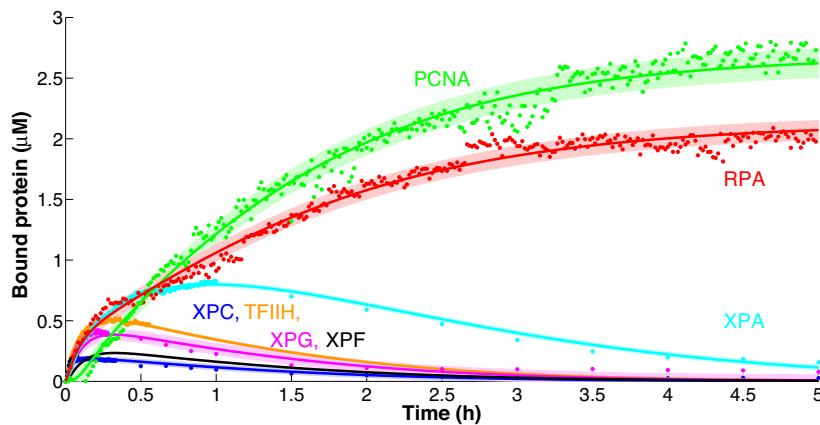
**Figure 2.8.: Parameter estimation for the 50 best fits after initial Latin hypercube sampling** Visualization of performance using 50 independent optimization runs (blue dots depict final optima of each individual run). Initial guesses (grey stars) were generated by Latin hypercube sampling. For illustrative purposes the global optimum is centred to 1. **put objective function on the left/axis handles bigger...**

two distinct minima can be distinguished suggesting a close local minimum, which 'hides' the rarely reached global minimum. For both parameter sets the model fits the experimental data set comprising accumulation, FLIP, perturbation and repair synthesis measurements with small estimated measurement errors (cf. Figure 2.9A, B, C). The accumulation kinetics (cf. Figure 2.9A) are depicted as concentrations scaled by the volume of the locally damaged area which was assumed to be 10% of the nuclear volume. We believe that this specification is more intuitively comprehensible compared to scaling by the whole nuclear volume as performed by Luijsterburg *et al.* [6]. Moreover it allows the realistic comparison between simulated and measured microscopy images of NER factor accumulation and EdU incorporation (cf. Figure 2.10).

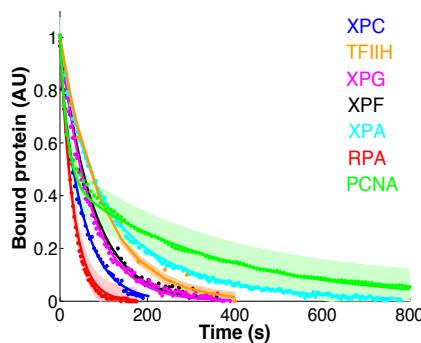
### 2.3.3. Identifiability analysis

In the following section we want to quantify the quality of the model fit and determine whether the current model structure is competent for reliable predictions concerning experimentally unobserved system behaviour. This capability depends on the structural and practical identifiability of the model, which can be influenced by functionally related parameters or limited amount and quality of the data, respectively [38, 39]. Both can be analysed and tested numerically by a formalism called profile likelihood estimation (PLE) [40–42], where

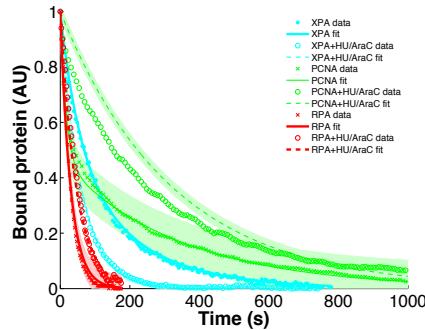
A



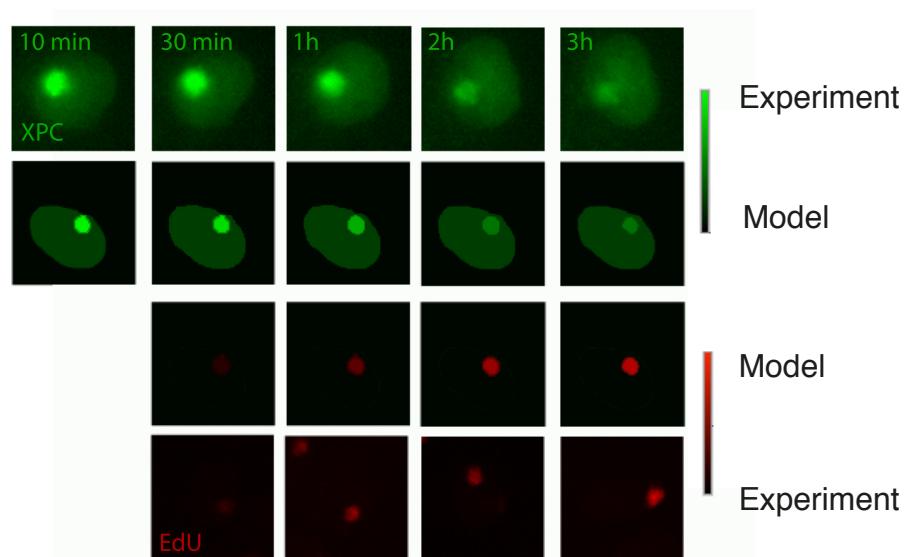
B



C



**Figure 2.9.: Quantitative NER model fits to accumulation and dissociation time courses.** A and B) Simulation (lines) and measurement (dots) of the accumulation and dissociation kinetics for the repair factors XPC, TFIID, XPG, XPA, XPF, RPA and PCNA. Estimated errors are depicted as shaded area. C) Fitted FLIP time courses of XPA, RPA and PCNA in the absence or presence of AraC and HU on locally irradiated cells.)



**Figure 2.10.: Comparison between measured and simulated single cell microscopy images** Simulated time courses (lines) and the associated microscopy images for GFP tagged XPC (two upper rows) and EdU incorporation (two lower rows). Simulated XPC expression intensities were normalized to the nuclear intensities of the microscopy images. For the EdU incorporation intensities were scaled according to the highest and lowest intensity values of the microscopy images.

the multi-dimensional model uncertainty inflicted by an individual parameter is projected to a one-dimensional profile likelihood (PL)

$$PL(\theta_i) = \max_{\forall j \neq i} [L(z^\dagger | \theta_j)]. \quad (2.12)$$

Accordingly, one parameter  $\theta_l$  with  $l \in 1 \dots N$  is gradually fixed along this dimension for different values of  $p$ . In each step  $X_{\theta_l}^2(p)$  is minimized fitting all other parameters  $\theta_k$  with  $k \in 1 \dots N; k \neq l$ . Subsequently, the identifiability of a parameter  $\theta_l$  can be determined by  $\Delta X_{\theta_l}^2(p)$ , which describes the difference between the global  $X^2$  minimum and the parameter dependent local minimum  $X_{\theta_l}^2(p)$ :

$$\begin{aligned} \Delta X_{\theta_l}^2(p) &= \min_{\{\theta_k | k=1, \dots, N; k \neq l\}} (X^2(\theta_1, \dots, \theta_{l-1}, p, \theta_{l+1}, \dots, \theta_N)) \\ &- \min_{\{\theta_k | k=1, \dots, N\}} (X^2(\theta_1, \dots, \theta_N)) \end{aligned} \quad (2.13)$$

Confidence bounds for the particular parameter depend on the threshold  $Q_{X^2}(1 - a, df)$ , which represent the  $(1 - a)$  quantile of the  $X^2$ -distribution with  $df$  degrees of freedom. The associated point-wise confidence intervals are defined as

$$CI_{1-a}(\theta_l) = \{p | \Delta X_{\theta_l}^2(p) \leq Q_{X^2}(1 - a, 1)\}. \quad (2.14)$$

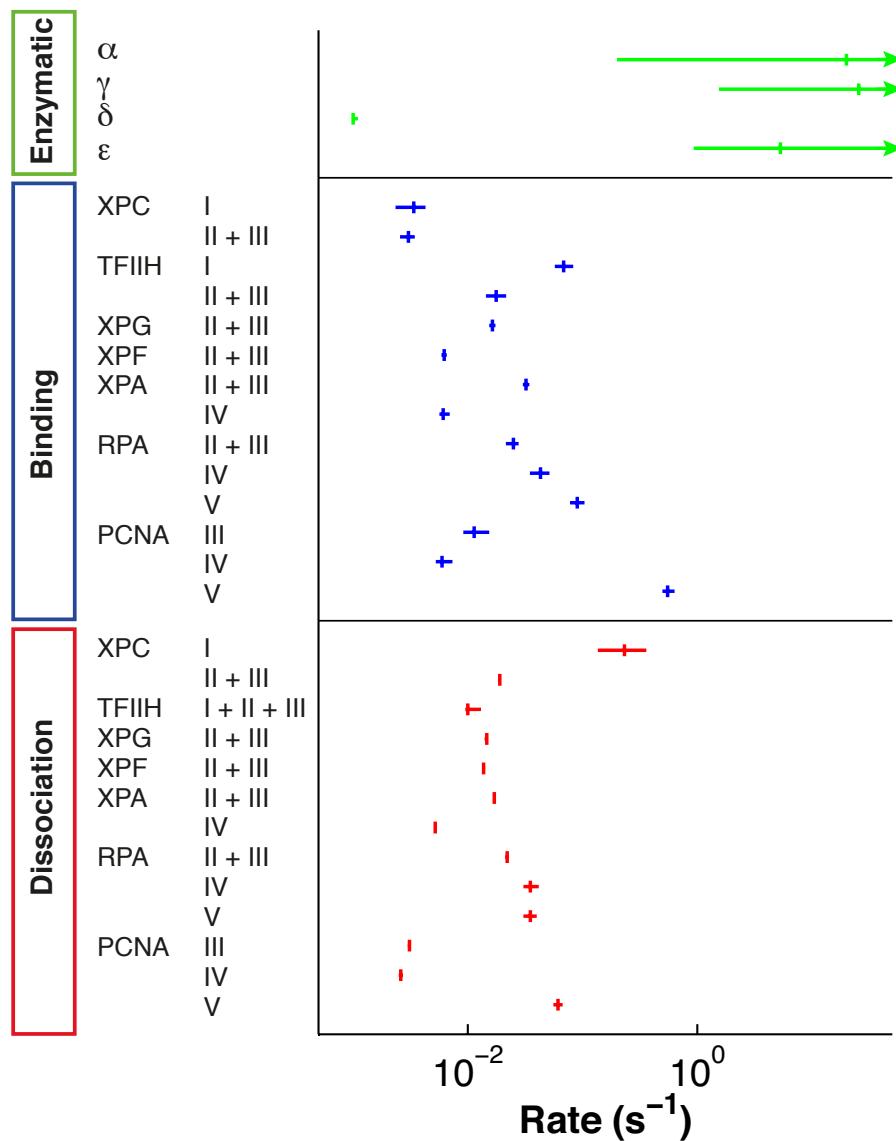
For one fixed parameter at a time and thus one degree of freedom we can derive the common confidence region  $CI_{95\%}$ , which corresponds to a  $X^2$ -distribution quantile of  $Q_{X^2}(95\%, 1) = 3.8$ . A parameter  $\theta_l$  is identifiable, if the confidence interval  $CI_{1-a}(\theta_l)$  is finite, which can be determined directly from the graph of the profile likelihood  $\Delta X_{\theta_l}^2(p)$  for different values of  $p$  (cf. Figure A.1).

We applied the identifiability analysis on our original NER model comprising 40 binding and dissociation parameters and 5 catalytic reaction constants. As it turns out, all binding and dissociation constants were identifiable with small bounds only under the assumption that the repair factor exchange at unwound and incised DNA is equal (cf. Figure 2.11). Besides the slow rate of rechromatinisation  $\delta$ , presumably identifiable due to the slow decrease in accumulated XPA (cf. Figure 2.9), all catalytic rates are fast. This is seen by the existence of lower bounds on the rate constants of the order of  $1 \text{ s}^{-1}$ . For the numerical values of

## *2. Mathematical Modelling of Nucleotide Excision Repair (NER)*

---

the parameters see Table A1 and Table A2. The  $K_d$  values, depicting the protein binding affinities to the respective DNA repair intermediate fall into a physiological realistic range between  $\sim 100$  nM and  $\sim 1$   $\mu$ M (cf. Table 2.3). Only XPC has a particular low affinity of 9  $\mu$ M which is consistent with previous findings reporting that the time until DNA incision is mainly determined by the slow lesion recognition [6].

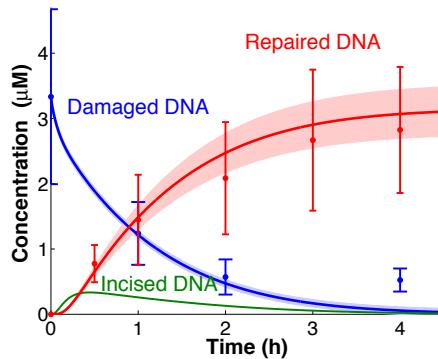


**Figure 2.11.: Parametrisation identifies realistic model for NER.** Mean values (vertical bars) and confidence intervals (horizontal bars) of catalytic (green), binding (blue) and dissociation constants (red) characterizing the dynamic assembly of NER factors at the successive repair intermediates (damaged DNA I, unwound DNA II, incised DNA III, resynthesised DNA IV, rechromatinised DNA V). Arrow heads indicate infinite confidence intervals.

Value	XPC	TFIIH	XPG	XPF	XPA	RPA	PCNA
<b>Damaged DNA</b>							
$K_d(\mu M)$	9.35 (3.46;16.09)	0.052 (0.044;0.072)	NA	NA	NA	NA	NA
<b>Unwound DNA</b>							
$K_d(\mu M)$	0.864 (0.635;1.006)	0.204 (0.163;0.259)	0.395 (0.373;0.419)	2.446 (2.344; 2.596)	0.147 (0.138;0.158)	1.222 (1.048;1.36)	NA
<b>Incised DNA</b>							
$K_d(\mu M)$	0.864 (0.635;1.006)	0.204 (0.163;0.259)	0.395 (0.373;0.419)	2.446 (2.344; 2.596)	0.147 (0.138;0.158)	1.222 (1.048;1.36)	0.388 (0.319;0.538)
<b>Resynthesized DNA</b>							
$K_d(\mu M)$	NA	NA	NA	NA	0.236 (0.222;0.27)	1.167 (0.924;1.521)	0.605 (0.531;0.747)
<b>Rechromatinized DNA</b>							
$K_d(\mu M)$	NA	NA	NA	NA	0.538 (0.438;0.645)	0.154 (0.134;0.182)	

**Table 2.3.:  $K_d$  Values.** NA, not applicable.  $K_d$  ( $k_{off}/k_{on}$ ) values are given for every repair protein and arranged in columns. Reference parameter set and 95% confidence intervals (in parentheses) are shown.

The narrow confidence bounds on the parameters allow to make valid computational predictions. For example, we can simulate the so far not observable fraction of incised DNA (cf. Figure 2.12, green trajectory). As the EdU incorporation measurement shows, the damaged (blue) and repaired DNA (red) kinetics are tightly coupled. This omits a higher accumulation of incised DNA. **hier koennte noch was kommen...**



**Figure 2.12.: Short delay between DNA damage removal and DNA repair synthesis omits accumulation of incised DNA.** Experimental (dots with error bars) and simulated (lines) time courses for damaged DNA (blue) and DNA repair synthesis (red). Simulated trajectories depict repair intermediate I+II for damaged DNA and repair intermediate IV+V for repaired DNA (cd. Figure 2.7). Estimated errors are depicted as shaded area. Model prediction for incised DNA (green) constitutes from DNA repair intermediate III.

### **3. Control analysis of the DNA repair rate**

In the previous chapter we introduced the incorporation of EdU upon UV-irradiation as a quantitative readout for the DNA repair synthesis kinetic (cf. Figure 2.4). We could show that despite the pathway's complexity DNA repair follows a mono-exponential kinetic of first order. Moreover, the repair of single lesions is distributed over a broad time span with 50% of the lesions being repaired after a half-time of  $\sim 1.2$  h. Notably, measurements on the single cell level revealed that also the half-time and hence the rate of repair is highly variable (cf. Figure 2.5B-E). So far the origin of this variability is unknown but also if and how the cell faces this highly fluctuating environment.

In the following chapter we will apply the previously introduced kinetic NER model and explore the nature of the repair rate variability. Simulating the effect of NER factor variation on the repair rate we find that the repair rate control is distributed among all NER factors. Exploiting the natural variability of endogenously expressed repair proteins we can experimentally corroborate the computationally derived finding for the NER factors XPC, TFIID, XPA, XPF and RPA. Apart from the variability in protein expression the model identifies the initial amount of inflicted DNA damage as major contribution determining the repair rate distribution. Both sources combined appear to be sufficient to explain the overall rate variability.

The work reported in this chapter has been done in close collaboration with Paul Verbruggen who planned and conducted the experiments.

#### **3.1. Kinetic NER model predicts collective rate control**

##### **3.1.1. Response coefficients**

The narrow confidence bounds derived by PLE analysis identified parameters in reasonable biological ranges (cf. Section 2.3.3) and allowed us to use the model for quantitative predictions. On the basis of the simplified model result, indicating that the fast and random enzyme exchange defines the slow first-order rate kinetics (cf. Section 2.2.1), we wanted to

test whether the concept of multi-protein rate control also applies on the realistic NER model. To determine the response of a system to changes in the environment one can calculate the response coefficients  $\tilde{R}_i$  [43, 44]. Accordingly, we can quantify the relative change in the repair rate  $\nu$  as a function of the relative change in the protein concentration  $C_i$  ( $i = \text{XPC}, \text{TFIIF}, \dots$ ).

$$\tilde{R}_i = \frac{\partial \ln \nu}{\partial \ln C_i}. \quad (3.1)$$

The inverse of  $\nu$  is referred to as characteristic time  $\tau$  so that Eqn. 3.1 can be rewritten as

$$\tilde{R}_i = \frac{C_i}{\tau^{-1}} \frac{\partial \tau^{-1}}{\partial C_i}. \quad (3.2)$$

$\tau$ , in turn, can be directly approximated from the distribution of repaired DNA states  $y_\pi^R(t)$ , which include the modelled DNA intermediates for resynthesised and rechromatinised DNA. By taking the ratio between the first ( $\mu^{(1)}$ ) and the zeroth ( $\mu^{(0)}$ ) central moment of the distribution we derive the reaction-specific mean time

$$\tau_R = \frac{\mu^{(1)}}{\mu^{(0)}}, \quad (3.3)$$

with

$$\mu^{(m)} = \int t^m y_\pi^R(t) dt. \quad (3.4)$$

To ensure the convergence of the integral we subtract all repair synthesis states (IV + V) from the initial amount of damages  $y_{00}^I = 3.33 \mu\text{M}$  at  $t = 0$ .

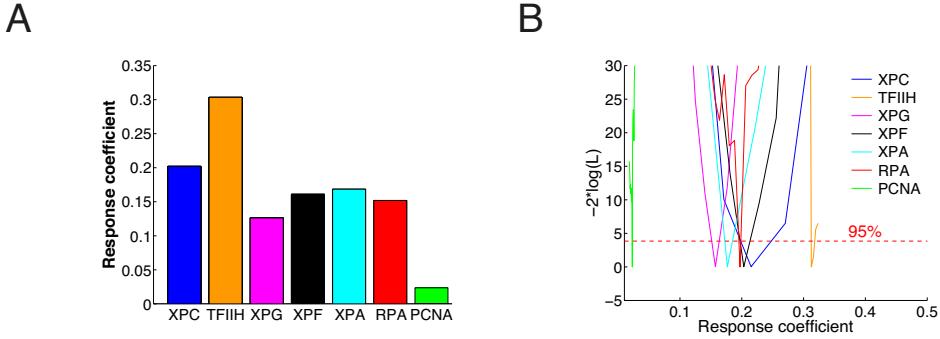
$$\tau_{\text{syn}} = \frac{\int (y_{00}^I(0) - (\sum_\pi y_\pi^{IV}(t) + \sum_\pi y_\pi^V(t)) dt) \cdot t dt}{\int y_{00}^I(0) - (\sum_\pi y_\pi^{IV}(t) + \sum_\pi y_\pi^V(t)) dt} \quad (3.5)$$

Using  $\tau_{\text{syn}}$  in Equation 3.2 we derive the response coefficients for the repair synthesis rate, which are uniformly small  $\sim 0.3$  and below (cf. Figure 3.1). This result implies that there is no single repair protein whose effect on the repair speed could be interpreted as rate-limiting. A similar result is obtained for the rate of re-synthesis response coefficients (cf. Appendix **tbd**), with the characteristic time

### 3. Control analysis of the DNA repair rate

---

$$\tau_{\text{inc}} = \frac{\int_0^\infty t \sum_{x=I}^{\text{III}} \sum_\pi y_\pi^x(t) dt}{\int_0^\infty \sum_{x=I}^{\text{III}} \sum_\pi y_\pi^x(t) dt}. \quad (3.6)$$



**Figure 3.1.: Collective control of the repair rate.** A) Response coefficients for seven repair factors (XPC, TFIIDH, XPG, XPA, XPF, RPA, PCNA) are small and uniformly distributed. B) Prediction profile likelihood indicate small prediction confidence bounds for each response coefficient. 95% threshold is given by the  $\chi^2$  distribution with one degree of freedom.

**Prediction profile likelihood** To obtain realistic confidence bounds for the response coefficients we performed a prediction profile likelihood estimation [45, 46]. Thereby the predicted response coefficients  $\tilde{R} = F(D_{\text{pred}}, \theta)$  are considered as model outcome  $F$  for a predicted experimental design  $D_{\text{pred}} = (z_{\text{pred}}, t_{\text{pred}}, u_{\text{pred}})$ .  $D_{\text{pred}}$  specifies a prediction observable  $z_{\text{pred}}$  at time point  $t_{\text{pred}}$  given the externally controlled stimulation  $u_{\text{pred}}$ . As defined in Eqn. 2.8  $z_{\text{pred}}(y(t), \theta)$  comprises a model simulation that can be mapped to experimentally observable quantities. Analogous to Eqn. 2.12 the prediction profile likelihood

$$PPL(\tilde{R}) = \max_{\theta^* \in \{\theta | F(D_{\text{pred}}, \theta) = \tilde{R}\}} L(z^\dagger | \theta^*, \tilde{r}) \quad (3.7)$$

is obtained by maximization over the model parameters satisfying the constraint that the model response  $F(D, \theta^*)$  after fitting is equal to the considered value  $\tilde{r}$  for the prediction  $\tilde{R}$  with respect to the measured data  $z^\dagger$ . This procedure is repeated for continuous variations of  $\tilde{R}$ . The model response can then be expressed as

$$\Delta X_{\theta_l}^2(\tilde{r}) = \min_{\{\theta, \tilde{r} \in \tilde{R}\}} (X^2(\theta, \tilde{r})) - \min_{\{\theta\}} (X^2(\theta)), \quad (3.8)$$

which describes the difference between the global  $X^2$  minimum and the best fit with  $r$  included into the objective function. Similar to Eqn. 2.14 we can determine prediction confidence bounds

$$PCI_{1-a}(D_{\text{pred}}, z^\dagger) = \{\tilde{r} | \Delta X_{\theta_l}^2(\tilde{r}) \leq Q_{X^2}(1 - a, 1)\}, \quad (3.9)$$

which include the set of predictions  $\tilde{R} = F(D_{\text{pred}}, \theta)$  for which  $-2\log(\text{PPL})$  is below the threshold given by the  $X^2$ -distribution. The PPLs were computed within the d2d-framework [34] using the CVODE package [36] for the numerical integration of the ODEs together with the sensitivity equations (cf. Section 2.3.1). By applying the PPL analysis for each response coefficient we find that all of them have small confidence bounds indicating how well the model predictions are determined by the data (cf. Figure 3.1B).

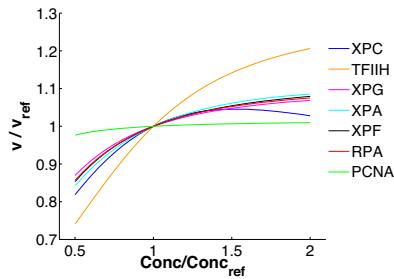
The moderate response predicted by the response coefficients also holds true for larger variations in the repair factor concentration (cf. Figure 3.2). The linear approximation (on which the response coefficients are based) yields a reasonable description for about two-fold concentration decreases or increases (corresponding to a knock-down or overexpression experiment), while for very large decreases the repair rate drops eventually to zero (corresponding essentially to a gene knock-out). From this result we can conclude that the kinetic NER model predicts a collective NER factor control of the repair rate. Consequently, the repair pathway appears robust against natural fluctuations in repair protein expression.

### 3.1.2. Exploiting natural variability in protein expression to quantify rate control

To corroborate the model prediction of a distributed repair rate control, we developed an experimental set-up for the investigation of single repair factors and their quantitative influence on the repair rate. In particular, we asked whether there is a measurable response in the repair rate due to the natural occurring variability in protein expression. By capturing the integrated nuclear fluorescence intensity of indirectly antibody-stained repair proteins we derive the expression values for XPC, TFIIH, XPA, XPF and RPA. (cf. Figure 2.2B). For

### 3. Control analysis of the DNA repair rate

---



**Figure 3.2.: Moderate repair rate response against natural NER factor expression variability.** Repair rate as a function of concentration changes in individual repair factors.

these five individually measured NER factors the expression variability was quantified by calculating the coefficient of variation (CV; standard deviation divided by the mean), which is on average  $\sim 0.37$  (cf. Figure 3.3A-E). Table 3.1 depicts mean and standard deviation representing at least three biologically independent measurements of the expressed protein cell-two-cell variability.

XPC	TFIID	XPA	XPF	RPA
CV: $0.34 \pm 0.05$	$0.33 \pm 0.02$	$0.33 \pm 0.03$	$0.4 \pm 0.04$	$0.44$

**Table 3.1.: Mean and standard deviation of the variability in nuclear XPC, TFIID, XPA and XPF expression.** Distributions for nuclear protein expression were acquired in  $n_b(XPC) = 5$ ,  $n_b(TFIID, XPA, XPF) = 3$  and  $n_b(RPA) = 2$  biological independent replicates. Within each measurement between  $n=250$  and  $n=572$  with an average of  $n=477$  cells were analysed.

To distinguish whether the measured variability is due to differences in nuclear expression or rather superimposed by measurement noise we co-analysed XPC-eGFP stably expressed in XP-C fibroblasts together with immunofluorescently labelled XPC in the same cell. Both quantities are strongly positively correlated (cf. Figure 3.3F) suggesting a large natural variability compared to much lower measurement noise. To quantitatively validate this observation we performed a principal component analysis [47]. Thereby both quantities, XPC-eGFP and antibody-stained XPC intensities, are orthogonally transformed into a new coordinate system where the new transformed variables are linearly uncorrelated. These new variables are referred to as principal components. In a two-dimensional case the variances

of both principal components define an error-ellipse as illustrated in Figure 3.3F. From this we estimated a relative measurement error of antibody labelling of  $\sim 11\%$  showing that the technique is suitable for quantification of nuclear NER factor concentrations.

Before exploring the direct relation between protein expression variability and the speed of repair we tested whether higher protein amounts in the nucleus correlate with the accumulation of NER factors in the locally damaged area. In correspondence to previous findings for XPG [6] there is a significant positive correlation between the nuclear XPC-eGFP expression and its local accumulation at the DNA lesions (cf. Figure 3.4A) thirty minutes after UV irradiation. Consequently, we conclude that the DNA lesions are not saturated so that a higher NER factor concentration could potentially accelerate the repair rate.

In fact, the nuclear protein concentrations for all five antibody-stained repair factors and the amount of incorporated EdU after one hour were significantly correlated (cf. Figure 3.4B-F). Remarkably, the dependency between both quantities, characterized by the slope of the regression line is very evenly distributed and also in the same order of magnitude as predicted computationally by the response coefficients (cf. Figure 3.1A). This agrees with the *in silico* finding that the kinetic control of the measured repair factors is uniformly distributed and that the rate of repair synthesis is robust against natural variations in the repair protein expression.

**Correlation analysis** The correlation analysis was performed in MATLAB (2012a, The Mathworks Inc., Natick, MA) using the inbuilt 'corrcoef' function, which also returns p-values indicating the significance of the correlation coefficient. The p-values are determined by a t-test, where the null-hypothesis states that the true correlation between two variables equals zero and that for its observed value  $r$  the quantity

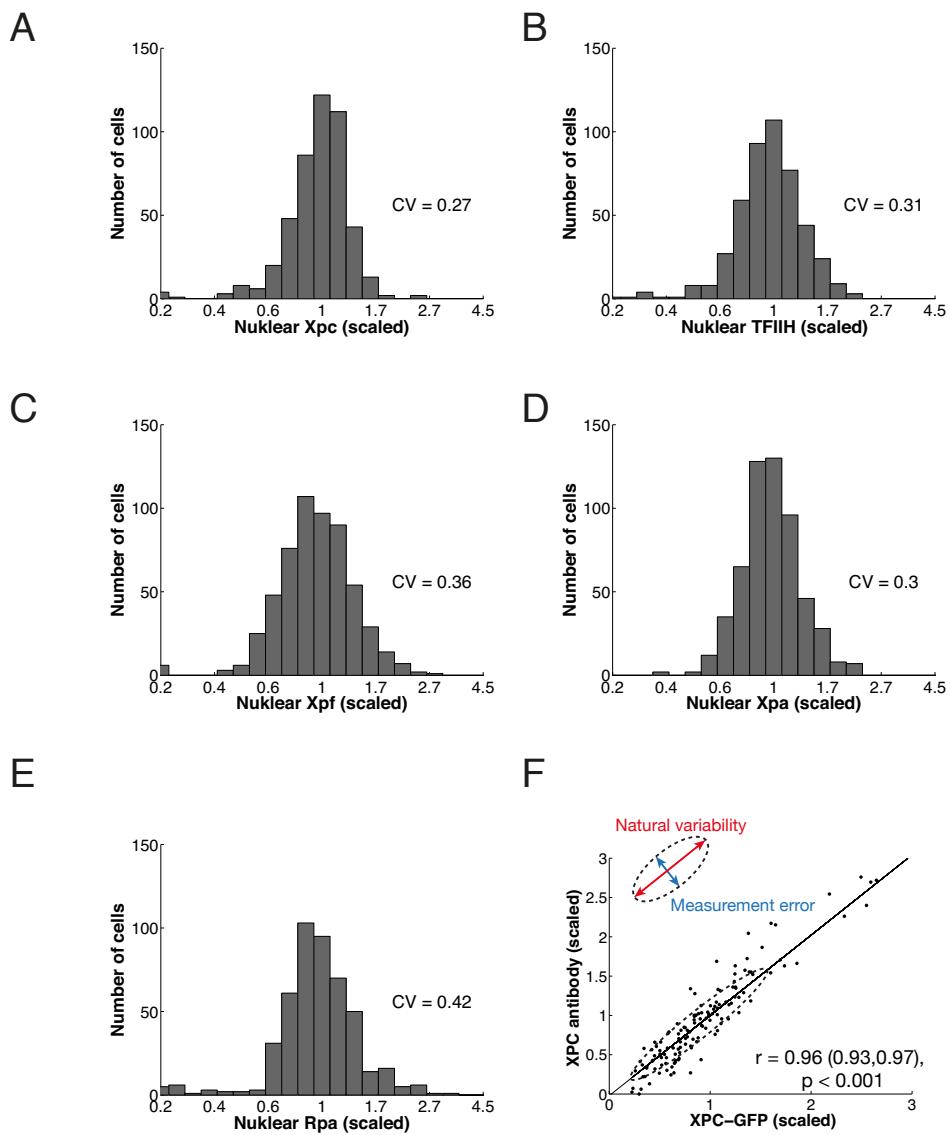
$$t = \frac{r}{\sqrt{\frac{(1-r^2)}{N-2}}} \quad (3.10)$$

follows approximately a  $t$ -distribution for  $N - 2$  degrees of freedom and a sample size  $N$  [48, 49]. A low p-value suggests the rejection of the null-hypothesis and thus, demonstrates the significance of the correlation.

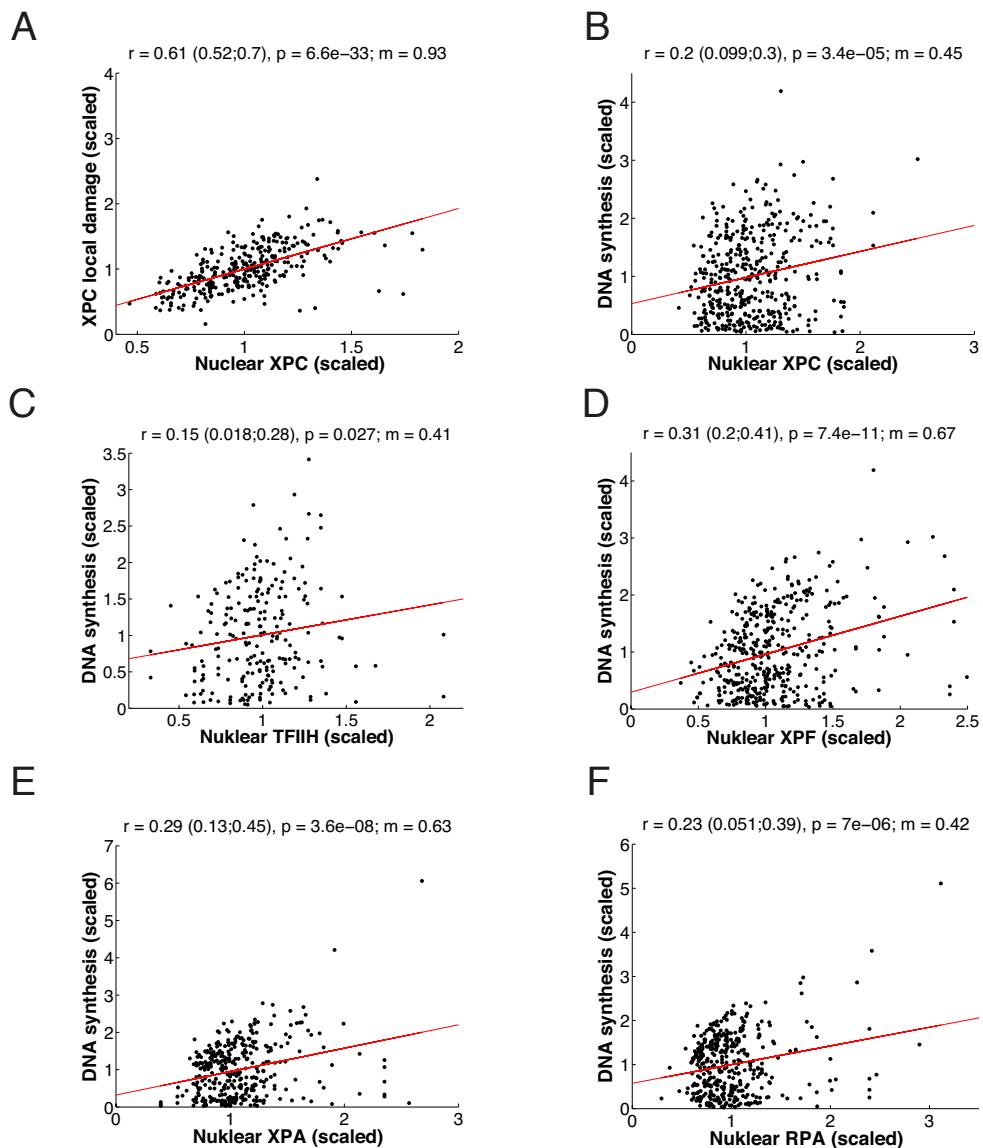
The confidence intervals for each correlation coefficient were estimated by non-parametric

### 3. Control analysis of the DNA repair rate

---



**Figure 3.3.: Natural variability in nuclear NER factor expression is significantly larger than measurement error** A-E) Histograms of nuclear protein concentrations of the antibody stained NER factors XPC, TFIIH, XPF, XPA and RPA. F) Scatter plot of antibody stained XPC against XPC-eGFP stably expressed in XP-C XPC-eGFP cells. The dashed error-ellipse illustrates the proportion of natural variability and measurement noise. **denote cell numbers**



**Figure 3.4.: Natural expression variability of the repair factors XPC, TFIID, XPF, XPA and RPA has small effect on DNA repair synthesis.** A) Scatter plot of accumulated XPC in local damage against the nuclear XPC concentration ( $n=303$ ) 60 minutes post irradiation. B-F) DNA synthesis correlated with the endogenous concentration of XPC, TFIID, XPF, XPA and RPA 60 minutes post-irradiation. A-F) Red lines represent linear regression with correlation coefficient  $r$ ,  $p$ -value and slope  $m$ . 95% confidence bounds of all correlation coefficients  $r$  were estimated by non-parametric bootstrap and are given in brackets. **denote correct cell numbers**

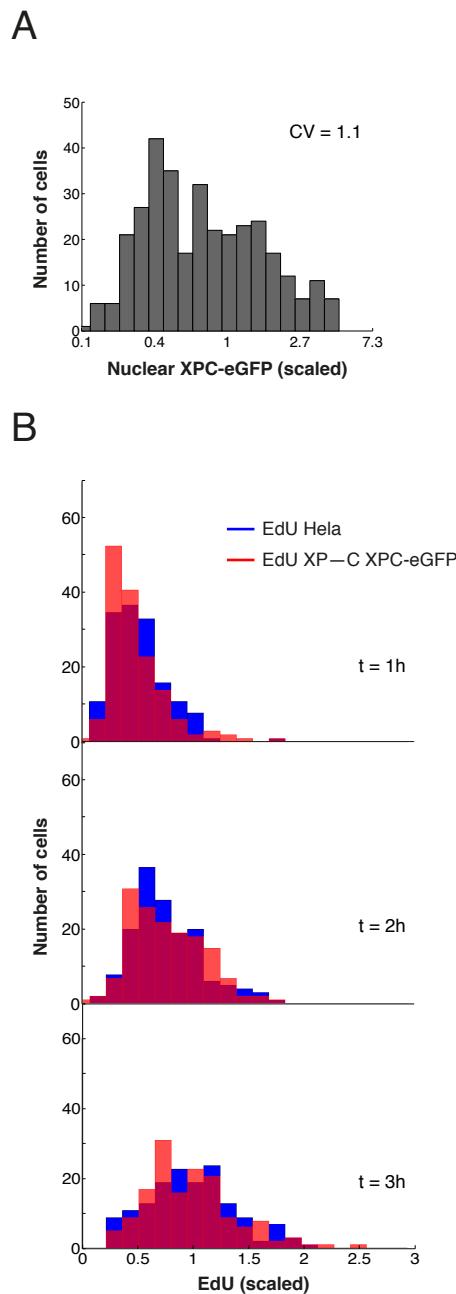
### *3. Control analysis of the DNA repair rate*

---

bootstrap [50]. Thereby the correlation coefficient was re-sampled for 10,000 times each time drawing n pseudoreplicates (with replacement), where n denotes the number of measured cells from the original data. The confidence borders mark the interval including 95% of the re-sampled correlation coefficients.

To reinforce the conclusion of a robustness DNA repair rate we compared the repair factor expression and its consequence on the repair variability in different cell lines. In contrast to indirectly labelled XPC in Hela cells [10] or in human primary fibroblasts (cf. Figure 3.3) the expression of stably transfected XPC-eGFP into XP-C patient cells is significantly broader distributed. Its CV has a value of  $\sim 1$ , which is up to four times larger compared to XPC variability in Hela or fibroblasts cells (cf. Figure 3.5A, Figure 3.3A and [10]). At the same time the distributions of accumulated EdU after 1, 2 and 3 hours at the sites of local damage is strikingly similar in both cell types (cf. Figure 3.5). The unaffected DNA repair dynamic, even for much larger NER factor fluctuations, suggests a general phenomenon of a robust repair in different cell types.

Returning to the comparison between the measured EdU incorporation as a function of NER factor concentration and the predicted response coefficients it should be noted that the former are consistently two to four-fold larger (cf. Figure 3.4B-F and Figure 3.1A). This result suggests that the measured repair-rate response is not entirely explained solely by the model-predicted response coefficients. It remains to clarify, whether this discrepancy is due to a so far unknown additional NER component intrinsically contributing to the pathway control. Or, is there an external mechanism coordinating the NER factor expression and thereby superimposing the mathematically predicted response coefficients?



**Figure 3.5.: Variability of repair synthesis is comparable between XPC-eGFP complemented XP-C patient cell lines and native NER.** A) Distribution of nuclear XPC-eGFP concentrations stably expressed in XP-C cells ( $n=332$ ). B) Comparison of incorporated EdU in local damages in HeLa cells (blue) and XPC-eGFP complemented XP-C cells after 1, 2 and 3 hours ( $n=154$  cells per time point per cell line).

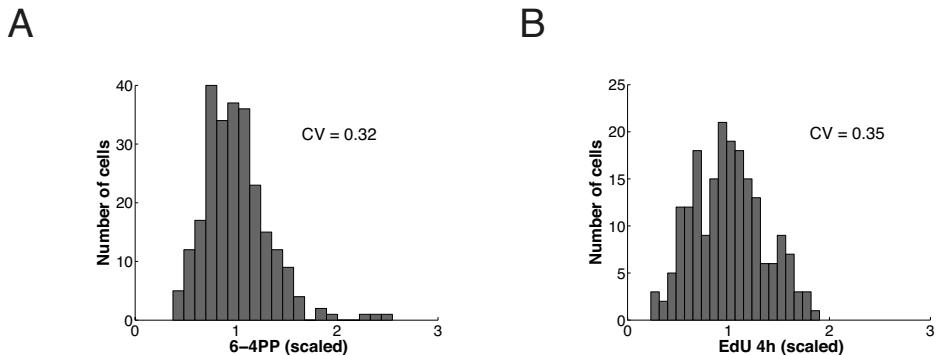
### 3. Control analysis of the DNA repair rate

---

## 3.2. Variability analysis of the repair rate

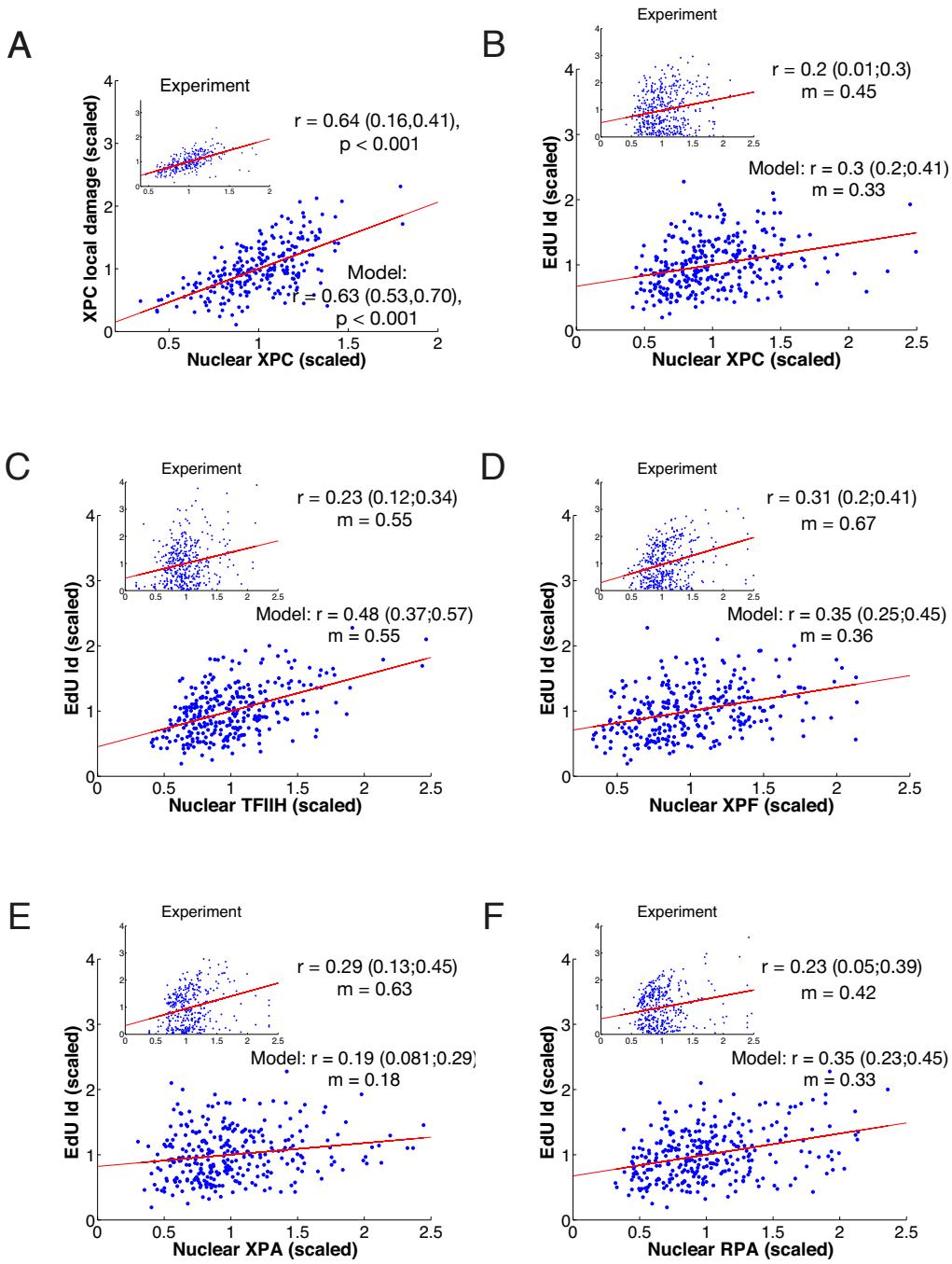
### 3.2.1. Variable NER factor expression and inflicted lesions account for the distribution of repair rates

While searching for factors affecting the repair rate control we noticed the broad scatter of incorporated EdU, which appears to increase for larger protein concentrations (cf. Figure 3.4B-F). This variability indicates that besides the repair proteins there might be further sources of cell-to-cell heterogeneity, which might be also involved in the regulation of the repair rate. To this end, we measured the amount of UV-inflicted DNA lesions by indirect (immuno)cytochemistry wondering how the distribution of pore sizes would propagate (cf. Figure 2.1B). The distributions of DNA damage captured immediately after UV-irradiation very much resembles the variable EdU incorporation after 4 h, which is also confirmed by the similar CV. These data suggest that the amount of DNA lesions contributes to the cell-to cell variability in the repair-rate.



**Figure 3.6.: Distributed amount of inflicted DNA damages contributes to repair rate variability.** A) XP-C XPC-eGFP cells were locally UV-irradiated with an intensity of 100 J/m<sup>2</sup>. DNA damages were quantified by indirect immunofluorescence microscopy in n=250 cells from five experiments. B) XP-C XPC-eGFP cells were locally irradiated and cultivated for 4 hours in the presence of EdU before fixation. Fluorescence signals were measured in n=198 cells derived from three independent experiments.

To find out whether the heterogeneity in nuclear NER factor expression together with the distributed amount of initial DNA damages would be sufficient to explain the variability in the repair rate we tried to reproduce the EdU response measurements with the model. Thus,



**Figure 3.7.: Quantitative comparison between measured and simulated repair synthesis in response to variable NER factor concentrations.** A) Simulated scatter plot of the of locally accumulated XPC against the nuclear XPC concentration. For comparison the experimentally measured scatter plot is given in reduced size. B-F) Comparison between measured (scatter plot in reduced size) and simulated correlation of DNA synthesis and nuclear XPC concentration. A-F) Red lines represent linear regression with correlation coefficient  $r$ , p-value and slope  $m$ . 95% confidence bounds of all correlation coefficients  $r$  were estimated by non-parametric bootstrap and are given in brackets. **EdU into DNA synthesis**

### *3. Control analysis of the DNA repair rate*

---

we simulated the model for several hundred times in correspondence with the number of measured cells. For each simulated cell nuclear NER factor concentrations and the amount of local damages were randomly drawn from log-normal distributions in accordance with the experimentally determined CV (cf. Figure 3.3 and Figure 3.6) and the previously measured mean values [6]. For PCNA and XPG we took the averaged  $CV \sim 0.33$  derived from the measured distributions. The simulated scatter plots for the accumulation of the nuclear repair factor at sites of local damage agree remarkably well with the measured data (cf. Figure 3.7A-E, compare enlarged correlation (simulation) with the smaller inlay (experiment)). Qualitatively, the same holds true for the comparison between simulated and measured EdU incorporation as a readout for DNA resynthesis in response to changes in the repair factor concentration (cf. Figure 3.7F-J). It should be noted that the simulated slopes representing the overall repair rate response are slightly elevated compared to the calculated response coefficients (cf. Figure 3.1), clearly marking the influence of the introduced distribution of DNA damages. However, the simulated slopes are still significantly smaller than the measured estimates (cf. Figure 3.7F-J, compare enlarged correlation (simulation) with the smaller inlay (experiment)).

To locate the lacking control, we first asked whether the individual contributions of the concentration distributions of the repair factors to the distribution of the repair rate 'sum up' according to the response coefficients predicted by the model. Assuming that the repair rate  $\nu$  is linearly dependent on the initial amount of inflicted lesions  $L$  and, as estimated in Section 2.1.4, also on the nuclear repair-factor concentrations  $C_i$  ( $i=1,2,\dots$ ) we can apply the general law of error propagation

$$\sigma(\nu) = \sqrt{\sum_i \left( \frac{\partial \nu}{\partial C_i} \sigma(C_i) \right)^2 + \left( \frac{\partial \nu}{\partial L} \sigma(L) \right)^2}. \quad (3.11)$$

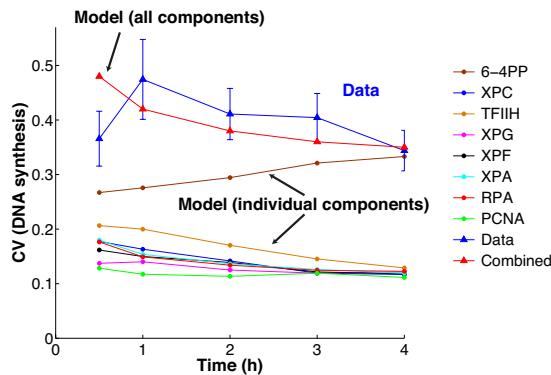
Introducing the predicted response coefficients  $R_i$  we can estimate the overall repair rate variability  $CV_\nu$  with

$$CV_\nu = \sqrt{CV_L^2 + \sum_i (R_i CV_i)^2}, \quad (3.12)$$

where  $CV_L$  and  $CV_i$  denote the coefficients of variation of the distributions of repair factors and initial amount of lesions. Obviously, the response coefficient for the initial amount of

inflicted lesions is 1. Including the variability for inflicted lesions  $CV_L \sim 0.32$  and the average  $CV_i \sim 0.33$  for each repair factor we can determine the repair rate variability with  $CV_\nu \sim 0.35$ . This is around 20% less than the measured  $CV_\nu \sim 0.45$  after one hour (cf. Figure 3.8).

To dissect the contribution of the individual factors to the overall variability, we computed the effect of heterogeneity in only one factor on repair synthesis. Accordingly, for each CV-trajectory in Figure 3.8 we determined the  $CV_\nu$  by simulating the repair rate for 500 cells and drawing only one factor randomly. The largest contribution can be allocated to the initial amount of damages which were responsible for about two thirds of the whole repair variability. Adding the heterogeneity of the nuclear expression values shrinks the gap between measured and simulated variability (cf. Figure 3.8 blue (data) and red (simulation) trajectory).



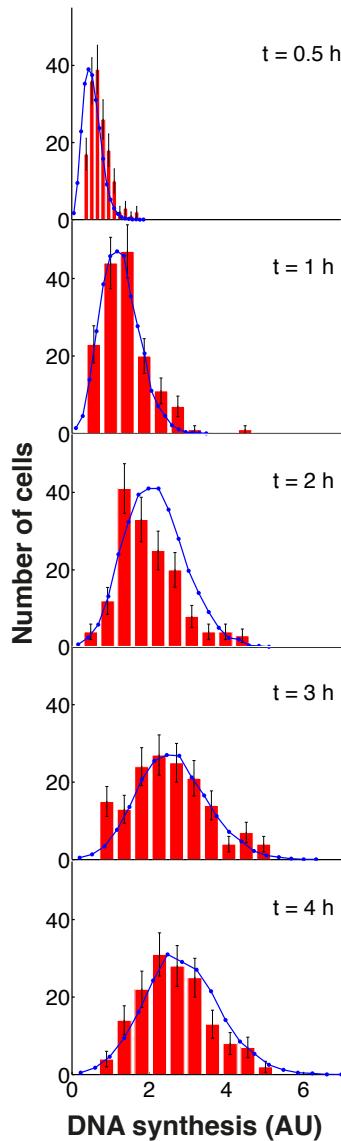
**Figure 3.8.: Variability in repair synthesis results from the aggregate of initial damage variability and concentration fluctuations of each NER factor.** Each time course represents the variability in repair synthesis produced by a single component (dots) or by all components together (triangles) at a given time. For comparison the experimental measured distribution of incorporated EdU is shown in blue. Error bars depict the 95% confidence intervals determined with non-parametric bootstrap.

This result becomes more explicit by comparing the strikingly matching temporal evolution of the measured and predicted distributions of repair synthesis (cf. Figure 3.9).

To summarize, we uncovered two sources of cell-to-cell variability in the rate of NER: (i) the amount of inflicted DNA lesions and (ii) the expression of NER factors. Together, both parameters explain most of the measured variability. However, the significant difference between measured and predicted repair rate responses leave space for interpretation

### 3. Control analysis of the DNA repair rate

---



**Figure 3.9.: Simulated repair rate distribution fits experimentally derived EdU incorporation** Temporal evolution of the repair rate variation i) measured in single cells after initial UV-irradiation (red histograms) and ii) predicted from model simulations (blue lines).

---

*3. Control analysis of the DNA repair rate*

whether there are additional factors involved in repair but have not been accounted for explicitly in the model.

## 4. Co-expression of NER repair factors

The aforementioned model-aided analysis of the DNA repair process revealed a link between the emergent phenomenon of rapidly exchanging and transiently interacting NER components with the experimentally observed slow first-order kinetics of repair. An important functional consequence of this kinetic design is that the control of the repair rate is shared by all repair factors. This manifests in the mathematical prediction of small, uniformly distributed response coefficients, which quantify the relative change of the repair rate in answer to changes in the nuclear repair protein concentration. Exploiting the natural variability in NER factor expression we experimentally corroborated the moderate control of the repair components on the repair rate. We found that compared to the model-predicted response coefficients the measured EdU response was significantly elevated.

In the following chapter we will show that this discrepancy can be explained assuming that the measured response is a superposition of firstly, the response due to concentrational changes for one repair factor and secondly, an additional cross-correlation among the repair factors. To test this assumption, we experimentally investigate the potential co-expression of five repair factors (XPC, TFIIH, XPA, XPF and RPA). Surprisingly, we find that the nuclear expression of these pairwise measured repair factors is indeed strongly positively correlated, whereas there is no correlation with the repair-independent cell cycle marker Ki67. This result suggests an additional control mechanism orchestrating NER factor expression on the pre-transcriptional level.

The experiments reported in this chapter have been planed and conducted in close collaboration with Martin Teichert, who is part of the division of Vascular Oncology and Metastasis led by Prof. Helmut Augustin.

## 4.1. Nuclear expression of NER factors is strongly correlated

### 4.1.1. Cross-correlation affects repair response

The finding that the measured response in DNA synthesis is about twofold higher than the model-predicted response coefficients (cf. Figure 3.1 and Figure 3.4) indicates that there is an additional factor regulating the repair response. A so far not considered explanation is a potential interdependence of the NER factor expression, introduced by an external gene-regulatory mechanism. This would result in a measurable cross-correlation of the nuclear NER factor expression. Mathematically, such a factor would extend the relation between measured EdU response (cf. Figure 3.4) and the model-predicted response coefficients to

$$\frac{C_i}{\nu} \frac{d\nu}{dC_i} = \frac{C_i}{\nu} \frac{\partial \nu}{\partial C_i} + \sum_{j \neq i} \left( \frac{C_j}{\nu} \frac{\partial \nu}{\partial C_i} \right) \left( \frac{C_i}{C_j} \frac{\partial C_j}{\partial C_i} \right), \quad (4.1)$$

where  $\nu$  represents the rate of EdU incorporation and  $C_i$  the nuclear concentration of the  $i^{\text{th}}$  repair factor. From Eqn. 4.1 we directly measured the EdU incorporation for varying NER factor concentrations after one hour (cf. Figure 3.4)

$$A_i = \frac{C_i}{\nu} \frac{d\nu}{dC_i}.$$

If we were able to measure the pairwise cross-correlations

$$B_{ij} = \frac{C_i}{C_j} \frac{\partial C_j}{\partial C_i}$$

additionally, we can derive the following system of equations

$$R_i + \sum_{j \neq i} B_{ij} R_j = A_i, \quad (4.2)$$

for  $i \in 1, \dots, N$  and

$$R_i = \frac{C_i}{\nu} \frac{\partial \nu}{\partial C_i}. \quad (4.3)$$

#### 4. Co-expression of NER repair factors

---

Assuming that all  $B_{ii} = 1$  Eqn. 4.2 translates into the matrix notation

$$\vec{A} = B\vec{R}. \quad (4.4)$$

By inverting  $B$  we can solve Eqn. 4.4 for the desired response coefficients directly.

##### 4.1.2. Microscopy analysis of co-staining experiments

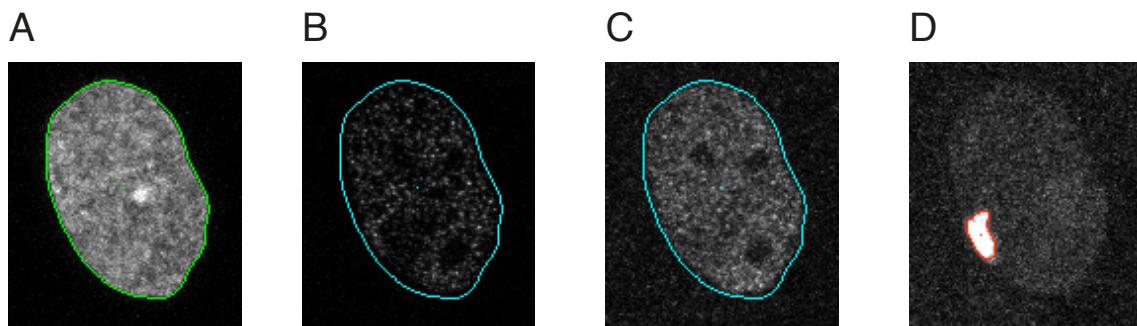
In order to measure the cross-correlation matrix  $B$  we made use of the fluorescence microscopy approach, which proved to result in accurate measurements of the nuclear repair factor expression and their UV-induced repair dynamics, before (cf. Figure 3.4). For a flexible combinatorial tagging of multiple antigens simultaneously we used an indirect antibody-labelling protocol and established five single cell double stainings (cf. Table 4.1) using the primary and secondary antibodies listed in Table **tbm**.

	XPC	TFIILH	XPA	XPF	RPA
XPC	X	X	X	X	X
TFIILH	-	-	X	-	
XPA		-	X	X	
XPF			-	-	
RPA				-	

**Table 4.1.: Immuno-fluorescence microscopy of NER factor co-expression** Crosses indicate pairwise measurement of indirectly antibody-labeled NER factor expression. Underlined crosses denote costainings involving directly labeled XPC.

Two additional cross-correlation measurements were possible by directly labelling XPC with a monoclonal mouse antibody and thereby avoiding potential cross-reactions between antibodies originating from the same host species [51, 52]. The cross-correlation analysis was performed in human diploid female fibroblasts which were grown to confluence on coverslips. Analogous to the description in Section 2.1.1 cells were UV-irradiated locally with a dose of 100 J/m<sup>2</sup>. Pre-incubated with serum-free medium containing 10 µM EdU cells were allowed to repair for 60 minutes in an incubator. After the subsequent direct or

indirect antibody labelling of two selected repair factors cells were also incubated with a DAPI solution, which visualizes the cell's chromatin and thereby gives rise to the contour of the nucleus. For each double staining microscopic 3-dimensional imaging was conducted on a Leica TCS SP5 II confocal microscope. All images were analysed following the protocol presented in Section 2.1.2. Segmentation of the nucleus was performed on the DAPI signal, due to the large signal-to-noise ratio in this channel (cf. Figure 4.1A). The segmented region was then projected to the other channels and used to quantify the signal emitted by the secondary antibodies at 488 and 647 nm (cf. Figure 4.1B and C). For the same reason segmentation of the locally UV-irradiated chromatin region was done with the area determined from the EdU signal and then projected onto the signal of accumulated protein (cf. Figure 4.1D).

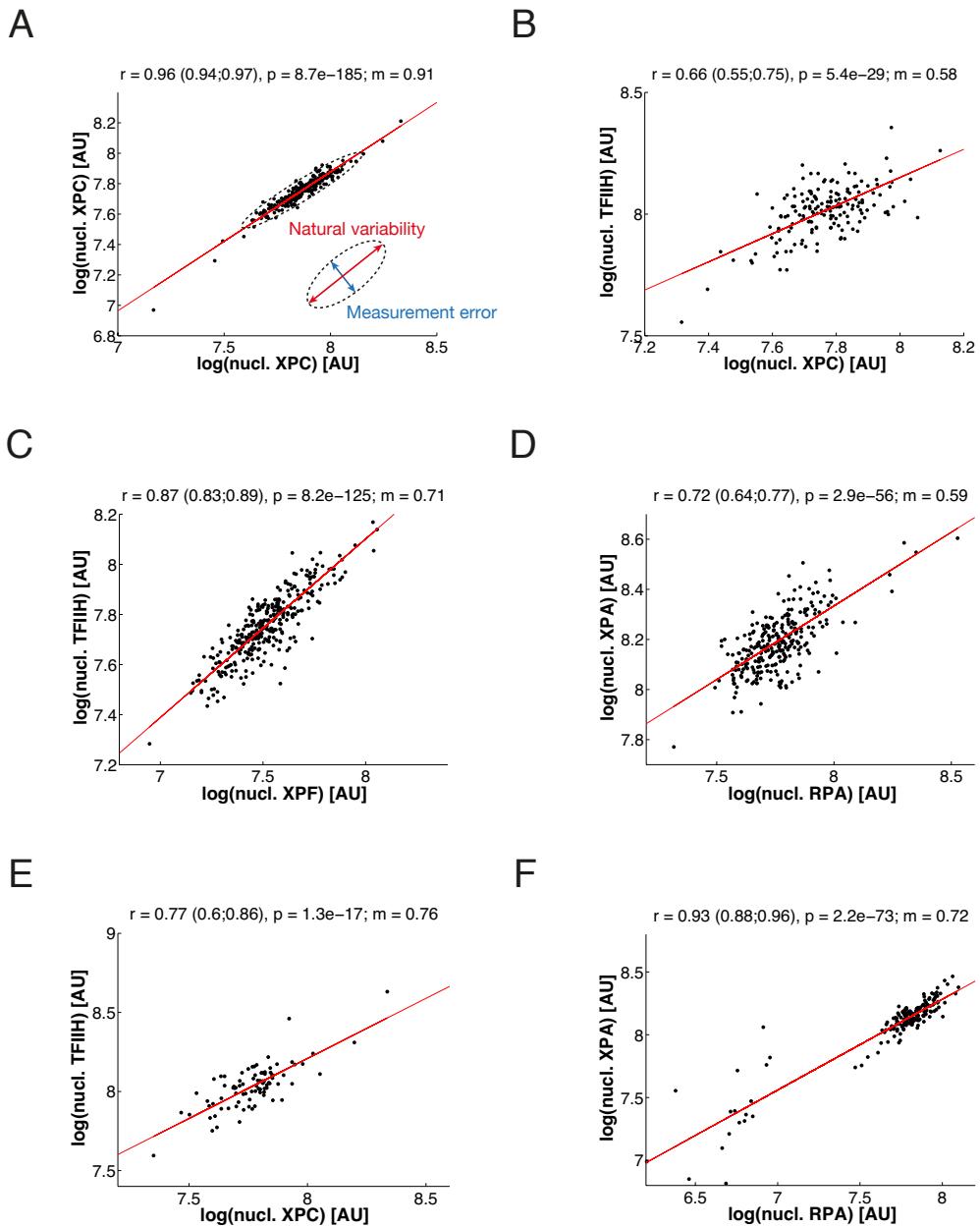


**Figure 4.1.: 3D fluorescence microscopy imaging of NER factor co-expression** A) DAPI-stained DNA content measured in human primary fibroblast (Life technologies). DAPI segmentation at 461 nm (green contour line) determines nuclear dimensions. B-C) Nuclear expression of indirectly immuno-stained repair factors at 488 nm (B) and at 647 nm (C). Signals are quantified according to the nuclear contour (cyan contour line) segmented from the DAPI signal. D) DNA locally damaged by UV-irradiation and subsequently incubated for 60 minutes in the presence of 10 µM EdU. Segmentation of the EdU signal is indicated by the red contour line.

To test whether the antibody co-staining experiment is suitable for the cross-correlation analysis we measured XPC expression with a directly labelled antibody together with an indirect immuno-staining and correlated both signals. Fitting an error ellipse to these data (cf. Figure 4.2A), as described in section 2.1.2, we estimated a relative measurement error of antibody labelling of 13%, showing that the technique is sufficiently accurate to exploit the natural variability in protein expression for the cross-correlation analysis.

As it turns out, all pairwise correlations of the measured co-staining experiments (cf.

#### 4. Co-expression of NER repair factors



**Figure 4.2.: NER factor cross-correlation is repair independent.** A) Scatter plot of indirectly antibody-stained XPC against a directly labelled antibody recognising XPC ( $n=336$ ) as determined by quantitative (immuno) fluorescence microscopy. B-D) Pairwise correlations of indirectly antibody-labelled XPB against XPC ( $n=220$ , B), XPB against XPF ( $n=410$ , C) and XPA against RPA ( $n=350$ , D) in locally damaged cells. Expression values represent fluorescence intensities originating from the nucleus including the damaged region (signal quantification was performed analogous to [6]). E-F) Scatter plots of the nuclear expression of XPB vs. XPC ( $n=85$ ) and XPA vs. RPA ( $n=170$ ) in undamaged cells. A-F) Red lines represent linear regression with correlation coefficient  $r$ , p-value and slope  $m$ . 95% confidence bounds of all correlation coefficients  $r$  were estimated by non-parametric bootstrap and are given in brackets.

Table 4.1) are strong positively correlated with correlation coefficients between 0.66 and 0.88 (cf. Figure 4.2B-D and Figure A.2). Notably, the result is irrespective of whether the acquired fluorescence signal is taken from the whole nucleus including the locally damaged area or only from the undamaged chromatin region (cf. Figure 4.2B-D and 4.2E-F). This suggests that the correlation of nuclear NER factors is independent of the ongoing repair. As a consequence, we suspect that the regulatory mechanism determining the protein concentrations lies on a preceding level such as transcription or translation.

#### 4.1.3. Flow cytometry

To further pursue this question we repeated the experiment by flow cytometry, this time, investigating the NER-factor expression in human brain pericytes. We established a double-staining protocol for XPA and RPA and observed that also in this cell type the two repair factors strongly correlate (cf. Figure 4.3A and B).

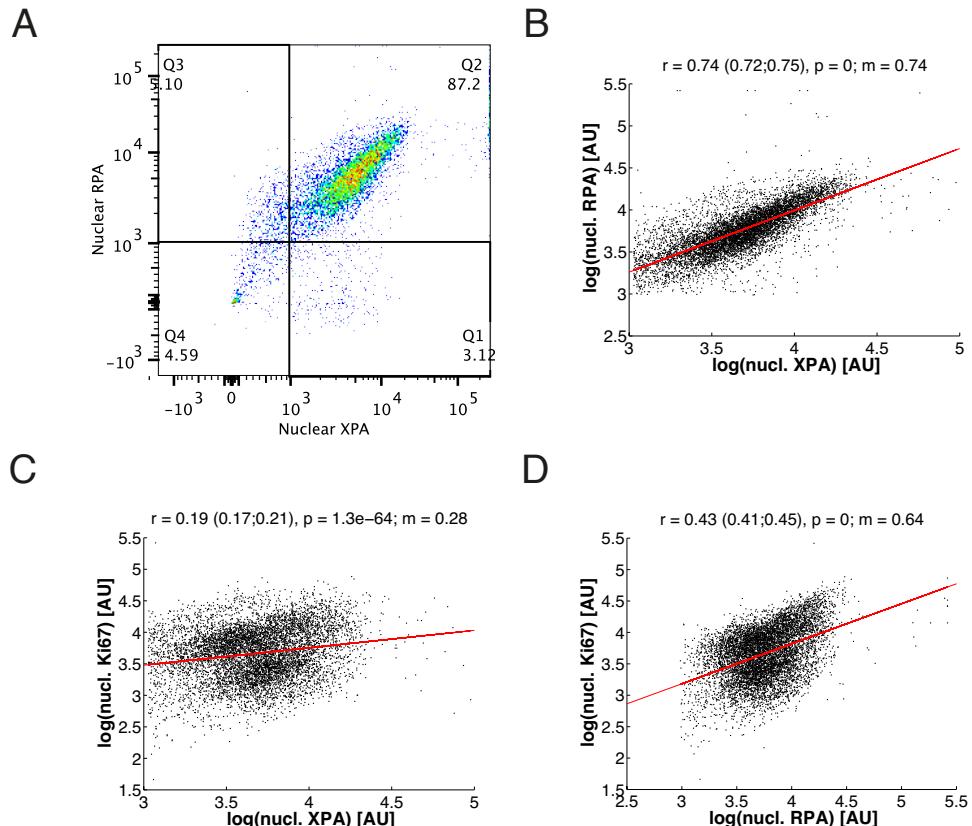
Quantitatively, the correlation coefficients of the double-staining signals in both cell types have the same order of magnitude. In particular, each value falls into the confidence interval of the other (cf. Figure 4.3B and Figure 4.2D). To test, whether this correlation is specific for proteins involved in DNA repair we measured the expression of the proliferation marker Ki67. Surprisingly, although the correlation between both repair factors and Ki67 is visibly reduced (XPA vs. Ki67 = 0.19 and RPA vs. Ki67 = 0.43 against XPA-RPA = 0.73) it is still significantly positive (cf. Figure 4.3C and D).

#### 4.1.4. Cell cycle independent cross-correlation of NER factor expression

As Ki67 increases during cell cycle we asked, to what extend the observed protein correlation of Ki67 and the NER factors is cell-cycle dependent? To answer this question, we stained the cell's DNA using FxCycle violet and gated them according to their DNA content (cf. Figure 4.4A). Two distinct peaks denote the portion of cells traversing the G1 or the synthesis and G2 phase, respectively. By sorting the protein expression values in accordance with their cell cycle phase we identified for each protein combination two contiguous regions revealing a general trend of increased protein expression during the S1

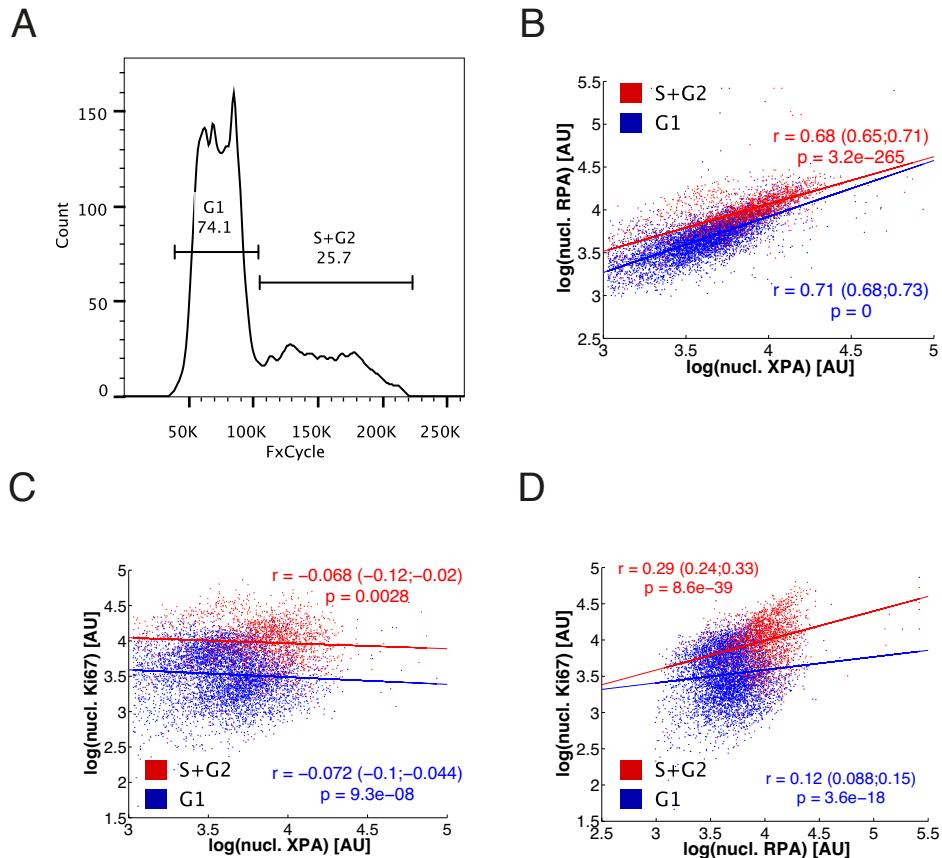
#### 4. Co-expression of NER repair factors

---



**Figure 4.3.: Correlated expression of RPA and XPA in human brain pericytes** A) Selection of XPA and RPA positive human brain pericytes (Q2) as determined by flow cytometry ( $n=8084$ ). B-D) Nuclear expression of indirectly antibody-labelled RPA against XPA (B), Ki67 against XPA (C) and Ki67 against RPA (D). B-D) Red lines represent linear regression with correlation coefficient  $r$ ,  $p$ -value and slope  $m$ . 95% confidence bounds of all correlation coefficients  $r$  were estimated by non-parametric bootstrap and are given in brackets.

and G2 phase in comparison to the G1 phase (cf. Figure 4.4B-D).



**Figure 4.4.: NER factor cross-correlation is robust against cell cycle progression.** A) Distribution of fluorescently labelled DNA proportional to the DNA content. Horizontal bars indicate the fractions of cells assigned to G1 (left peak, n=5469) and S+G2 (right peak, n=1945). B-D) Individual regression analysis of the nuclear expression values sorted according to the corresponding cell-cycle phase (G1 - blue; S+G2 - red) for RPA vs. XPA (B), Ki67 vs. XPA (C) and Ki67 vs. RPA (D). B-D) Red and blue lines represent linear regression with correlation coefficient r and p-value. 95% confidence bounds of all correlation coefficients r were estimated by non-parametric bootstrap and are given in brackets.

Remarkably, whereas the correlation coefficients between XPA and RPA remain constant in both regimes (G1: 0.71, S+G2: 0.68, all: 0.74) the correlation between XPA and Ki67 disappear (G1: -0.072, S+G2: -0.068). Between RPA and Ki67 there is close to no correlation in the G1 phase but a significant small positive correlation in the S and G2 phase (G1: 0.12, S+G2: 0.29). These results strengthen the conclusions derived from the cross-correlation analysis in fibroblasts that NER factor expression is functionally

#### 4. Co-expression of NER repair factors

---

co-regulated. In particular, the missing correlation between the nuclear factors and the cell cycle marker after taking the cell-cycle into account suggests that the measured correlation is specific to NER proteins.

##### 4.1.5. Measured average response agrees with model prediction

To determine potential mutual dependencies in NER factor expression we performed two independent cross-correlation experiments, firstly based on microscopy of human fibroblast and secondly using flow cytometry in human brain pericytes. For both approaches we found conclusive evidence for a positive pairwise correlation of the nuclear repair factor concentration. As described in Section 4.1.1 the measured cross-correlations represent the entries of matrix  $B$  in Eqn. 4.4. In theory, the product of the inverse of  $B$  and  $A$  define the vector of response coefficients. However, due to the strong multicollinearity the matrix  $B$  has a variance inflation factor of **240**. This indicates that the standard error of the response coefficient estimates is about  $\sqrt{240} \cong 15$  times as large as it would be if the matrix  $B$  was uncorrelated.

Hence, we are not able to derive a reliable result for each response coefficient. However, it is possible to make an estimation of the average size of  $R$ . Consider  $\nu$  as a function of the concentration of the repair factors  $C_i$  and the initial amount of DNA lesions  $L$ :

$$\nu = Lf(C_1, \dots, C_N). \quad (4.5)$$

Using, once more, the standard law for the propagation of uncertainty analogue to Eqn. 3.11

$$\sigma(\nu) = \sqrt{\sum_i \left( \frac{\partial \nu}{\partial C_i} \sigma(C_i) \right)^2 + \left( \frac{\partial \nu}{\partial L} \sigma(L) \right)^2 + \left( \frac{\partial \nu}{\partial C_i} \frac{\partial \nu}{\partial C_j} \text{cov}_{ij} \right)^2} \quad (4.6)$$

and introducing the response coefficients from Eqn. 4.3 we derive

$$CV_\nu^2 = CV_L^2 + \sum_i (R_i CV_i)^2 + \sum_{j \neq i} R_i R_j \frac{\text{cov}_{ij}}{\langle C_i \rangle \langle C_j \rangle}. \quad (4.7)$$

With the simplifying assumption that  $R = R_i$  we can solve Eqn. 4.7 for  $R$  and thereby derive an estimate for the average response coefficient:

$$R = \sqrt{\frac{CV_\nu^2 - CV_L^2}{\sum_i CV_i^2 + \sum_{j \neq i} \frac{\text{cov}_{ij}}{\langle C_i \rangle \langle C_j \rangle}}}. \quad (4.8)$$

On the right-hand side we measured most of the CVs and about a third of the NER factor covariations. The missing entries were added by taking the average of the known elements. By inserting all measured and estimated quantities (summarized in Table...tbm) into Eqn. 4.8 we calculated an average response coefficient of  $R = 0.16$ . Remarkably, this result agrees exactly with the average response coefficients predicted by the model (cf. Figure 3.1) and thus, reconfirms the hypothesis of a small and collectively shared repair-rate control.

## 5. Discussion

The continuation of our DNA-repair analysis was motivated by the ability of explicitly observing the kinetics of resynthesized DNA during nucleotide-excission repair. The quantification of this repair intermediate by following the incorporation of fluorescently labelled nucleotides on the single cell level portrays the repair process as a slow, first-order kinetic. The new data supplement our comprehensive knowledge about the dynamic behaviour of single repair proteins comprising their individual exchange kinetics at the chromatin fibre, which eventually lead to the formation of catalytic multi-protein machineries. Based on this information we were able to identify a realistic kinetic model of NER, which established a quantitative link between fast and random assembly of the NER factors at the DNA template and the emergent phenomenon of a slow overall repair time. Most importantly, the model predicts a collective control of the repair rate by the repair factors, which rebuts the presumption of a rate-limiting step in the pathway. This finding has been corroborated experimentally exploiting the natural variability of the nuclear NER factor concentrations, which emphasizes the robustness of DNA repair against fluctuations in NER factor expression.

In fact, we observe that the substantial heterogeneity in repair synthesis is generated primarily by the distribution of inflicted DNA lesions. This factor is complemented by the variability contributed from each individual repair factor as well as by mutual dependencies in NER factor expression. The quantitative impact of these three parameters on the cell-to-cell variability of the repair rate is consistent with the robustness of the repair rate against concentration fluctuations of individual repair factors.

**Slow first order kinetics as a general systems property for chromatin associated processes.** During our analysis of the NER kinetic we focused on the removal of one of the major UV-induced DNA lesions, the 6-4PPs, which are repaired significantly faster ( $\sim$ 2-3 hours) compared to CPDs, which are still present after  $\sim$  24 hours [6, 53, 54]. 6-4PPs are cut out by incision of the damaged DNA strand, which we were able to monitor by following the incorporation of the fluorescently tagged uracil analogue EdU. The nucleotides are labelled by copper-catalysed covalent addition of the fluorophore for which the cells

have to be fixed. Hence, EdU accumulation cannot be measured continuously in one cell. Instead, we followed EdU incorporation stepwise in increasingly longer time intervals, each starting immediately after UV-irradiation. In this way we were able to quantify the 6-4PP repair time course which fits to a first order kinetic with a half-life of  $\sim$  1.2 hours. This long time scale (hours) stays in striking contrast to several live-cell imaging studies, which determined the time of engagement on damaged DNA for early and late repair factors in the range of minutes [9, 13, 21, 22, 26, 55, 56]. Recently, this result could be generalized for all core NER factors combining different photobleaching and fluorescence life-time microscopy approaches [6].

The present work strengthens a prominent explanation of this apparent contradiction using the comprehensive biological knowledge about NER together with mathematical modelling. In agreement with a simplified model simulating the assembly of a catalytic protein repair-complex [11] we find that reversible exchanging NER factors resemble the key property that makes NER an apparent first-order process. Accordingly, the long time-scale arises from the concurrence of stochastic assembly and reversible exchange, which leads to many non-functional preliminary stages before eventually the catalytic complex is formed. As suggested by Luijsterburg *et al.* (2010) [6], the slow repair time might even present an advantage for the specificity of the repair process due to differential dissociation kinetics in case of a false-positive lesion detection. This is in contrast to a sequence of irreversible binding steps, which will create sigmoidal kinetics with an even sharper delay with increasing numbers of assembling components [11].

The phenomenon of rapid protein exchange is a widespread property also found in other chromatin associated processes like replication, chromatin remodelling or transcription [57–60]. Analogue to NER, these proteins are involved in finding the specific reaction site first before the catalytic machinery can assemble at the same location. In particular in case of DNA transcription, transient protein dynamics were experimentally demonstrated for transcription factor binding and transcription machinery assembly [61]. The time scale for these reactions (several seconds to minutes) is comparable to our results for DNA repair, which encourages to hypothesise whether the slow time scales of transcriptional bursting in mammalian cells (tens of minutes to hours; [62, 63]) also arise through the reversible assembly of large macromolecular complexes.

## 5. Discussion

---

**An identifiable and hence predictive kinetic model of DNA repair.** Augmenting the extensive kinetic data of the binding and dissociation of individual components of the NER machinery [6] with a direct readout for DNA repair synthesis we were able to develop a predictive model of NER. To the best of our knowledge it is thereby the first model of a DNA-associated process, for which the model parameters could be identified (i.e., parameter values uniquely assigned with narrow confidence bounds) from the experimental data. A prerequisite for the computational feasibility of the identifiability analysis (calculating the profile likelihoods for each parameter) is the runtime optimization of the parameter estimation algorithm. Therefore, we integrated the model into a dynamic modelling framework that applies a deterministic trust region algorithm implemented in MATLAB [42]. Accuracy and computational performance of the implementation benefit immensely from the simultaneous solution of the sensitivity equations [34, 64, 65], which provide the required derivatives of the objective function with respect to the parameters. In fact, solving this extended system of ordinary differential equations with a CVODE solver implemented in C accelerated the speed for one estimation run by  $2 \times 10^2$  fold compared to the previously applied stochastic Markov Chain Monte Carlo (MCMC) optimization algorithm. The runtime spent for one fit on a 12-core processor with ...GHz each is now reduced to  $\sim 10$  minutes. This allowed us to sample the parameter space very effectively using Latin hypercube sampling, which makes the deterministic optimization algorithm more robust against local optima [34]. Simultaneous to the parameters determining the model dynamics we also estimated the parameters that characterize the measurement noise. This approach has the advantage that it avoids an inaccurate estimate of the experimentally determined error due to low numbers of replicates and thus has more exact assessment of the model parameters than using a preprocessing of the experimental data [34].

To achieve an identifiable NER model required a number of simplifications concerning the model structure. These were partly imposed by the limited collection of NER factors, which are currently available for live-cell imaging. In addition, the profile likelihood analysis itself allocated model substructures, which are not sufficiently described by the measurements. Accordingly, the two most significant modifications compared to the work of Luijsterburg *et al.* (2010) [6] concern the removal of the 'partially unwound' repair intermediate and the neglect of protein-protein interactions, in particular the sequential binding of XPA and XPF, for which

there was no evidence in the data. Due to the lack of a direct readout for incised DNA the incision of the DNA lesion has been assumed practically instantaneous once the pre-incision complex has been completely assembled. Despite the rate of rechromatinisation, this uncertainty certainly also affects all other catalytic rates, indicated by the existence of lower bounds, only. Consequently, the individual rate constants turn out to be fast, which agrees with a recent direct measurement in bacteria where DNA synthesis and ligation take seconds [66].

The current status, where the model structure is fully reconcilable with the experimental data, denotes a concrete progression beyond previous NER models [6, 23, 67]. At the same time this enclosed data-model entanglement defines the predictive power of the model and thus determines what can be said quantitatively about the regulation of NER. Consequently, gaining further mechanistic understanding about system properties requires a parallel and intertwined advancement where the addition of molecular detail is consecutively balanced with appropriate quantitative measurements. We expect that the general dynamic behaviour of the model (e.g. repair rate, robustness against protein-expression noise and fidelity of lesion recognition) will prove robust with respect to such a development as we observed these properties already for a simplified model of repair (cf. Figure 2.6).

**Model parametrization and its implications** The model fit yields biochemically plausible estimates for the kinetic parameters of the individual assembly and dissociation kinetics *in vivo* that account for both the long-term accumulation and the rapid exchange of the NER factors. Notably, the affinity of the lesion recognition factor XPC is remarkably low, which is mainly caused by the high  $k_{\text{off}}$  rate. This is consistent with previous work where high molecular off-rates have been described as a general property of self-organizing systems, which allows for efficient exploration of an assembly landscape and selection of a functional steady state [68]. As discussed in depth at Luijsterburg and coworkers (2010) [6] the strong reversibility is also beneficial for the specificity and regulability of the system, for example by preventing the trapping of NER proteins in incomplete (and thus unproductive) repair complexes. The same applies for the trade-off between specificity and efficiency of the mechanism, which equally plays a role in chromatin remodelling and transcription [69, 70].

The parameter set is complemented by the fast catalytic rate constants, which indicate a

## 5. Discussion

---

tight coupling between lesion excision and repair synthesis. This implies that the lesions are repaired without much delay immediately after their recognition by XPC, which consequently prevents the accumulation of incised DNA. This *in vivo* finding stays in contrast to *in vitro* experiments that have found a delay before repair synthesis [24, 71]. The predicted prevention of the single-stranded repair intermediate corresponds with our intuition that an elevated abundance of incised DNA would trigger DNA degradation to avoid the risk of an inflammatory response or autoimmunity [72].

**Collective control of the DNA repair rate** A consequence of the NER model architecture, expressing the sequential organization of this chromatin-associated process through cycles of protein recruitment, is the absence of a rate-limiting factor for repair synthesis. The control is rather homogeneously distributed with small contributions by each repair factor. This result is similar to the outcome of previous flux control analyses that found no experimental support for the existence of a unique rate-limiting factor [44]. Moreover, response coefficients with values much smaller than 1 indicate that DNA repair is robust against natural fluctuation of the protein concentrations in the cell [73]. It is important to note that the degree of control as a kinetic property of a particular NER factor is independent of the order of engagement in the repair process. Thus, the lesion recognition factor XPC, which binds first, can have the same control as XPA or RPA that bind much later to protect the unwound DNA single-strands. As a biochemical network can never be insensitive against all possible perturbations [73, 74], also DNA repair will be sensitive to larger fluctuations in the concentration of NER factors. In this respect, the model predicts that the rate of repair eventually drops to zero in case of a strong concentration reduction of any repair factor. This agrees with various studies showing that a significant reduction in nuclear XPC and XPA levels leads to decreased cell survival and/or decreased lesion removal [75–77]. In contrast, increased levels of ERCC1 and XPA lead to increased resistance in certain tumours against cisplatin, which induces DNA interstrand crosslinks specifically removed by NER [75–79].

To analyse the effect of natural changes in NER factor expression on the repair rate experimentally we exploited the natural variability of the nuclear protein concentration instead of changing the protein expression by gene knock-down or overexpression, which usually cause the partial or complete disruption of essential system functions [80]. Despite

the weak correlation in the data we saw a significantly positive dependency between DNA resynthesis and protein expression for all five measured NER factors. The regression slopes were evenly distributed and below unity, which corroborates the finding of a shared moderate repair-rate control.

Not surprisingly, there are studies indicating that the collective control of a systems property (e.g. repair rate) is not an exclusive concept for the DNA repair pathway, but also applies to other chromatin-associated processes. For example, it was shown that the probability to express interferon- $\beta$  was significantly increased when five out of six transcription factors, necessary for transcriptional activation, were overexpressed [81]. A very similar result was recently obtained, which suggested the elevated expression of IL-2 by the concerted interplay of four transcription factors together with FOXP3 [82]. These findings emphasize the probability of a collective process also for transcription.

**Co-regulation of the nuclear NER factor expression** For the mechanistic interrelation between robustness as a systems property and the corresponding biochemical network several regulatory motifs on the molecular level (e.g. negative feedback, incoherent feed-forward loops and functional redundancy) have been proposed [83–86]. However, most of these functional building blocks come not into question for promoting robust DNA repair as the NER pathway manages without any reported signalling events (shown in this work), nor should transcriptional regulation play a role on the considered time scale [7]. The individual repair reactions purely depend on the repair complex assembly time and thus are determined by the NER factor binding affinity (this work and [6]). Due to the long, random and reversible, assembly process of the whole repair complex minor delays or speed-ups in one assembly step caused by small fluctuations of one repair factor are negligible on the global time scale. In this way the robustness of DNA repair is directly linked to the dynamic design of the repair pathway.

Intriguingly, we found experimental evidence for the regulated expression of the NER factors themselves, indicated by a strongly positive pairwise correlation for all seven measured repair factor combinations. In contrast to previous studies showing correlated mRNA levels for different NER factors in cancers cells [87,88] we observe this cross-correlated expression in single cells of different type under unperturbed, NER independent conditions. Despite

## *5. Discussion*

---

a general increase in the nuclear NER factor concentration during the S-phase and G2 phase the cross-correlation persists independently of the cell cycle. Together, these results imply an additional regulative motive that indirectly controls the repair rate already on the transcriptional level.

Remarkably, the co-expression of NER factors has strong conceptual similarity to bacterial signal transduction systems where different regulatory enzymes are encoded on the same operon [?, 89]. This co-localisation leads to co-expression, which in turn provides concentration robustness against variability of other pathway components [73]. So far such a transcriptional coupling mechanism is unknown for the repair proteins involved in NER. It would be therefore interesting to investigate whether this co-expression can be perturbed pre-transcriptionally to further study its impact on the rate of repair.

**Concluding remarks** The accurate repair of the nuclear DNA is pivotal for the enduring vitality of the cell and in particular for the fidelity of all chromatin-associated processes. It is therefore not surprising to find two cooperating molecular approaches that ensure robustness of the repair rate against natural variability. First, as a consequence to the dynamic design of the repair complex assembly there is only a moderate and uniformly distributed control of the repair rate by each NER factor, which translates into a low systems response in the event of NER factor fluctuations. Second, a so far unknown extrinsic mechanism regulates the coordinated NER factor expression presumably on the transcriptional level, which limits the degree of protein variability already pre-transcriptionally. Together both mechanisms might be a widespread mode of robust regulation of chromatin-associated processes, given that chromatin remodelling, transcription and translation have similar dynamic features as DNA repair.



## 6. Bibliography

- [1] Hoeijmakers, J. H. J. DNA damage, aging, and cancer. *The New England journal of medicine* **361**, 1475–85 (2009).
- [2] Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* **362**, 709–715 (1993).
- [3] Lhomme, J., Constant, J.-F. & Demeunynck, M. Abasic DNA Structure , Reactivity , and Recognition. *Biopolymers* **52**, 65–83 (1999).
- [4] De Bont, R. Endogenous DNA damage in humans: a review of quantitative data. *Mutagenesis* **19**, 169–185 (2004).
- [5] Sander, M., Cadet, J., Casciano, D. a., Galloway, S. M., Marnett, L. J., Novak, R. F., Pettit, S. D., Preston, R. J., Skare, J. a., Williams, G. M., Van Houten, B. & Gollapudi, B. B. Proceedings of a workshop on DNA adducts: biological significance and applications to risk assessment Washington, DC, April 13-14, 2004. *Toxicology and applied pharmacology* **208**, 1–20 (2005).
- [6] Luijsterburg, M. S., von Bornstaedt, G., Gourdin, A. M., Politi, A. Z., Moné, M. J., Warmerdam, D. O., Goedhart, J., Vermeulen, W., van Driel, R. & Höfer, T. Stochastic and reversible assembly of a multiprotein DNA repair complex ensures accurate target site recognition and efficient repair. *The Journal of Cell Biology* **189**, 445–63 (2010).
- [7] Moné, M. J., Volker, M., Nikaido, O., Mullenders, L. H., van Zeeland, a. a., Verschure, P. J., Manders, E. M. & van Driel, R. Local UV-induced DNA damage in cell nuclei results in local transcription inhibition. *EMBO reports* **2**, 1013–7 (2001).
- [8] Houtsmuller, A. & Vermeulen, W. Macromolecular dynamics in living cell nuclei revealed by fluorescence redistribution after photobleaching. *Histochemistry and cell biology* **115**, 13–21 (2001).
- [9] Moné, M. J., Bernas, T., Dinant, C., Goedvree, F. a., Manders, E. M. M., Volker, M., Houtsmuller, A. B., Hoeijmakers, J. H. J., Vermeulen, W. & van Driel, R. In vivo dynamics of chromatin-associated complex formation in mammalian nucleotide excision repair. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 15933–7 (2004).
- [10] Verbruggen, P., Heinemann, T., Manders, E., von Bornstaedt, G., van Driel, R. & Höfer, T. Robustness of DNA Repair through Collective Rate Control. *PLoS Computational Biology* **10**, e1003438 (2014).
- [11] Terstiege, G. Mammalian DNA Repair: a Model for Multi-protein Complex Assembly on Chromatin. Ph.D. thesis, University of Heidelberg (2010).
- [12] Araújo, S. J., Nigg, E. A. & Wood, R. D. Strong Functional Interactions of TFIIH with XPC and XPG in Human DNA Nucleotide Excision Repair , without a Preassembled Repairosome. *Molecular and cellular biology* **21**, 2281–2291 (2001).
- [13] Houtsmuller, a. B. Action of DNA Repair Endonuclease ERCC1/XPF in Living Cells. *Science* **284**, 958–961 (1999).

- [14] Limsirichaikul, S., Niimi, A., Fawcett, H., Lehmann, A., Yamashita, S. & Ogi, T. A rapid non-radioactive technique for measurement of repair synthesis in primary human fibroblasts by incorporation of ethynyl deoxyuridine (EdU). *Nucleic acids research* **37**, e31 (2009).
- [15] Essers, J., Theil, A. F., Baldeyron, C., Wiggert, A., Cappellen, V., Houtsmuller, A. B., Kanaar, R. & Cappellen, W. A. V. Nuclear Dynamics of PCNA in DNA Replication and Repair. *Molecular and cellular biology* **25**, 9350–9359 (2005).
- [16] Sporbert, A., Gahl, A., Ankerhold, R., Leonhardt, H. & Cardoso, M. C. DNA Polymerase Clamp Shows Little Turnover at Established Replication Sites but Sequential De Novo Assembly at Adjacent Origin Clusters. *Molecular cell* **10**, 1355–1365 (2002).
- [17] Evans, E., Moggs, J. G., Hwang, J. R., Egly, J.-m. & Wood, R. D. Mechanism of open complex and dual incision formation by human nucleotide excision repair factors. *EMBO Journal* **16**, 6559–6573 (1997).
- [18] Mu, D., Hsu, D. S. & Sancar, A. Reaction Mechanism of Human DNA Repair Excision Nuclease. *Journal of Biological Chemistry* **271**, 8285–8294 (1996).
- [19] Polo, S. E., Roche, D. & Almouzni, G. New histone incorporation marks sites of UV repair in human cells. *Cell* **127**, 481–93 (2006).
- [20] Tapias, A., Auriol, J., Forget, D., Enzlin, J. H., Schärer, O. D., Coin, F., Coulombe, B. & Egly, J.-M. Ordered conformational changes in damaged DNA induced by nucleotide excision repair factors. *The Journal of biological chemistry* **279**, 19074–83 (2004).
- [21] Rademakers, S., Volker, M., Hoogstraten, D., Nigg, A. L., Moné, M. J., Zeeland, A. A. V., Hoeijmakers, J. H. J., Houtsmuller, A. B. & Vermeulen, W. Xeroderma Pigmentosum Group A Protein Loads as a Separate Factor onto DNA Lesions. *Molecular and cellular biology* **23**, 5755–5767 (2003).
- [22] Zotter, A., Luijsterburg, M. S., Warmerdam, D. O., Ibrahim, S., Nigg, A., van Cappellen, W. a., Hoeijmakers, J. H. J., van Driel, R., Vermeulen, W. & Houtsmuller, A. B. Recruitment of the nucleotide excision repair endonuclease XPG to sites of UV-induced dna damage depends on functional TFIIH. *Molecular and cellular biology* **26**, 8868–79 (2006).
- [23] Politi, A., Moné, M. J., Houtsmuller, A. B., Hoogstraten, D., Vermeulen, W., Heinrich, R. & van Driel, R. Mathematical modeling of nucleotide excision repair reveals efficiency of sequential assembly strategies. *Molecular cell* **19**, 679–90 (2005).
- [24] Riedl, T., Hanaoka, F. & Egly, J.-M. The comings and goings of nucleotide excision repair factors on damaged DNA. *The EMBO journal* **22**, 5293–303 (2003).
- [25] Yokoi, M., Masutani, C., Maekawa, T., Sugasawa, K., Ohkuma, Y. & Hanaoka, F. The xeroderma pigmentosum group C protein complex XPC-HR23B plays an important role in the recruitment of transcription factor IIH to damaged DNA. *J Biol Chem* **275**, 9870–9875 (2000).
- [26] Volker, M., Moné, M. J., Karmakar, P., van Hoffen, a., Schul, W., Vermeulen, W., Hoeijmakers, J. H., van Driel, R., van Zeeland, a. a. & Mullenders, L. H. Sequential assembly of the nucleotide excision repair factors in vivo. *Molecular cell* **8**, 213–24 (2001).

- [27] O'Donovan, A., Davies, A. A., Moggs, J. G., West, S. C. & Wood, R. D. XPG endonuclease makes the 3' incision in human DNA nucleotide excision repair. *Nature* **371**, 432–435 (1994).
- [28] Sijbers, A. M., de Laat, W. L., Ariza, R. R., Biggerstaff, M., Wei, Y. F., Moggs, J. G., Carter, K. C., Shell, B. K., Evans, E., de Jong, M. C., Rademakers, S., de Rooij, J., Jaspers, N. G., Hoeijmakers, J. H. & Wood, R. D. Xeroderma pigmentosum group F caused by a defect in a structure-specific DNA repair endonuclease. *Cell* **86**, 811–822 (1996).
- [29] Winkler, G. S., Sugasawa, K., Eker, a. P., de Laat, W. L. & Hoeijmakers, J. H. Novel functional interactions between nucleotide excision DNA repair proteins influencing the enzymatic activities of TFIIH, XPG, and ERCC1-XPF. *Biochemistry* **40**, 160–5 (2001).
- [30] de Laat, W. L., Appeldoorn, E., Sugasawa, K., Weterings, E., Jaspers, N. G. & Hoeijmakers, J. H. DNA-binding polarity of human replication protein A positions nucleases in nucleotide excision repair. *Genes Dev* **12**, 2598–2609 (1998).
- [31] Moser, J., Volker, M., Kool, H., Alekseev, S., Vrieling, H., Yasui, A., van Zeeland, A. a. & Mullenders, L. H. F. The UV-damaged DNA binding protein mediates efficient targeting of the nucleotide excision repair complex to UV-induced photo lesions. *DNA repair* **4**, 571–82 (2005).
- [32] Shivji, M. K., Podust, V. N., Hübscher, U. & Wood, R. D. Nucleotide excision repair DNA synthesis by DNA polymerase epsilon in the presence of PCNA, RFC, and RPA. *Biochemistry* **34**, 5011–5017 (1995).
- [33] Moser, J., Kool, H., Giakzidis, I., Caldecott, K., Mullenders, L. H. & Fousteri, M. I. Sealing of chromosomal DNA nicks during nucleotide excision repair requires XRCC1 and DNA ligase III alpha in a cell-cycle-specific manner. *Mol Cell* **27**, 311–323 (2007).
- [34] Raue, A., Schilling, M., Bachmann, J., Matteson, A., Schelker, M., Schelke, M., Kaschek, D., Hug, S., Kreutz, C., Harms, B. D., Theis, F. J., Klingmüller, U. & Timmer, J. Lessons learned from quantitative dynamical modeling in systems biology. *PLoS one* **8**, e74335 (2013).
- [35] Leis, J. R. & Kramer, M. A. The Simultaneous Solution and Sensitivity Analysis of Systems Described by Ordinary Differential Equations. *ACM Transactions on Mathematical Software* **14**, 45–60 (1988).
- [36] Hindmarsh, A. C., Brown, P. N., Grant, K. E., Lee, S. L., Serban, R., Shumaker, D. A. N. E. & Woodward, C. S. SUNDIALS : Suite of Nonlinear and Differential / Algebraic Equation Solvers. *ACM Transactions on Mathematical Software* **31**, 363–396 (2005).
- [37] Owen, A. B. A Central Limit Theorem for Latin Hypercube Sampling. *Royal Statistical Society* **54**, 541–551 (2014).
- [38] Cobelli, C. & DiStefano, J. J. I. Parameter and structural identifiability concepts and ambiguities: a critical review and analysis. *The American Physiological Society* **239**, 7–24 (1980).
- [39] Swameye, I., Müller, T., Timmer, J., Sandra, O. & Klingmüller, U. Identification of nucleocytoplasmic cycling as a remote sensor in cellular signaling by databased modeling. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 1028–1033 (2003).

- [40] Venzon, D. J. & Moolgavkar, S. H. A method for computing profile-likelihood-based confidence intervals. *Applied Statistics* **37**, 87–94 (1988).
- [41] Murphy, S. A. & van der Vaart, A. W. On Profile Likelihood. *Journal of the American Statistical Association* **95**, 449–485 (2000).
- [42] Raue, a., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U. & Timmer, J. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics (Oxford, England)* **25**, 1923–9 (2009).
- [43] Hofmeyr, J.-H. S. & Cornish-Bowden, A. Quantitative assessment of regulation in metabolic systems. *European Journal of Biochemistry* **200**, 223–236 (1991).
- [44] Fell, D. A. Metabolic Control Analysis : a survey of its theoretical and experimental development. *Biochemical Journal* **286**, 313–330 (1992).
- [45] Kreutz, C., Raue, A. & Timmer, J. Likelihood based observability analysis and confidence intervals for predictions of dynamic models. *BMC systems biology* **6**, 120 (2012).
- [46] Hinkley, D. Predictive Likelihood. *The Annals of Statistics* **7**, 718–728 (1979).
- [47] Pearson, K. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine* **2**, 559–572 (1901).
- [48] Kendall, M. G. & Stuart, A. The Advanced Theory of Statistics. McMillan, London, 4th ed. (1979).
- [49] Fisher, R. Statistical Methods for Research Workers. Hafner, 13th ed. (1958).
- [50] Efron, B. Bootstrap methods: another look at the jackknife. *Annual Statistics* **7**, 1–26 (1979).
- [51] Burry, R. W. Controls for immunocytochemistry: an update. *The journal of histochemistry and cytochemistry* **59**, 6–12 (2011).
- [52] Giepmans, B. N. G., Adams, S. R., Ellisman, M. H. & Tsien, R. Y. The fluorescent toolbox for assessing protein location and function. *Science (New York, N.Y.)* **312**, 217–24 (2006).
- [53] Smerdon, M. J., Tlsty, T. D. & Lieberman, M. W. Distribution of ultraviolet-induced DNA repair synthesis in nuclelease sensitive and resistant regions of human chromatin. *Biochemistry* **17**, 2377–2386 (1978).
- [54] van Hoffen, A., Venema, J., Meschini, R., van Zeeland, A. A. & Mullenders, L. H. Transcription-coupled repair removes both cyclobutane pyrimidine dimers and 6-4 photoproducts with equal efficiency and in a sequential way from transcribed DNA in xeroderma pigmentosum group C fibroblasts. *EMBO J* **14**, 360–367 (1995).
- [55] Hoogstraten, D., Nigg, A. L., Heath, H., Mullenders, L. H. F., van Driel, R., Hoeijmakers, J. H. J., Vermeulen, W. & Houtsmuller, A. B. Rapid switching of TFIIH between RNA polymerase I and II transcription and DNA repair in vivo. *Molecular cell* **10**, 1163–74 (2002).
- [56] Hoogstraten, D., Bergink, S., Ng, J. M. Y., Verbiest, V. H. M., Luijsterburg, M. S., Geverts, B., Raams, a., Dinant, C., Hoeijmakers, J. H. J., Vermeulen, W. & Houtsmuller, a. B. Versatile DNA damage detection by the global genome nucleotide excision repair protein XPC. *Journal of Cell Science* **121**, 3991–3991 (2008).

- [57] McNairn, A. J., Okuno, Y., Misteli, T. & Gilbert, D. M. Chinese hamster ORC subunits dynamically associate with chromatin throughout the cell-cycle. *Experimental cell research* **308**, 345–56 (2005).
- [58] Erdel, F. & Rippe, K. Binding kinetics of human ISWI chromatin-remodelers to DNA repair sites elucidate their target location mechanism. *Nucleus (Austin, Tex.)* **2**, 105–12 (2011).
- [59] Sonneville, R., Querenet, M., Craig, A., Gartner, A. & Blow, J. J. The dynamics of replication licensing in live *Caenorhabditis elegans* embryos. *The Journal of cell biology* **196**, 233–46 (2012).
- [60] Stasevich, T. J. & McNally, J. G. Assembly of the transcription machinery: ordered and stable, random and dynamic, or both? *Chromosoma* **120**, 533–45 (2011).
- [61] Hager, G. L., McNally, J. G. & Misteli, T. Transcription dynamics. *Molecular cell* **35**, 741–53 (2009).
- [62] Harper, C. V., Finkenstädt, B., Woodcock, D. J., Friedrichsen, S., Semprini, S., Ashall, L., Spiller, D. G., Mullins, J. J., Rand, D. a., Davis, J. R. E. & White, M. R. H. Dynamic analysis of stochastic transcription cycles. *PLoS biology* **9**, e1000607 (2011).
- [63] Suter, D. M., Molina, N., Gatfield, D., Schneider, K., Schibler, U. & Naef, F. Mammalian genes are transcribed with widely different bursting kinetics. *Science (New York, N.Y.)* **332**, 472–4 (2011).
- [64] Conn, A. R., Scheinberg, K. & Vicente, L. N. Introduction to derivative-free optimization, vol. 8. Siam (2009).
- [65] Ramachandran, R. & Barton, P. I. Effective parameter estimation within a multi-dimensional population balance model framework. *Chemical Engineering Science* **65**, 4884–4893 (2010).
- [66] Uphoff, S., Reyes-Lamothe, R., Garza de Leon, F., Sherratt, D. J. & Kapanidis, A. N. Single-molecule DNA repair in live bacteria. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 8063–8 (2013).
- [67] Kesseler, K. J., Kaufmann, W. K., Reardon, J. T., Elston, T. C. & Sancar, A. A mathematical model for human nucleotide excision repair: damage recognition by random order assembly and kinetic proofreading. *Journal of theoretical biology* **249**, 361–75 (2007).
- [68] Kirschner, M., Gerhart, J. & Mitchison, T. Molecular 'Vitalism'. *Cell* **100**, 79–88 (2000).
- [69] Cook, P. R. A model for all genomes: the role of transcription factories. *Journal of molecular biology* **395**, 1–10 (2010).
- [70] Voss, T. C., Schiltz, R. L., Sung, M.-H., Yen, P. M., Stamatoyannopoulos, J. a., Biddie, S. C., Johnson, T. a., Miranda, T. B., John, S. & Hager, G. L. Dynamic exchange at regulatory elements during chromatin remodeling underlies assisted loading mechanism. *Cell* **146**, 544–54 (2011).
- [71] Mocquet, V., Lainé, J. P., Riedl, T., Yajin, Z., Lee, M. Y. & Egly, J. M. Sequential recruitment of the repair factors during NER: the role of XPG in initiating the resynthesis step. *The EMBO journal* **27**, 155–67 (2008).
- [72] Takeuchi, O. & Akira, S. Pattern recognition receptors and inflammation. *Cell* **140**, 805–20 (2010).
- [73] Blüthgen, N. & Legewie, S. Robustness of signal transduction pathways. *Cellular and molecular life sciences : CMLS* **70**, 2259–69 (2013).

- [74] Csete, M. E. & Doyle, J. C. Reverse engineering of biological complexity. *Science (New York, N.Y.)* **295**, 1664–9 (2002).
- [75] Köberle, B., Masters, J. R., Hartley, J. a. & Wood, R. D. Defective repair of cisplatin-induced DNA damage caused by reduced XPA protein in testicular germ cell tumours. *Current Biology* **9**, 273–278 (1999).
- [76] Köberle, B., Roginskaya, V. & Wood, R. D. XPA protein as a limiting factor for nucleotide excision repair and UV sensitivity in human cells. *DNA repair* **5**, 641–8 (2006).
- [77] Renaud, E., Miccoli, L., Zacal, N., Biard, D. S., Craescu, C. T., Rainbow, A. J. & Angulo, J. F. Differential contribution of XPC, RAD23A, RAD23B and CENTRIN 2 to the UV-response in human cells. *DNA repair* **10**, 835–47 (2011).
- [78] Stewart, D. J. Mechanisms of resistance to cisplatin and carboplatin. *Critical reviews in oncology/hematology* **63**, 12–31 (2007).
- [79] Arora, S., Kothandapani, A., Tillison, K., Kalman-Maltese, V. & Patrick, S. M. Downregulation of XPF-ERCC1 enhances cisplatin efficacy in cancer cells. *DNA repair* **9**, 745–53 (2010).
- [80] Moriya, H., Shimizu-Yoshida, Y. & Kitano, H. In vivo robustness analysis of cell division cycle genes in *Saccharomyces cerevisiae*. *PLoS genetics* **2**, e111 (2006).
- [81] Apostolou, E. & Thanos, D. Virus Infection Induces NF-kappaB-dependent interchromosomal associations mediating monoallelic IFN-beta gene expression. *Cell* **134**, 85–96 (2008).
- [82] Bendfeldt, H., Benary, M., Scheel, T., Steinbrink, K., Radbruch, A., Herz, H. & Baumgrass, R. IL-2 Expression in Activated Human Memory FOXP3(+) Cells Critically Depends on the Cellular Levels of FOXP3 as Well as of Four Transcription Factors of T Cell Activation. *Frontiers in immunology* **3**, 264 (2012).
- [83] Alon, U. Network motifs: theory and experimental approaches. *Nature reviews. Genetics* **8**, 450–61 (2007).
- [84] Bleris, L., Xie, Z., Glass, D., Adadevoh, A., Sontag, E. & Benenson, Y. Synthetic incoherent feedforward circuits show adaptation to the amount of their genetic template. *Molecular systems biology* **7**, 519 (2011).
- [85] Yi, T.-M., Huang, Y., Simon, M. I. & Doyle, J. Robust perfect adaptation in bacterial chemotaxis through integral feedback control. *Proceedings of the National Academy of Sciences* **97**, 4649–4653 (2000).
- [86] Muzzey, D., Gómez-Uribe, C. A., Mettetal, J. T. & van Oudenaarden, A. A systems-level analysis of perfect adaptation in yeast osmoregulation. *Cell* **138**, 160–171 (2011).
- [87] Damia, G., Guidi, G. & Incalci, M. D. Original Paper Expression of Genes Involved in Nucleotide Excision Repair and Sensitivity to Cisplatin and Melphalan in Human Cancer Cell Lines. *European Journal of Cancer* **34**, 1–6 (1998).
- [88] Cheng, L., Spitz, M. R., Hong, W. K. & Wei, Q. Reduced expression levels of nucleotide excision repair genes in lung cancer : a case-control analysis Epidemiological studies have indicated that reduced DNA repair capacity and increased DNA adduct levels are associated with increased risk of lung cancer. *Carcinogenesis* **21**, 1527–1530 (2000).

- [89] Kollmann, M., Løvdok, L., Bartholomé, K., Timmer, J. & Sourjik, V. Design principles of a bacterial signalling network. *Nature* **438**, 504–7 (2005).



## A. Appendix

### A.1. Parameter estimation

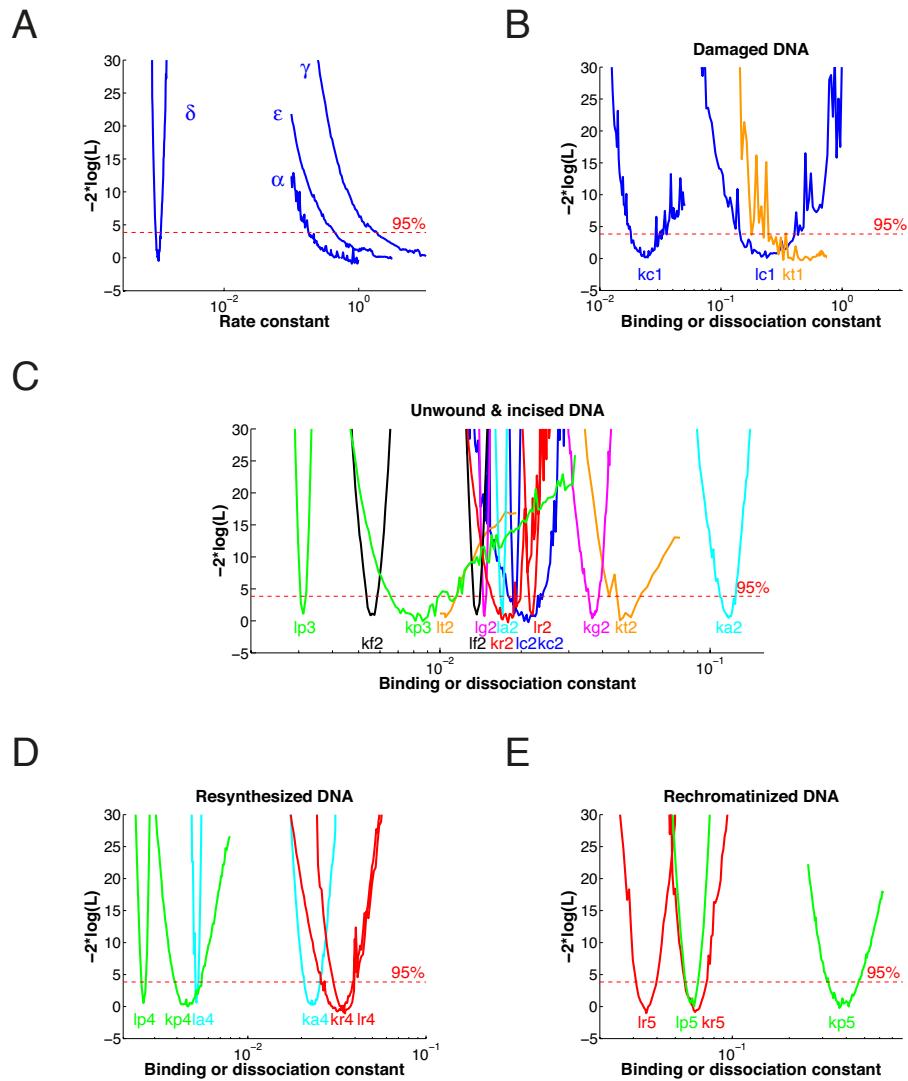


Figure A.1.:

## A. Appendix

Table S1. Values of binding and dissociation rate constants

Value	XPC	TFIIFH	XPG	XPF	XPA	RPA	PCNA
Concentration ( $\mu\text{M}$ )	0.140	0.360	0.440	1.110	0.170	1.110	1.110
<b>Damaged DNA</b>							
$k_{\text{on}}(\mu\text{M}^{-1}\text{s}^{-1})$	0.025 (0.013;0.036)	0.2 (0.152;0.288)	NA	NA	NA	NA	NA
$k_{\text{off}}(\text{s}^{-1})$	0.231 (0.136;0.36)	0.01 (0.009;0.012)	NA	NA	NA	NA	NA
$K_d(\mu\text{M})$	9.35 (3.46;16.09)	0.052 (0.044;0.072)	NA	NA	NA	NA	NA
<b>Unwound DNA</b>							
$k_{\text{on}}(\mu\text{M}^{-1}\text{s}^{-1})$	0.022 (0.016;0.025)	0.049 (0.04;0.06)	0.037 (0.036;0.039)	0.0056 (0.0054;0.0059)	0.116 (0.109;0.125)	0.018 (0.016;0.02)	NA
$k_{\text{off}}(\text{s}^{-1})$	0.019 (0.019;0.019)	0.01 (0.009;0.012)	0.0146 (0.0140;0.0149)	0.0137 (0.0134;0.0141)	0.017 (0.0167;0.0172)	0.022 (0.0213;0.0224)	NA
$K_d(\mu\text{M})$	0.864 (0.635;1.006)	0.204 (0.163;0.259)	0.395 (0.373;0.419)	2.446 (2.344; 2.596)	0.147 (0.138;0.158)	1.222 (1.048;1.36)	NA
<b>Incised DNA</b>							
$k_{\text{on}}(\mu\text{M}^{-1}\text{s}^{-1})$	0.022 (0.016;0.025)	0.049 (0.04;0.06)	0.037 (0.036;0.039)	0.0056 (0.0054;0.0059)	0.116 (0.109;0.125)	0.018 (0.016;0.02)	0.008 (0.0066;0.0111)
$k_{\text{off}}(\text{s}^{-1})$	0.019 (0.019;0.019)	0.01 (0.009;0.012)	0.0146 (0.0140;0.0149)	0.0137 (0.0134;0.0141)	0.017 (0.0167;0.0172)	0.022 (0.0213;0.0224)	0.0031 (0.003;0.0032)
$K_d(\mu\text{M})$	0.864 (0.635;1.006)	0.204 (0.163;0.259)	0.395 (0.373;0.419)	2.446 (2.344; 2.596)	0.147 (0.138;0.158)	1.222 (1.048;1.36)	0.388 (0.319;0.538)
<b>Resynthesized DNA</b>							
$k_{\text{on}}(\mu\text{M}^{-1}\text{s}^{-1})$	NA	NA	NA	NA	0.022 (0.021;0.025)	0.032 (0.025;0.037)	0.004 (0.0038;0.005)
$k_{\text{off}}(\text{s}^{-1})$	NA	NA	NA	NA	0.0052 (0.0051;0.0053)	0.035 (0.031;0.042)	0.0026 (0.0025;0.0027)
$K_d(\mu\text{M})$	NA	NA	NA	NA	0.236 (0.222;0.27)	1.167 (0.924;1.521)	0.605 (0.531;0.747)
<b>Rechromatinized DNA</b>							
$k_{\text{on}}(\mu\text{M}^{-1}\text{s}^{-1})$	NA	NA	NA	NA	NA	0.065 (0.056;0.075)	0.396 (0.359;0.457)
$k_{\text{off}}(\text{s}^{-1})$	NA	NA	NA	NA	NA	0.035 (0.031;0.04)	0.061 (0.056;0.067)
$K_d(\mu\text{M})$	NA	NA	NA	NA	NA	0.538 (0.438;0.645)	0.154 (0.134;0.182)

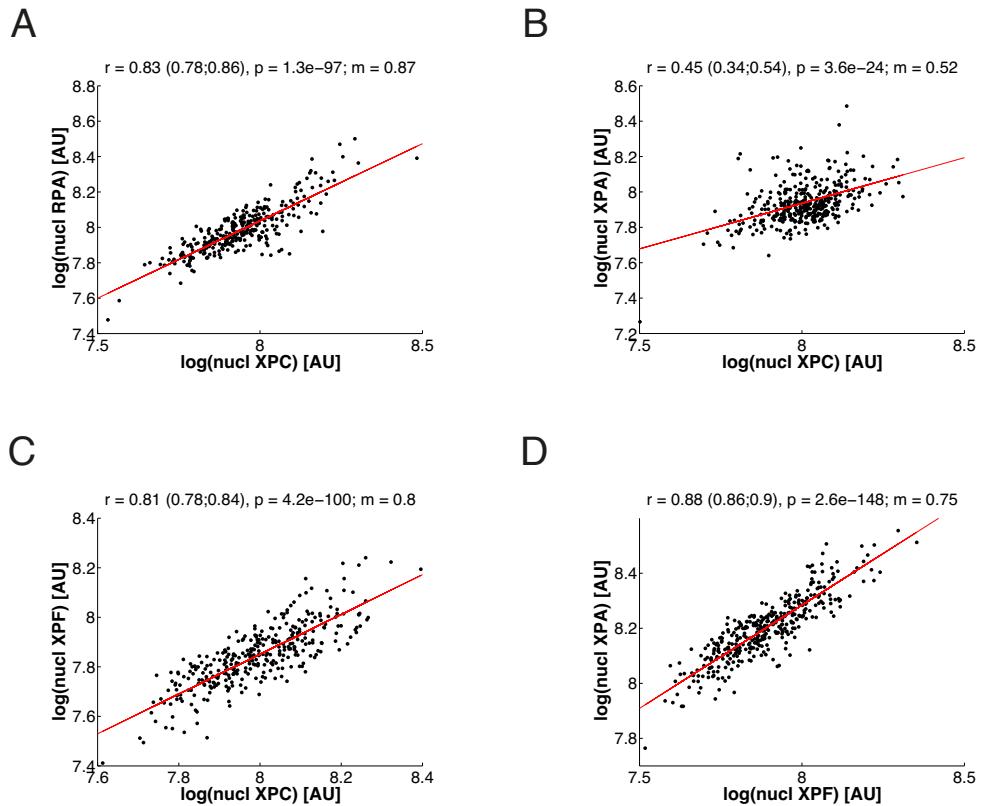
Table A1.: Blubb

Table SII. Values of the enzymatic rate constants

Enzymatic rate	$k_{\text{cat}}$
	$\text{s}^{-1}$
Unwinding $\alpha$	19.9 (>0.2)
Resynthesis $\gamma$	25.5 (>1.5)
Rechromatinization $\delta$	0.001 (0.001;0.0011)
Reannealing $\epsilon$	5.3 (>0.9)

Table A2.: Blubb

## A.2. Cross-correlation



**Figure A.2.:**