

Inaugural - Dissertation
submitted to the
Combined Faculty 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

Robust Regulation of DNA Repair

Referees: **Prof. Dr. Thomas Höfer**
Prof. Dr. Ursula Kummer

Contents

Zusammenfassung	i
Summary	iii
1. Introduction	1
1.1. DNA damage and its consequences	1
1.2. NER as a model system for chromatin-associated processes	2
1.3. Research objectives	5
2. Fluorescence time-lapse imaging of the NER process	7
2.1. Locally UV-inflicted DNA damage	7
2.2. Repair factor accumulation and dissociation occur on different time scales	8
2.3. Direct measurement of DNA resynthesis	12
2.4. Repair rate follows first-order rate kinetics	12
3. Kinetic model of NER	15
3.1. Slow first-order reaction kinetic due to many fast interacting components	15
3.2. Model structure and parametrization	17
3.3. A maximum likelihood approach for efficient model fitting	22
3.4. NER model fits accumulation, FLIP and repair synthesis measurements	23
3.5. Profile likelihood analysis identifies realistic model of NER	26
4. Control analysis of the DNA repair rate	31
4.1. Kinetic NER model predicts collective rate control	31
4.2. Exploiting natural variability in protein expression to quantify rate control	34
4.3. Variable NER factor expression and inflicted lesions account for the distribution of repair rates	38
5. Co-expression of NER repair factors	47
5.1. Cross-correlation affects repair response	47
5.2. Microscopy analysis of co-staining experiments	49
5.3. Cell cycle independent cross-correlation of NER factor expression	52
5.4. Measured average response agrees with model prediction	53

6. Discussion	57
7. Bibliography	65
8. Abbreviations	73
A. Appendix	77
A.1. Parameter estimation	77
A.2. Correlation analysis	79
A.3. Control analysis	79
A.4. Cross-correlation	80

Zusammenfassung

Summary

DNA repair is indispensable for the intracellular protection against environmental and endogenous damaging agents. This is reflected in an increased susceptibility to cellular ageing and cancer development as a consequence to impaired repair. The repair process is carried out by enzymatic macromolecular complexes assembling at damaged sites on the chromatin fibre. Fluorescence imaging of the individual repair factor dynamics showed a rapid protein exchange at repair sites, while the overall repair time is significantly longer. How these fast molecular interactions regulate the emergence of higher-level system-properties, such as the repair rate and its robustness, is poorly understood.

In this thesis we quantify the relation between the protein dynamics and the resulting repair by measuring the time course of nucleotide-excision DNA repair (NER) in mammalian cells. We find that, despite the pathway's molecular complexity, DNA repair follows a slow first-order reaction with a half-time of 1.2 hours.

To gain functional insight, we develop an identifiable mathematical model of the NER pathway. According to a model prediction repair synthesis is not controlled by one rate-limiting component alone. Instead, all NER factors collectively control the repair-rate, which is indicated by uniformly small response coefficients. Harnessing the natural cell-to-cell variability of protein expression, we show qualitative experimental support for the shared control of the core NER factors XPC, TFIIH, XPA, XPF and RPA. However, quantitatively the measured average repair-rate response is significantly larger than the predicted mean response coefficient. In search of additional sources contributing to the repair-rate control, we observe that the nuclear NER factor expression is strongly correlated suggesting a pre-transcriptional complementary control mechanism regulating NER factor expression. Notably, including the cross-correlation in the control analysis resolves the discrepancy between measured and predicted repair-rate response.

These findings portray two complementary modes of robust regulation in DNA repair that, given their similar dynamical design, may also be valid approaches in other chromatin-associated processes such as transcription or replication.

1. Introduction

1.1. DNA damage and its consequences

The impeccable preservation of DNA is an essential requirement for the correct functioning of cellular processes. Without an active repair mechanism DNA would accumulate on average $\sim 10^5$ damages per cell on a single day [1]. The inflicted lesions can be grouped into three different categories: 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. These species react with the DNA and form molecular distortions ranging from several kinds of single-strand breaks to various oxidative base and sugar products [4, 5]. And third, DNA damages may be caused by external physical or chemical agents such as sunlight, which induces approximately 30,000 pyrimidine dimers per hour in each exposed keratinocyte [6].

If unrecognised, these DNA aberrations persist and accumulate in the genetic code and may lead to the activation of oncogenes or the inactivation of tumour suppressor genes and thereby facilitate the risk to develop cancer [7]. Or, alternatively, the damage impairs essential cellular processes such as transcription or replication. This is usually the case for inter and intra-strand cross-links as well as double-strand breaks induced by ionizing irradiation. Notably, already relatively low damage levels will lead to the inhibition of transcription [8]. The occurrence of such lesions causes cell death, which particularly in proliferative tissues promotes symptoms of premature ageing [9]. Depending on the location and number of lesions, the cell type, stage in the cell cycle and differentiation cancer development and cell death can happen concurrently. In fact, these two cell fates may be linked considering that damage-induced cell-death protects the body from cancer [1].

DNA damage response DNA is the only known biological macromolecule, which cannot be replaced if parts of it are destroyed. This has the inevitable consequence that any mutation that escapes detection will persist over the lifetime of the cell [1, 9]. For this reason, cells contain an elaborate genomic maintenance apparatus, the DNA damage response (DDR) [10], which includes various DNA-damage sensors and the enzyme machin-

ery catalysing the individual repair reactions. DDR also regulates a range of downstream mediators and signalling molecules that control cell cycle progression and cell fate. Moreover DDR comprises a number of complementary repair systems, each of which responsible for a specific type of damage [9, 10]. Subtle chemical alterations such as oxidative lesions and alkylation products are removed by base-excision repair (BER) [11, 12]. More complex lesions like inter-strand crosslinks (ICLs) or helix distorting lesions are repaired by ICL repair or nucleotide excision DNA repair (NER), respectively [1, 13], while double-strand breaks in the DNA are processed either by non-homologous end joining (NHEJ) or homologous repair (HR) [12, 14].

Spatiotemporal regulation of DNA repair For all repair mechanisms their spatiotemporal regulation follows the general organization of chromatin-associated processes: a) detection of the target site (e.g. a DNA lesion) b) assembly of the components to a multi-protein complex at the DNA and c) catalysis of an enzymatic reaction with the DNA as a substrate [6, 15–17]. Of particular interest is thereby the role of the assembly phase, which was until recently understood as a sequential, cooperative recruitment of components [18]. A competing perception considers the complex formation as a combined process of randomly diffusing and rapidly exchanging components [6]. The latter hypothesis is supported by experimental evidence for DNA repair but also transcription initiation showing that dissociation and rebinding of the constituents occur on the time scale of seconds to minutes [19, 20]. Similar findings are reported for the clustering of RNA and DNA polymerases during transcription and replication [21, 22]. However, it remains poorly understood how the dynamic interplay of these multi-protein machineries effects essential system properties like fidelity, timing and robustness.

1.2. NER as a model system for chromatin-associated processes

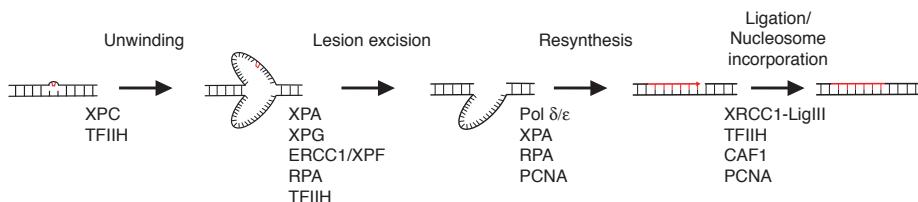
To gain more insight into the assembly and functionality of protein complexes, we have studied the mammalian NER pathway. NER is involved in the elimination of a wide range of structurally unrelated DNA lesions, including cyclobutane-pyrimidine dimers (CPDs) and 6-4 pyrimidine-pyrimidone photoproducts (6-4PPs), both induced by ultraviolet (UV) light; various bulky chemical adducts (man-made and natural); intrastrand crosslinks caused by drugs such as cisplatin; and cyclopurines generated by reactive oxygen species (ROS) [9]. Consequently, the NER machinery is in itself a composition of intertwined repair pathways, which give rise to an exceptional clinical heterogeneity associated with the impairment of

NER components.

Molecular mechanism of NER The process of damage detection by NER is distinguished into transcription-coupled NER (TC-NER) [16, 23], where bulky DNA adducts block the transcription machinery, and global genome NER (GG-NER), where the same adducts are eliminated throughout the whole genome in an transcription-independent pathway [24]. In this work we will focus on GG-NER, whose core repair factors have all been identified and intensively studied. Figure 1.1A shows the sequence of experimentally distinguishable repair intermediates and the core NER factors that bind to the corresponding intermediate. For GG-NER the initial lesion detection is performed by the single damage-sensing protein *xeroderma pigmentosum*, group C (XPC) [18, 25]. This is in contrast to TC-NER, where damage detection is mediated by the RNA polymerase itself [26]. The binding of XPC promotes the association of the TFIIH (transcription initiation factor IIH) complex to the lesion site [18, 27, 28]. TFIIH consists of 10 components out of which the subunits XPB and XPD have helicase activity with opposite polarity allowing the DNA to be unwound [29, 30]. Upon DNA unwinding XPA can bind to lesions on the damaged single strand while the replication protein A (RPA) protects the undamaged strand. Subsequently the endonucleases XPG and ERCC1/XPF excise approximately 30 nucleotides around the lesion [31–35]. Subsequently, the proliferating cell nuclear antigen (PCNA) loads the polymerase δ , which refills the single-strand DNA gap [15, 36, 37]. After ligation by LigIII-XRCC1 the DNA strand is then rechromatinised by CAF1, which completes the repair process [38, 39].

Data-driven kinetic modelling of NER Defects in GG-NER are associated with the prototypical nucleotide-excision-repair disorder *xeroderma pigmentosum* (XP), which is associated with pigmentation abnormalities and an increased risk of skin cancer and internal tumours [1]. XP is caused by gene-defects in one of the seven XP-genes, XPA to XPG. Despite the severe symptoms related to one of these defects all patients are viable, which allowed for the isolation of patient cell lines where the endogenous NER factor is either knocked-out or the translated NER factors are defective. Importantly, these cell lines could be transfected with a GFP-tagged construct of the missing protein enabling the fluorescence real-time measurement of the NER factor kinetics. The DNA repair-capacity of the 'restored' is similar to intact cells with functional repair systems thus justifying and validating the use of fluorescent derivatives [40–43]. Taken together, the clearly defined UV-mediated inducibility of NER combined with the availability of the fluorescently-tagged core repair factors makes NER a paradigm for the research on chromatin-associated processes.

A



B

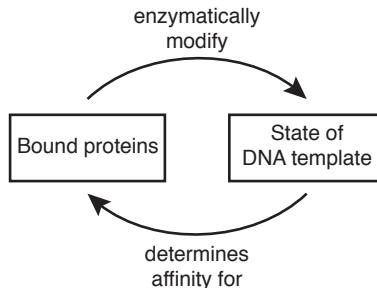


Figure 1.1.: NER as a paradigm for chromatin-associated processes. A) Graphical scheme of the NER pathway. Shown is the sequence of repair intermediates together with the repair factors that, if fully assembled, catalyse the next reaction. B) 'Recruiting-reaction' cycle illustrating the principle of the sequential organization of chromatin-associated processes. Therein, recruited process components catalytically modify the DNA template and thereby redefine the binding affinity for the next complex.

Based on accumulation and dissociation time course data for each NER factor Luijsterburg *et al.* [6] developed a quantitative model of the NER process. This model suggests for the formation of enzymatically active protein complexes a cyclic principle (*cf.* Figure 1.1B): Repair factors bind reversibly and assemble stochastically at the DNA template; once the complex is fully assembled it modifies the DNA substrate and thereby changes the binding affinity for the next enzymatic complex. Traversing multiple of these 'recruiting-reaction' cycles translates the recurring stochastic protein exchange at the chromatin site into an ordered sequence of regulatory events [17].

This model unravels a number of interesting relations between assembly dynamics and process functionality, for instance the rapid and reversible protein binding allows for high specificity of the pathway through kinetic proofreading [6]. In agreement, combinations of live cell experiments and mathematical modelling have led to equally intriguing results for transcription. As indicated by Gorski *et al.* [44] already a moderate increase in the dwell-time of RNA polymerase I (RNA Pol I) subunits enhanced the assembly efficiency and the initiation rate of RNA Pol I. Another study describes a mechanism termed 'assisted loading', where

1. Introduction

a weak transcription factor-DNA interaction facilitates the binding of a second transcription factor and thereby obviates the need for simultaneous and cooperative binding [45].

A prevailing problem affecting the value of quantitative model predictions is concerned with the non-identifiability of the model parameters by the *in vivo* data (binding and dissociation constants as well as enzymatic rate constants). As a consequence the parameter uncertainties are propagated to the model predictions, which limits their significance [46]. To overcome these limitations further development in the model-driven analysis of chromatin-associated processes will depend on the integration of model calibration, uncertainty analysis and experimental design.

1.3. Research objectives

In this study, we want to expand the theoretical analysis of the NER process and thereby increase our understanding of the interrelation between protein-complex assembly dynamics and the functionality of chromatin-associated machineries. This analysis will be guided by the following questions:

- What is the quantitative relation between the 'recruitment-reaction' design principle of functional repair complexes and the overall repair kinetic of the DNA lesions?
- Which pathway components contribute to the control of the DNA repair rate?
- Which molecular mechanisms are responsible for the robust regulation of DNA repair?

To establish a mathematical model for which the model parameters are identifiable within narrow confidence intervals, we first generated accurate time-resolved measurements of DNA repair synthesis in single cells augmenting a comprehensive data set including the protein dynamics for all core NER factors. The measured repair-synthesis time course follows a slow first-order kinetic, which proved compatible with the current understanding of rapid reversible protein assembly. Based on this data, we were able to identify a realistic model for NER and thus make quantitative predictions about the repair process. One particularly captivating model outcome contradicts the existence of a rate-limiting repair factor but instead emphasises a shared and moderate control of the repair rate by all components. Exploiting the natural variability in repair factor expression, we could confirm this prediction experimentally. Intriguingly, protein expression of the repair factors is correlated independent of the amount of ongoing repair and the cell cycle stage. Thus, our results suggest that the robustness of DNA repair is regulated by a combination of collective

NER factor rate control and a so far unknown mechanism regulating NER protein expression.

This work has been conceived and carried out in close collaboration with Dr. Paul Verbruggen and Prof. Emer. Roel van Driel (University of Amsterdam).

2. Fluorescence time-lapse imaging of the NER process

Moné and colleagues (2001) advanced the research on nucleotide excision repair (NER) essentially by developing a technique to UV-irradiate a cell's nucleus within a locally confined area [8]. Until then, the *in-vivo* analysis of NER was limited to measurements of enzyme mobility and exchange dynamics at steady state performed with photobleaching experiments after irradiation of the whole nucleus [19, 47, 48]. In contrast, the enclosed amount of single-stranded DNA damages causes the local activation of a multi-protein repair machinery, which in turn allows us to study the *de-novo* assembly and dissociation kinetics of individual repair factors. Exploiting this standard experimental set-up, Luijsterburg *et al.* collected a comprehensive time-series dataset comprising the accumulation and dissociation of seven individual repair enzymes.

In the following chapter we will introduce this data together with the experimental methods, which represent the basis for our quantitative analysis of the NER process. We were able to augment the measurements with a direct readout for newly-synthesized DNA following the incorporation of the fluorescently modified nucleotide analogue 5-ethynyl-2'-deoxyuridine (EdU). It turns out that the DNA repair process is essentially a slow first-order reaction with a half-life of 1.2 hours.

The experiments reported in this chapter were planed and conducted by Martijn Luijsterburg (accumulation and fluorescence loss in photobleaching (FLIP) experiments; Leiden University) and Paul Verbruggen (EdU incorporation; University of Amsterdam).

2.1. Locally UV-inflicted 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) [8], in which NER-competent cells are covered with a polycarbonate ultraviolet (UV) filter containing pores with a given diameter (*cf.* Figure 2.1A). For our purposes, cells were grown on uncoated 24 mm coverslips before being overlaid with a mask with pores of 5 µm diameter [49]. Figure 2.1B illustrates the realistic deviation of pore diameters due to manufacturing inaccuracies. The expected size deviation has a very small coefficient of variation (*CV*) of 0.14 indicating a negligible effect on the filter's transmissivity. Filter-covered cells were immediately irradiated with a dose of 100 J/m² of UV-C (we used a germicidal UV-C light source with peak

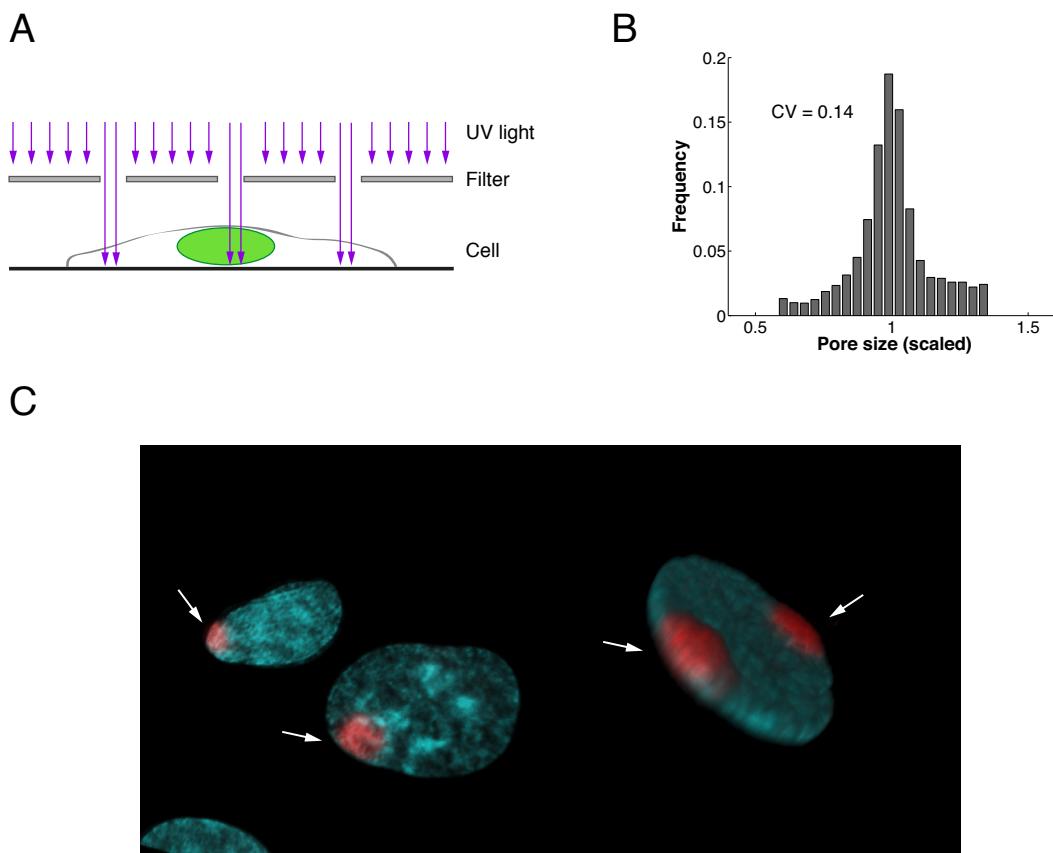


Figure 2.1.: Local irradiation leads to spatially confined DNA damage. A) Schematic illustration of the experimental set-up. UV-light (purple arrows) is transmitted through a UV filter containing pores with 5 μm diameter. Local irradiation of the chromatin occurs if a pore is located above a cell nucleus. (reprinted Figure from [50]) B) Distribution of the pore sizes within the UV irradiation mask ($n=5008$) 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. (Experiments by P. Verbruggen)

output at 254nm) with a fluency of 3.85 W/m². Due to the specific pore density (4×10^5 pores/cm²), we could observe nuclei containing only either one damage spot or no spot at all. Two spots per cell were only observed very rarely (cf. Figure 2.1C, white arrows indicate the damage spot). With this technique we were able to measure DNA repair in damaged nuclei as well as undamaged control nuclei in the same experiment.

2.2. 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 for the study of accumulation and exchange behaviour of fluorescently-tagged repair proteins under dif-

2. Fluorescence time-lapse imaging of the NER process

ferent experimental conditions. In the following, we describe the comprehensive dataset acquired by Luijsterburg *et al.* (2010) [6], which represents the basis for our quantitative analysis of the NER process. In total, the kinetics of seven repair factors were measured: a) XPC, the lesion recognition factor, b) TFIIH, the helicase responsible for DNA unwinding, c) XPA and RPA, which bind and thereby protect single stranded DNA against cleavage, d) the exonucleases XPF/ERCC1 and XPG performing the incision of the damaged DNA strand and, e) PCNA, which loads the DNA polymerase and hence indirectly provides insight into the DNA repair-synthesis kinetics.

Each repair factor was exogenously tagged with a green fluorescent protein (GFP) (or its 'enhanced' derivative eGFP) and expressed at physiological levels within the cell nucleus. Before the cells have been UV-irradiated at time $t = 0$, the repair proteins were homogeneously distributed at steady state. Immediately after UV-irradiation, the repair factors accumulated at the damaged chromatin sites, which led 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 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_{LD} , I_{outspot} , I_{nucleus} and $I_{\text{background}}$ represent the average fluorescence intensity within the locally damaged spot (LD), an equally sized area in the non-damaged nucleus, the whole nucleus and the background, respectively (*cf.* Figure 2.2B). A_{LD} and A_{nucleus} give the size of the damaged area and the size of the nucleus. Finally, the concentrations of accumulated protein are calculated assuming a damaged nuclear volume of 0.3 pL [6].

Protein	Concentration	Bound fraction
XPC	0.140 μM	13%
TFIIH	0.360 μM	10%
XPG	0.440 μM	9%
XPA	1.110 μM	7%
XPF/ERCC1	0.170 μM	7%
RPA	1.110 μM	15%
PCNA	1.110 μM	20%

Table 2.1.: Nuclear concentrations of NER factors (in μM). All nuclear quantities are based on published data or on previous estimates [48, 51, 52]. The nuclear concentrations were taken from Luijsterburg *et. al.* [6] were a nuclear volume of 0.3 pL was assumed. The bound fractions were determined by Eqn. 2.1.

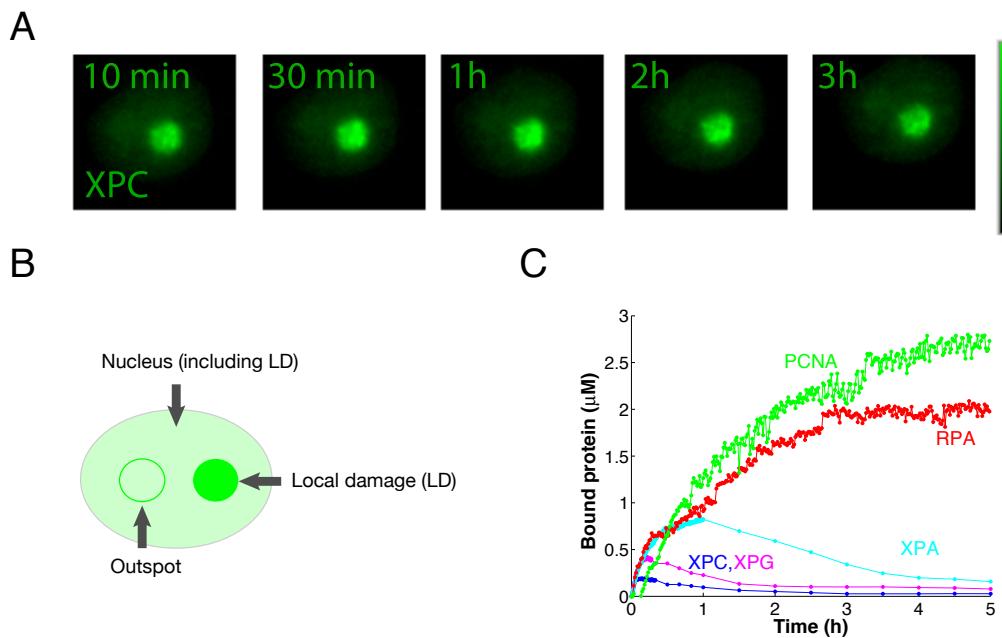


Figure 2.2.: Fluorescently labelled NER factors accumulate at locally irradiated nucleus. A) XPC-eGFP stably expressed in XP-C cells after local irradiation with UV-C (100 Jm^{-2} through 5- μm -diameter pores). XPC-eGFP accumulates at sites of local DNA damage. B) Scheme of a locally UV-irradiated nucleus.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

During the timespan of DNA repair, NER factor concentrations rise at the sites of local damage and then gradually decrease from their plateau levels at different rates (*cf.* Figure 2.2C). For example, the half-time $t_{1/2}$ for XPC- and XPG-eGFP is $\sim 1 \text{ h}$, whereas XPA-eGFP has a longer half-time of $\sim 2.5 \text{ h}$ [6]. In contrast, PCNA and RPA stay present in the damage spot even after lesion removal. These results 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.* Figure 2.3A, white rectangle). At the same time, fluorescent proteins are probed at low laser power (to prevent photobleaching) elsewhere within the nucleus. Repair proteins dissociating from the local damage spot have a high probability of being bleached before rebinding due to their large diffusivity. Accordingly, binding of the repair proteins seems to be rate limiting for the dwell time, not diffusion [6].

2. Fluorescence time-lapse imaging of the NER process

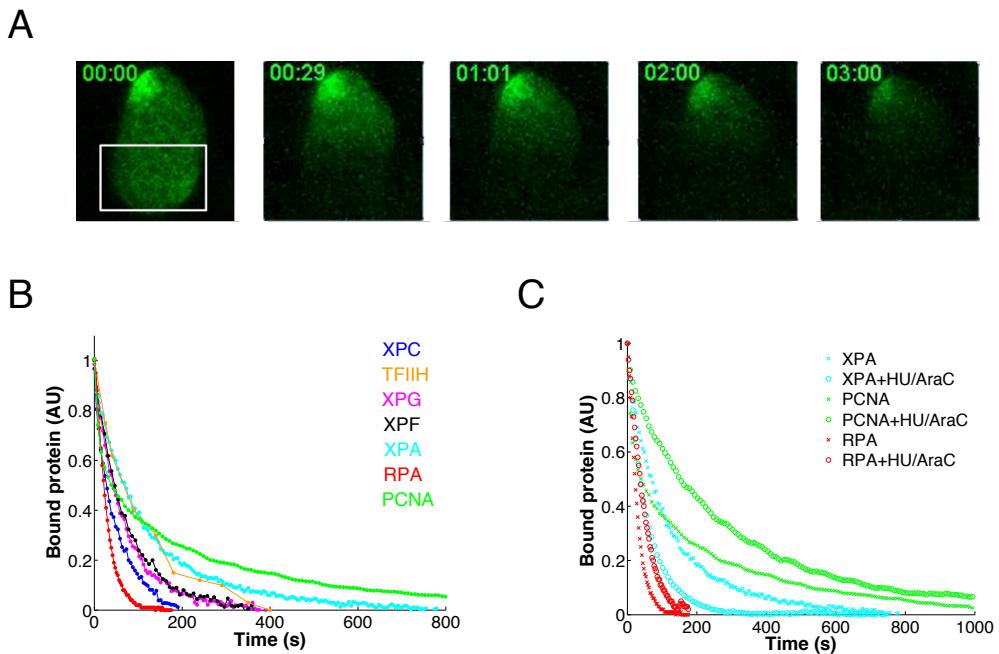


Figure 2.3.: Rapid dissociation of NER factors from damaged DNA. A) FLIP experiment in patient XP2OS cells stably expressing eGFP-XPA after 2 h of local irradiation. The nucleus is continuously bleached in an undamaged region (white rectangle). Fluorescence loss is recorded at the sites of local UV-irradiation B) Amount of bound NER-factors (XPC-eGFP, TFIIH-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. (Experiments by M. Luijsterburg)

However, compared to the long timespan (in the order of hours), where repair protein levels are still elevated at the 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 (TFIIH, 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, DNA resynthesis was stalled by addition of hydroxyurea (HU) and cytosine- β -arabinofuranoside (AraC). NER factors, which reached their plateau levels earlier, were not affected [6], so omitting the repair progression at this late stage slowed down the dissociation of PCNA and RPA (*cf.* Figure 2.3C). In contrast, XPA's half-life decreased by two fold indicating its higher affinity to repair synthesis intermediates. This demonstrates 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.

2.3. 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 [49]. DNA resynthesis reflects the kinetics of the post-incision repair process, complementing the measurement of DNA lesion removal, which captures the system's behaviour of the pre-incision steps (*cf.* section 1.2). To mark newly-incorporated DNA shortly before and after local UV irradiation, cells were incubated in microscopy medium supplemented with 10 μ M 5-ethynyl-2'-deoxyuridine (EdU). Due to its excessive presence in solution, the DNA polymerase integrates EdU (a thymidine analogue) instead of the endogenous thymidine into the new DNA strand (*cf.* Figure 2.4A). After incubation for the desired time, cells were fixed, stopping EdU incorporation, and subsequently permeabilized. EdU was then tagged with the fluorescent azide (AlexaFluor-555, Life Technologies) forming a covalent bond by click chemistry [53]. 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 and thereupon resynthesized 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 and constitutively expressed XPC-eGFP (*cf.* Figure 2.4B, lower row). Replicating cells could be excluded from the analysis due to the prevalent EdU incorporation distributed over the entire nucleus.

2.4. Repair rate follows first-order rate kinetics

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 2.3, 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 kinetics 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 hours, which coincides with the removal of 6-4PP (*cf.* [6] and Figure 2.5B). This agrees with the observation that NER is not primarily engaged in the removal of cyclobutane pyrimidine dimers (CPD) which are repaired

2. Fluorescence time-lapse imaging of the NER process

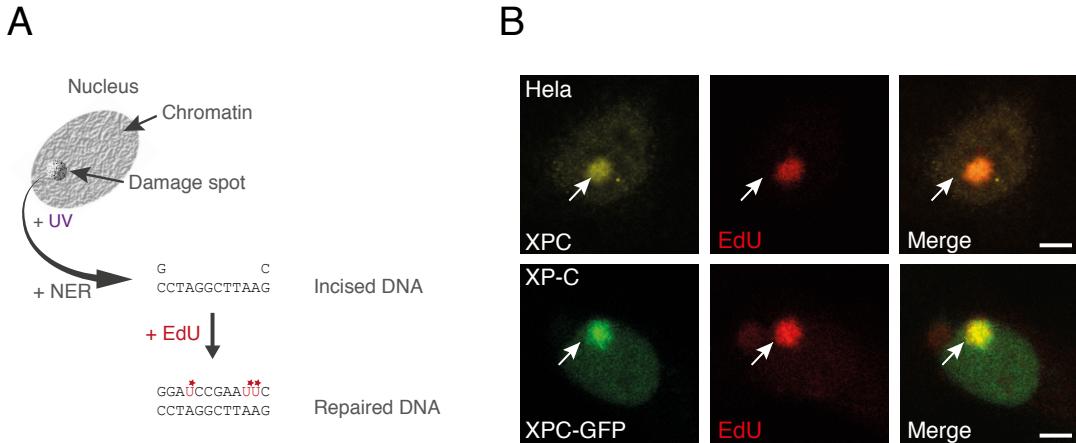


Figure 2.4.: EdU is incorporated at sites of local DNA damage. A) Schematic illustration depicting the experimental procedure of EdU incorporation at sites of local damage. B) Antibody-stained endogenous XPC in HeLa cells (upper panel) or stably expressed XPC-eGFP in XP-C cells (lower panel) accumulate in the LD spot and co-localize with incorporated EdU. (Experiments by P. Verbrugge)

on a much longer time scale [6]. 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). We observed that the amount of EdU incorporation per time interval (EdU rate) 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 [6,36,54]. Taken together, these data establish EdU incorporation as a direct and quantitative measure for DNA repair synthesis in locally damaged nuclei.

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} give the amount of incorporated EdU at time $t \in \mathbb{R}_{\geq 0}$ and its value at saturation, respectively. The time constant was determined as $\lambda=0.58 (\pm 0.07) \text{ h}^{-1}$ using the polyfit function (pre-implemented in MATLAB). 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.

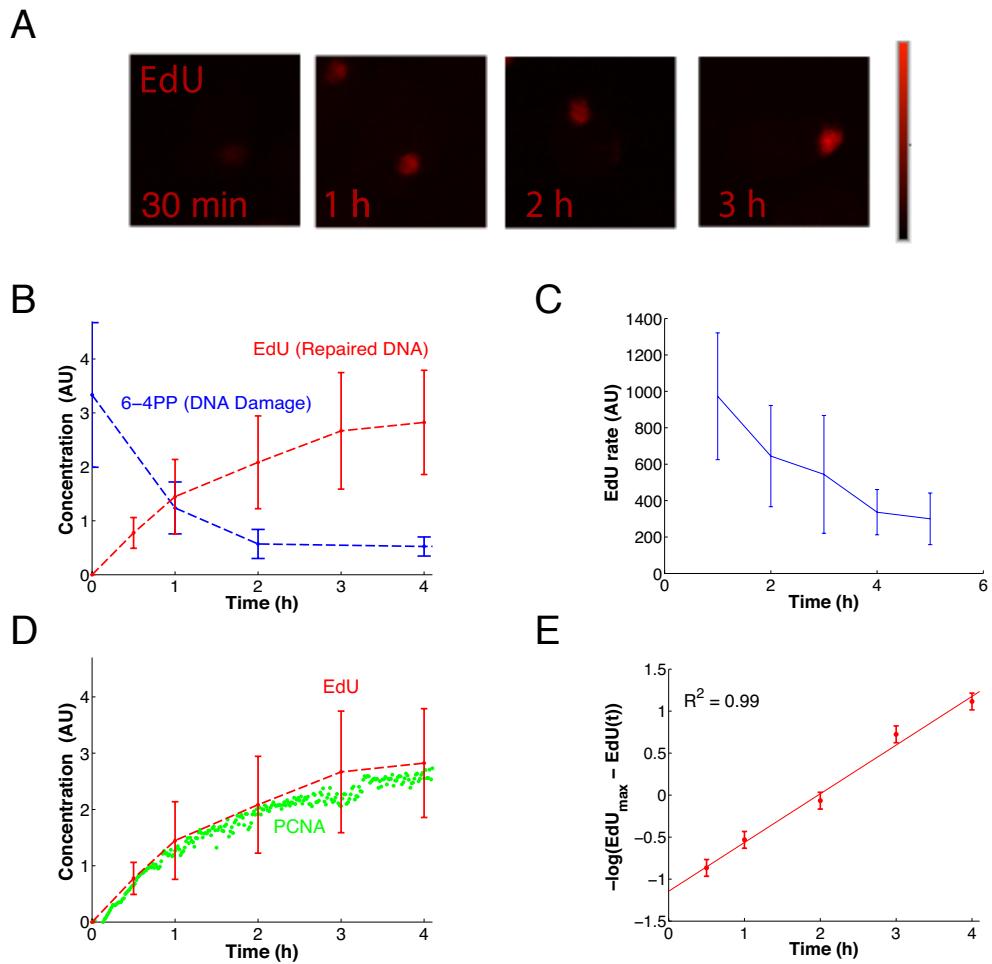


Figure 2.5.: EdU incorporation reflects DNA repair synthesis quantitatively. A) EdU signal at various time points after local UV-C irradiation. B) Repair DNA synthesis (EdU, red curve) coincides with damage removal (6-4PP, blue curve) measured previously by Luijsterburg *et al.* (2010) [6]. Graphs display the means \pm SD, of $n=50\text{-}70$ locally damaged cells per time point for damaged DNA and $n=150$ locally damaged cells (derived from three independent experiments) per time point for repaired DNA 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 represents mean \pm SD of EdU distributions derived from 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. (Experiments by P. Verbruggen and M. Luijsterburg)

Taken together, our experimental insight of the NER process got extended by following the kinetics of a repair intermediate directly. Now we are able to integrate our knowledge about the dynamic behaviour of individual repair factors and study its association with the overall repair rate quantitatively.

3. Kinetic model of NER

Luijsterburg *et al.* developed a mathematical NER model by integrating a large data set for all seven core NER factors [6]. However, despite the comprehensive collection of time-resolved *in vivo* measurements of the individual protein dynamics the model parameters comprising binding, dissociation and catalytic constants were not fully identified by the data. This is presumably caused by the missing readout of DNA repair, which was so far not measured on a similar time-scale as the protein assembly rates and the dwell times. Unfortunately, the non-identifiability sets a limitation for the predictive power of the model. Based on the EdU incorporation data introduced in chapter 2 we present here an evolved mathematical model of NER. It explains the connection between the dynamic exchange of individual repair factors at damaged chromatin sites and the overall time scale of the entire repair process. We calibrated the parametrization of the model iteratively with a profile likelihood estimation leading to identifiable parameters and hence an increased predictive power of the model. As a consequence, our fully identifiable model can reliably predict the behaviour of the experimentally inaccessible repair intermediate of incised DNA.

3.1. 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, Dr. Gesa Terstiege supervised by Prof. Thomas Höfer performed this analysis [49, 50], considering the complex formation with a simplified model of repair including N repair factors. They examined different scenarios distinguishing between random and sequential protein assembly; reversible and irreversible protein binding, or mixed mechanisms specific for each DNA repair intermediate (*cf.* Figure 3.1A, [49]). Under appropriate assumptions, the mean repair time of such a process can be generally described with:

$$\tau = \underbrace{\frac{1}{k} \sum_{i=0}^{N-1} K_i \left(\frac{l}{k}\right)^i}_{\text{first assembly}} + \underbrace{\frac{1}{\rho} \sum_{i=0}^N J_i \left(\frac{l}{k}\right)^i}_{\text{reassembly and reaction}}, \quad (3.1)$$

where k denotes the pseudo first-order association rate constant of NER factors, l the dissociation rate constant of NER factors and ρ the rate of the repair reaction. K_i and J_i differ

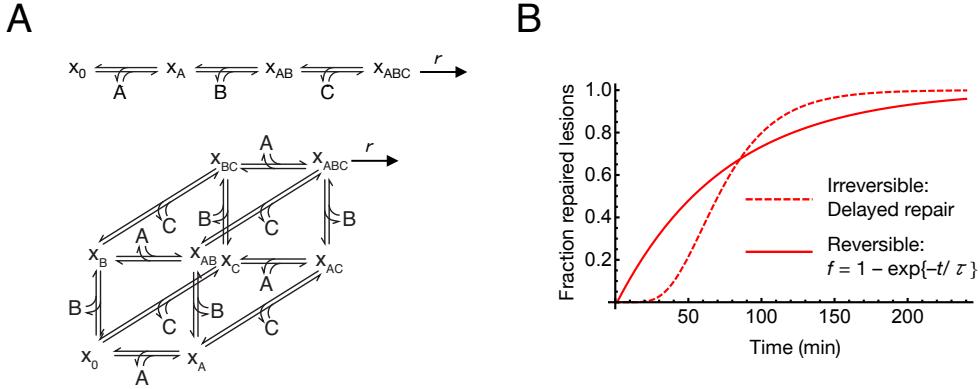


Figure 3.1.: Simplified analytical model of the repair complex formation. A) Sequential (upper panel) and random (lower panel) assembly scheme of the repair factors A, B and C to the DNA template x_i , $i \in \{A, B, C, AB, AC, BC, ABC\}$. x_0 denotes the empty lesion (e.g. damaged DNA). When the complex 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 τ , 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). (Simulations by G. Terstiege [50])

between random assembly

$$K_i^{\text{rand}} = \sum_{j=1}^{N-i} \frac{1}{i+j} \frac{\binom{N}{j-1}}{\binom{N}{i+j}}, \quad J_i^{\text{rand}} = \binom{N}{i} \quad (3.2)$$

and sequential assembly

$$K_i^{\text{seq}} = N - i, \quad J_i^{\text{seq}} = 1. \quad (3.3)$$

Considering reversible protein binding ($l = k = 1$), sequential and random assembly occur on a similar time scale for fewer than or equal to 10 components [50], which coincides with the number of core NER factors [6]. The theoretical component number where both assembly strategies have a similar timing is even higher for increasing reaction rates ($\rho \gg k, l$). Given that all repair factor dwell times are in the order of $\sim 1 \text{ min}$ ($k = l = 1 \text{ min}^{-1}$), the model was simulated for $N = 9$ components. The resulting trajectory followed a single-exponential kinetic with a half-time of $\sim 1 \text{ hour}$ in the case of balanced reversibility. In contrast, for an irreversible process ($l/k = 0$) the repair kinetic resembles a sigmoid time course (cf. Figure 3.1B). 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. This leads to a large number of non-functional repair complexes before the catalytically active complex is assembled.

3.2. Model structure and parametrization

To examine the relation between rapid repair factor exchange and the slow first-order reaction kinetic, we extended the previously performed analysis [6, 50], by considering a more realistic NER model. Conceptually, the model developed here, follows on the model introduced by Luijsterburg *et al.* (2010) [6]. In their model NER factors bind transiently to DNA repair intermediates to form catalytic complexes that, if complete, perform the next repair step. The latter are usually irreversible reactions embodying the sequential characteristics of this pathway (*cf.* Figure 1.1A).

The nature of these distinguishable repair intermediates has been widely investigated *in vitro* and *in vivo* [29, 39, 55, 56]. The evolved model, presented here, distinguishes five DNA repair intermediates: damaged DNA, unwound DNA, incised DNA, resynthesised DNA and rechromatinised DNA. The molecular state of these intermediates defines the binding affinity for a specific set of repair proteins (*cf.* Figure 3.2). Table 3.1 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 indicated that diffusion is not rate limiting for protein binding and therefore, we do not include it in the model (*cf.* Section 2.2 and [42, 43]).

Theoretical results suggest that the detection of damaged DNA sites by a single element instead of multiple elements simultaneously is crucial for the efficient initiation of the repair process [18, 57]. This prominent role within the NER pathway plays the lesion recognition factor XPC. Besides this experimentally verified exception, we assumed random and non-cooperative binding for all NER factors.

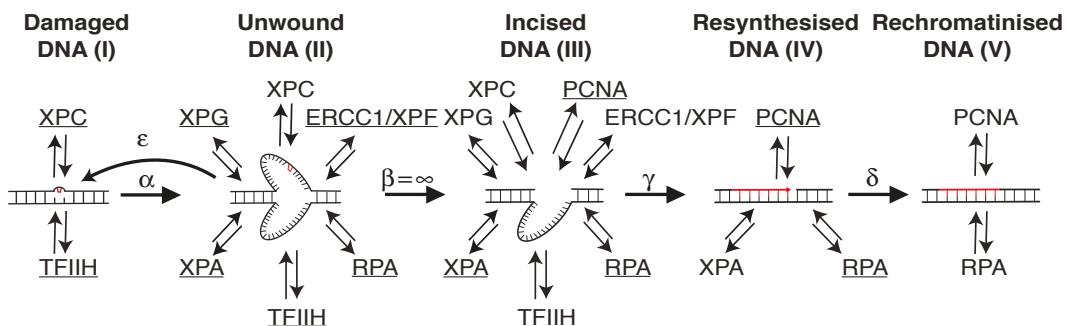


Figure 3.2.: Schematic model description of the DNA repair mechanism. The model distinguishes five individual repair intermediates: Damaged (I), unwound (II), incised (III), resynthesised (IV) and rechromatinised DNA (V). As indicated, specific tuples of repair proteins bind reversibly to the intermediates. Catalytic reactions, denoted by Greek letters, occur when the catalytic reaction complex (underlined proteins) is assembled.

Repair intermediate	Binding proteins	Catalysed process Required proteins	Remarks	Ref.
(I) Damaged DNA	XPC,TFIIH (3 states)	Unwinding (reaction α) XPC and TFIIH	Initiation by binding of XPC and subsequent recruitment of TFIIH.	[55] [28] [27] [43] [18]
(II) Unwound DNA	XPC,TFIIH, XPG, XPA, ERCC1/XPF, RPA (64 states)	Dual incision (reaction β) TFIIH,XPG, XPA, RPA and ERCC1/XPF	If the DNA becomes devoid of any protein, it will re-anneal (reaction ε). Dual incision requires the endonucleases XPG and ERCC1/XPF and is stimulated by TFIIH, XPA, RPA and possibly XPC.	[55] [58] [59] [60] [32]
(III) Incised DNA	XPC,TFIIH, XPG, XPA, ERCC1/XPF, RPA, PCNA (128 states)	Repair-synthesis (reaction γ) 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).	[55] [60]
(IV) Resynthesised DNA	XPA, RPA, PCNA (8 states)	Rechroma- tinization (reaction δ) RPA and PCNA	Accumulation data imply that XPA binds to repaired DNA while the pre-incision proteins do not (Figure 2.2C).	[61] [62] [6]
(V) Rechromatin- ised DNA	RPA, PCNA (4 states)		RPA and PCNA associate with repaired intermediate I, as levels of bound eGFP- PCNA and eGFP-RPA are high up to at least 4 h after UV irradiation, while other repair proteins are no longer bound.	[28] [6]

Table 3.1.: Model assumptions. Adapted from Terstiege *et al.* (2010) [50]

3. Kinetic model of NER

The model structure introduced above and shown in Figure 3.2 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_π^R (cf. Eqn. 3.4) associated with a specific repair intermediate $R = I, II, \dots, V$ (damaged (I), unwound (II), incised (III), resynthesised (IV), rechromatinised (V)). The index π 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), TFIH (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 3.1) with: $\pi_{R=I} = \{C, T\}$; $\pi_{R=II} = \{C, T, G, A, F, R\}$; $\pi_{R=III} = \{C, T, G, A, F, R, P\}$; $\pi_{R=IV} = \{A, R, P\}$; $\pi_{R=V} = \{R, P\}$. The time course of this model is described by

$$\frac{d}{dt} y_\pi^R = \sum_{p \text{ in } R} \eta \left((-1)^{\pi(p)} l_p^R y_\pi^R |_{\pi(p)=1} + (-1)^{1+\pi(p)} k_p^R C_p(t) y_\pi^R |_{\pi(p)=0} \right) + E(y_\pi^R), \quad (3.4)$$

where η represents a cooperativity ensuring the inclusion of all cooperative binding events:

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

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 ordinary differential equations for $p \in \{C, T, G, A, F, R, P\}$:

$$\frac{d}{dt} C_p = \sum_{R=I}^V \sum_{\pi} \xi (\delta_{\pi(p)1} l_p^R y_\pi^R - \delta_{\pi(p)0} k_p^R C_p y_\pi^R) \quad (3.5)$$

Analogous to η , ξ governs the sequential binding of XPC and TFIH by:

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

The Kronecker delta

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

ensures that a protein only binds to complexes, where it is yet missing and only leaves complexes, where it is actually present.

Finally, when an enzymatic complex has fully assembled at the DNA template, it catalyses the next repair step, which is represented in the model by the term $E(y_\pi^R)$ in Eqn. 3.4. After damage recognition by XPC, damaged DNA is unwound by the helicase TFIIH with the unwinding activity α , 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 ($R = I$):

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

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

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

Once the lesion strand is excised with incision rate β , the remaining repair steps are irreversible. Incised DNA is resynthesised with the rate γ 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

3. Kinetic model of NER

RPA but not XPC and TFIIH [37]. Evidence for the importance of PCNA and RPA for the resynthesis reaction was shown by Shivji *et al.* (1995) [62], and therefore we incorporate the following catalytic reactions for incised DNA ($R = \text{III}$):

$$E(y_{1111111}^{\text{III}}) = \beta y_{011111}^{\text{II}} \quad \text{and} \quad E(y_{0001011}^{\text{III}}) = -\gamma y_{0001011}^{\text{III}}.$$

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 ($R = \text{IV}$):

$$\begin{aligned} E(y_{111}^{\text{IV}}) &= \gamma y_{0001011}^{\text{III}}, \\ E(y_{011}^{\text{IV}}) &= -\delta y_{011}^{\text{IV}} \quad \text{and} \\ E(y_{11}^{\text{V}}) &= \delta y_{011}^{\text{IV}}. \end{aligned}$$

Due to the non-cooperativity assumption of all NER factors, there are a total of 2^N repair states where N denotes the number of repair proteins assembling to the particular repair intermediate. The only exception is the sequential assembly of TFIIH, after lesion detection by XPC, which reduces the number of states for damaged DNA to $2^2 - 1 = 3$ states. For the remaining repair intermediates we derive $2^6 = 64$ states for unwound DNA; $2^7 = 128$ states for incised DNA; $2^3 = 8$ states for resynthesised DNA and $2^2 = 4$ 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. The initial conditions are the measured free protein concentrations denoted in Table 2.1 [6,50] and the initial amount of inflicted damages was estimated with $y_{00}^1 = 3.33 \mu\text{M}$ [49]. The remaining initial states were assumed to be zero. To reproduce the FLIP kinetic for a specific repair factor, all corresponding dissociation constants were set to zero when the FLIP experiment started (at the time where protein accumulation at LD reached the plateau level). Accordingly, FLIP kinetics were acquired after 600 s for XPC and ERCCC1/XPF, 900 s for XPG and TFIIH, 2000 s for XPA and 7200 s for PCNA.

3.3. A maximum likelihood approach for efficient model fitting

To find a realistic parametrisation for the temporal development of the repair states y_π^R (cf. Eqn. 3.4) the model was mapped to m observables z_k via the function f_{z_k} :

$$z_k(t_i, \theta) = f_{z_k}(t_i, y(t_i, \theta), \theta). \quad (3.6)$$

The observables z_k are parametrised by θ and resemble experimentally-derived quantities at time t_i . θ depicts binding, dissociation and catalytic constants. Each model observable $z_k(t_i, \theta)$ corresponds to the measured data $z_k^\dagger(t_i)$ with intrinsic noise ϵ_{ki} . The data is usually the sum of measurement noise combined with the naturally occurring biological variability. Assuming additive, normally distributed noise leads to $z_k^\dagger(t_i) = z_k(t_i, \theta) + \epsilon_{ki}$ with $\epsilon_{ki} \sim N(0, \sigma_{ki}^2)$. To calibrate the measured data $z_k^\dagger(t_i)$ with the model observables $z_k(t_i, \theta) + \epsilon_{ki}$, we applied a maximum likelihood approach as distance measure which translates, considering normally distributed noise, into:

$$L(z_k^\dagger | \theta) = \prod_{k=1}^m \prod_{i=1}^{d_k} \frac{1}{\sqrt{2\pi\sigma_{ki}^2}} \exp \left(-\frac{1}{2\sigma_{ki}^2} (z_{ki}^\dagger(t_i) - z_k(t_i, \theta))^2 \right). \quad (3.7)$$

Here, d_k denotes the number of distinct experimental data-sets (t_i, z_k^\dagger) for each measured observable $k = 1, \dots, m$, measured at time points t_i with $i = 1, \dots, d_k$. σ_{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 multiplied with 2: $-2 \log(L(z_k^\dagger | \theta))$, which we will refer to in the following as χ^2 .

$$\chi_\theta^2 = -2 \log(L(z_k^\dagger | \theta)). \quad (3.8)$$

For the minimization of χ^2 the choice for θ is controlled by σ_{ki}^2 (cf. Eqn. 3.7). As shown by Raue and colleagues [46], for a reliable estimation of the model parameters θ , the simultaneous approximation of σ_{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 [46]. Accordingly σ_{ki}^2 can be considered as parametrised function

$$\sigma_k(t_i, \theta) = f_{\sigma_k}(t_i, z_k(t_i, \theta), \theta), \quad (3.9)$$

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 θ .

For reasons of fitting-speed efficiency we implemented our model into the online-available D2D software environment [46] 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. 3.7). 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 into the form of sensitivity equations

$$\frac{d}{dt} \frac{dy(t, \theta)}{d\theta} = \frac{\partial f_y}{\partial y} \frac{dy(t, \theta)}{d\theta} + \frac{\partial f_y}{\partial \theta}, \quad (3.10)$$

which represent additional ordinary differential equations (ODEs) that are solved in parallel to the original ODE system (*cf.* Eqn. 3.4 and Eqn. 3.5) [63]. Applying sensitivity equations instead of a simple finite difference approximation proved to be numerically more accurate and computationally faster [46]. Both, model and sensitivity ODEs were solved with the CVODES algorithm written in ANSI standard C [64].

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) [65]. In contrast to a random sampling approach, LHS provides a better coverage of the sampling space by maximizing the distance between successive parameter drawings [46].

3.4. NER model fits accumulation, FLIP and repair synthesis measurements

To derive faithful fitting results we reiterated the optimization process 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 within one order of magnitude compared to the global χ^2 minimum (*cf.* Figure 3.3).

Compared to the average initial χ^2 -value, the average fit improvement was about eight orders of magnitude. For the optimal parameter set, the model fits the experimental data comprising accumulation, FLIP, perturbation and repair synthesis measurements (*cf.* Figure 3.4A, B, C; estimated measurement errors are shown as shaded area around the fit). The accumulation kinetics (*cf.* Figure 3.4A) are depicted as concentrations scaled by the volume of the locally damaged area, which were estimated with 10% of the nuclear volume (average derived

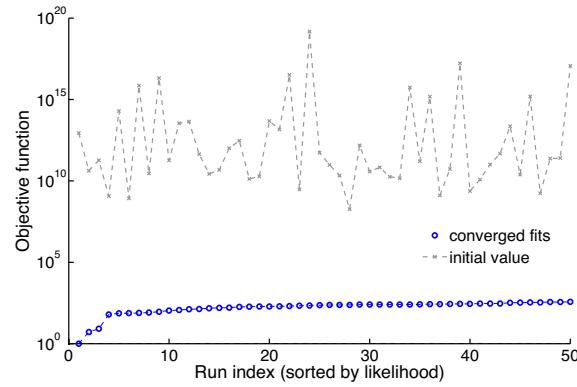


Figure 3.3.: Parameter estimation for the 50 best fits after initial Latin hypercube sampling. Visualization of performance using the 50 (out of 250) best independent optimization runs (blue dots depict final optima of each individual run). Initial guesses (grey stars) were generated by Latin hypercube sampling. For illustrative purposes the global optimum is centred to 1.

from segmentation results of the nucleus and the LD foci). We believe that this specification is more intuitively comprehensible compared to scaling by the whole nuclear volume as performed by Luijsterburg *et al.* [6]. It allows the realistic comparison between simulated and measured microscopy images of NER factor accumulation and EdU incorporation (*cf.* Figure 3.5).

3. Kinetic model of NER

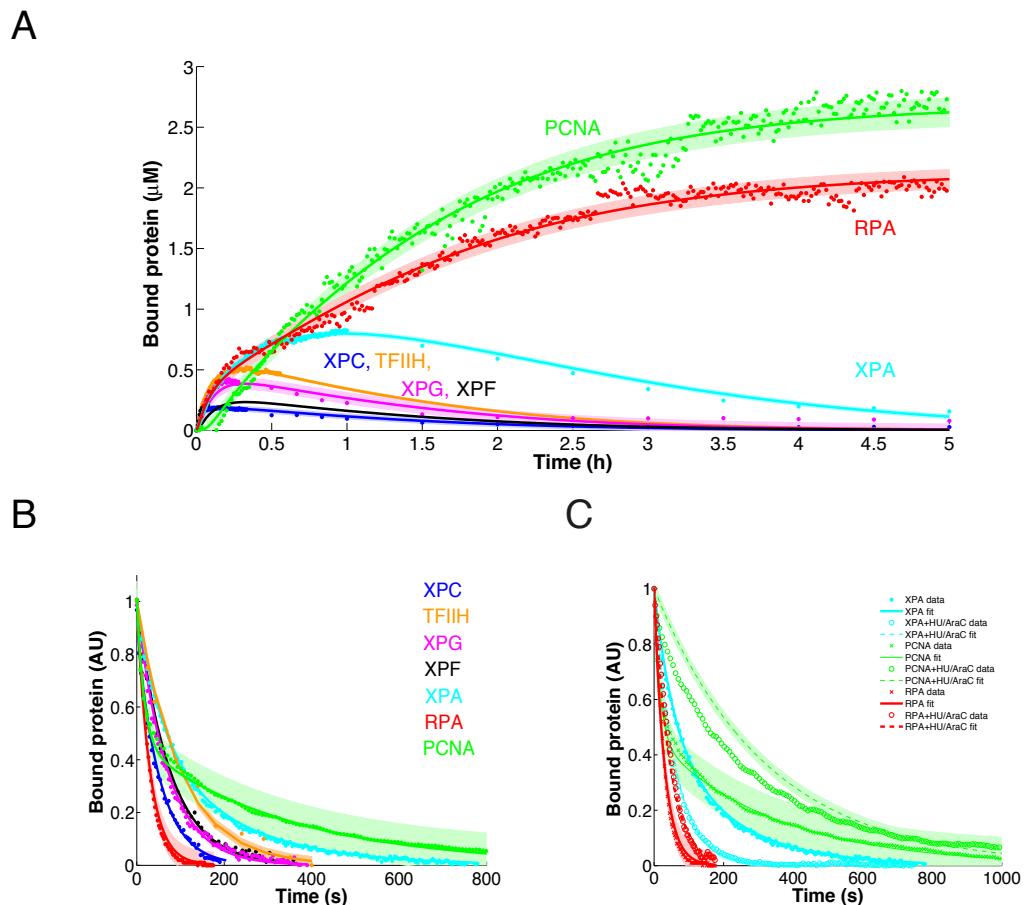


Figure 3.4.: Quantitative NER model fits to accumulation and dissociation time courses. A) and B) Simulation (lines) and measurement (dots) of the accumulation and dissociation kinetics for the repair factors XPC, TFIID, XPG, XPA, XPF, RPA and PCNA. Estimated errors are depicted as shaded area. C) Fitted FLIP time courses of XPA, RPA and PCNA in the absence or presence of AraC and HU on locally irradiated cells. (Experiments by M. Luijsterburg)

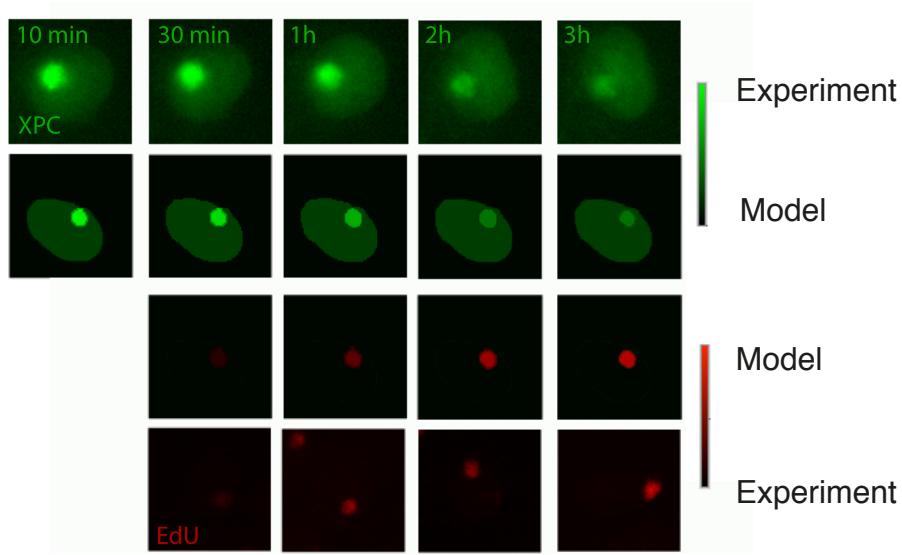


Figure 3.5.: Comparison between measured and simulated single cell microscopy images. Simulated time courses and the associated microscopy images for GFP tagged XPC (two upper rows) and EdU incorporation (two lower rows). Simulated XPC expression intensities were normalized to the nuclear intensities of the microscopy images. For the EdU incorporation intensities were scaled according to the highest and lowest intensity values of the microscopy images. (Experiments by P. Verbrugge)

3.5. Profile likelihood analysis identifies realistic model of NER

In this section we 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 by the limited amount and quality of the data, respectively [66,67]. Both can be analysed and tested numerically by a formalism called profile likelihood estimation (PLE) [68–70], where the multi-dimensional model uncertainty inflicted by an individual parameter is projected to a one-dimensional profile likelihood (PL)

$$PL(\theta_i) = \max_{\forall j \neq i} [L(z_k^\dagger | \theta_j)]. \quad (3.11)$$

One parameter θ_l with $l \in \{1, \dots, N\}$ at a time, is gradually fixed along this dimension for different values of p . In each step the negative log-likelihood $X_{\theta_l}^2(p)$ is minimized (*cf.* Eqn. 3.8), fitting all other parameters θ_k with $k \in \{1, \dots, N\}; k \neq l$. Subsequently, the identifiability of a parameter θ_l can be determined by $\Delta X_{\theta_l}^2(p)$, which describes the difference between

3. Kinetic model of NER

the parameter dependent local minimum $X_{\theta_l}^2(p)$ and the global χ^2 minimum

$$\begin{aligned}\Delta X_{\theta_l}^2(p) &= \min_{\{\theta_k | k=1, \dots, N; k \neq l\}} (\chi^2(\theta_1, \dots, \theta_{l-1}, p, \theta_{l+1}, \dots, \theta_N)) \\ &- \min_{\{\theta_k | k=1, \dots, N\}} (\chi^2(\theta_1, \dots, \theta_N)).\end{aligned}\quad (3.12)$$

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

$$CI_{1-a}^{\theta_l} = \{p | \Delta X_{\theta_l}^2(p) \leq Q_{\chi^2}(1 - a, 1)\}. \quad (3.13)$$

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 χ^2 -distribution quantile of $Q_{\chi^2}(95\%, 1) = 3.8$. A parameter θ_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 NER model (*cf.* Eqn. 3.4 and Eqn. 3.5) comprising 40 binding and dissociation parameters and 4 catalytic reaction constants. However, only under the assumption that the repair factor exchange at unwound and incised DNA is equal (*cf.* Figure 3.6 with the exception of PCNA) all binding and dissociation constants were identifiable. The same holds true for the dissociation kinetics of TFIIH from damaged and unwound DNA. This reduces the number of fitted parameters to 14 binding, 13 dissociation and 4 catalytic rate constants.

Besides the slow rate of rechromatinisation δ , presumably identifiable due to the slow decrease in accumulated XPA (*cf.* Figure 3.4), all catalytic rates are fast. This is seen by the existence of lower bounds on the rate constants of the order of 1 s^{-1} . For the numerical values of the parameters see Table A1 and Table A2.

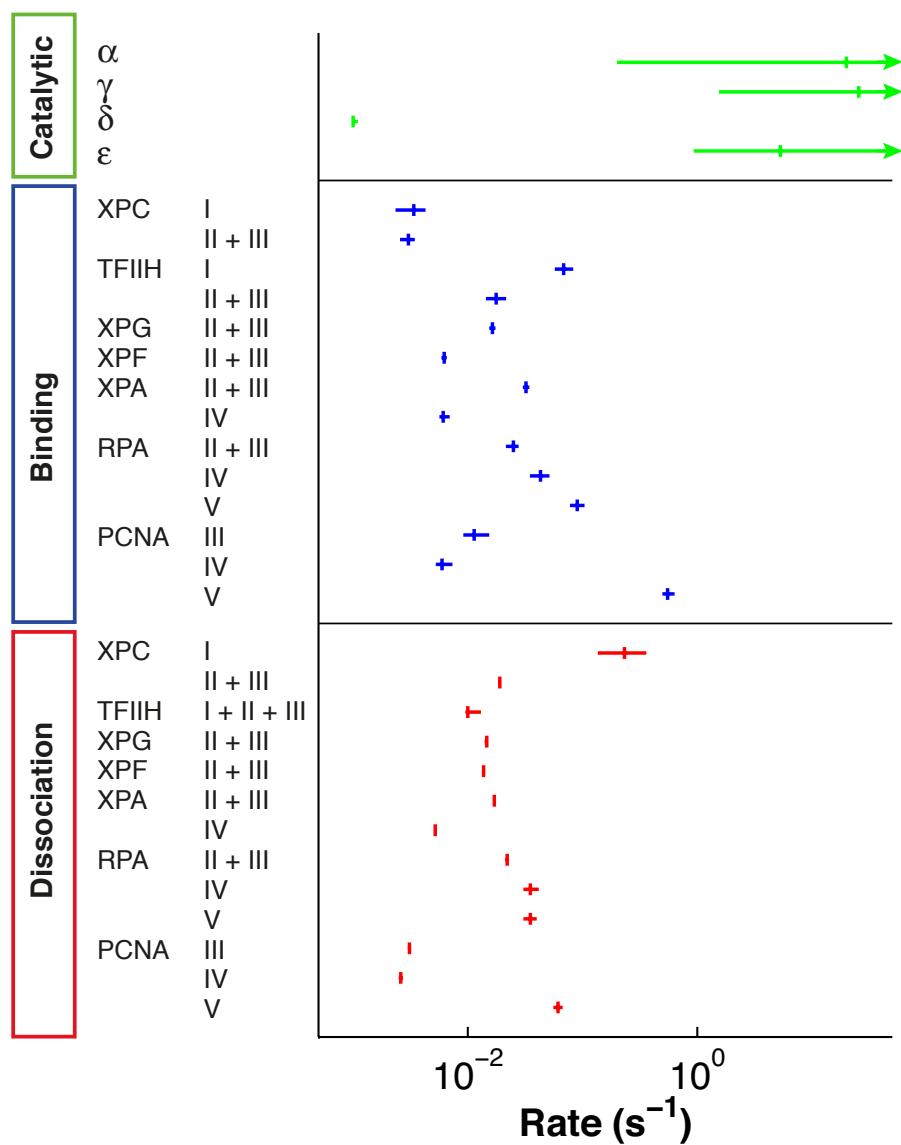


Figure 3.6.: Parametrisation identifies realistic model for NER. Mean values (vertical bars) and confidence intervals (horizontal bars) of catalytic (green), binding (blue) and dissociation constants (red) characterizing the dynamic assembly of NER factors at the successive repair intermediates (damaged DNA I, unwound DNA II, incised DNA III, resynthesised DNA IV, rechromatinised DNA V). Arrow heads indicate infinite confidence intervals.

Value	XPC	TFIIE	XPG	XPF	XPA	RPA	PCNA
Damaged DNA							
$K_d(\mu M)$	9.35 (3.46;16.09)	0.052 (0.044;0.072)	NA	NA	NA	NA	NA
Unwound DNA							
$K_d(\mu M)$	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
Incised DNA							
$K_d(\mu M)$	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)
Resynthesised DNA							
$K_d(\mu M)$	NA	NA	NA	NA	0.236 (0.222;0.27)	1.167 (0.924;1.521)	0.605 (0.531;0.747)
Rechromatinised DNA							
$K_d(\mu M)$	NA	NA	NA	NA	0.538 (0.438;0.645)	0.154 (0.134;0.182)	

Table 3.2.: 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 K_d values ($k_{\text{off}}/k_{\text{on}}$), depicting the protein binding affinities to the respective DNA repair intermediate fall into a physiological realistic range between ~ 100 nM and ~ 1 μM (cf. Table 3.2). Only XPC has a particular low affinity of 9 μM , which is consistent with previous findings reporting that the time until DNA incision is mainly determined by the slow lesion recognition [6].

Having an identifiable model allows us to make more precise computational predictions. For example, we can simulate the unobservable fraction of incised DNA (cf. Figure 3.7, green trajectory). As the EdU incorporation measurement shows, the damaged (blue) and repaired (red) DNA kinetics are tightly coupled. This omits a higher accumulation of incised DNA.

In summary, for the development of a realistic model of the DNA repair pathway we integrated live-cell imaging data of the assembly and dissociation kinetics of all core NER factors and linked it directly to the DNA repair kinetics. The extended dataset identifies a total of 31 model parameters including 14 binding, 13 dissociation and 4 catalytic rate constants. Therefore, quantitative predictions with the model on the functioning of NER are no longer limited.

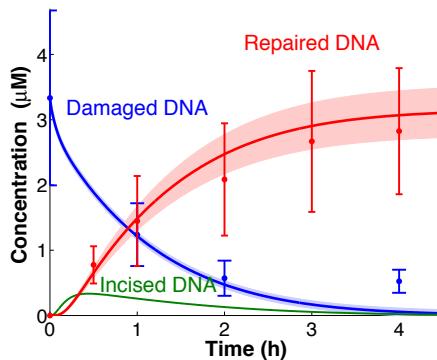


Figure 3.7.: Short delay between DNA damage removal and DNA repair synthesis omits accumulation of incised DNA. Experimental (dots with error bars) and simulated (lines) time courses for damaged DNA (blue) and DNA repair synthesis (red). Simulated trajectories depict repair intermediate I+II for damaged DNA and repair intermediate IV+V for repaired DNA (cf. Figure 3.2). Estimated errors are depicted as shaded area. Model prediction for incised DNA (green) constitutes from DNA repair intermediate III. (Experiments by P. Verbruggen)

4. Control analysis of the DNA repair rate

In chapter 2 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 with a half-time of ~ 1.2 h. Notably, measurements on the single cell level revealed also that the half-time and hence the rate of repair is highly variable (*cf.* Figure 2.5B-E). So far, not only is the origin of this heterogeneity unknown, but also how the cell faces this highly variable environment.

In this chapter we apply the established kinetic NER model and explore the nature of the repair rate variability. By 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 repair proteins we can experimentally corroborate the computationally-derived finding for the NER factors XPC, TFIIH, 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 appears 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.

4.1. Kinetic NER model predicts collective rate control

Using PLE analysis, we identified model parameters in reasonable biological ranges (*cf.* Section 3.5), which allows 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 3.1), we wanted to test whether this 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 [71, 72]. Accordingly, we can quantify the relative change in the repair rate ν as a function of the relative change in the protein concentration C_i ($i = \text{XPC, TFIIH, XPG, XPA, XPF, RPA, PCNA}$).

$$\tilde{R}_i = \frac{\partial \ln \nu}{\partial \ln C_i}. \quad (4.1)$$

The inverse of ν is the characteristic time τ_R that represents the time duration for the repair reaction R . Hence, Eqn. 4.1 can be rewritten as

$$\tilde{R}_i = \frac{C_i}{\tau_R^{-1}} \frac{\partial \tau_R^{-1}}{\partial C_i}. \quad (4.2)$$

τ_R , in turn, can be directly approximated from the distribution of repaired DNA states $y_\pi^R(t)$ (*cf.* Eqn. 3.4), 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

$$\tau_R = \frac{\mu^{(1)}}{\mu^{(0)}}, \quad (4.3)$$

with

$$\mu^{(m)} = \int_0^\infty t^m y_\pi^R(t) dt. \quad (4.4)$$

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 time $t = 0$.

$$\tau_{\text{syn}} = \frac{\int_0^\infty t (y_{00}^I(0) - (\sum_\pi y_\pi^{IV}(t) + \sum_\pi y_\pi^V(t))) dt}{\int_0^\infty y_{00}^I(0) - (\sum_\pi y_\pi^{IV}(t) + \sum_\pi y_\pi^V(t)) dt} \quad (4.5)$$

Using τ_{syn} in Eqn. 4.2 we derive the response coefficients for the repair synthesis rate, which are uniformly small with ~ 0.3 and below (*cf.* Figure 4.1). 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.* Figure A.2), with the characteristic time

$$\tau_{\text{inc}} = \frac{\int_0^\infty t \sum_{x=I}^{\text{III}} \sum_\pi y_\pi^x(t) dt}{\int_0^\infty \sum_{x=I}^{\text{III}} \sum_\pi y_\pi^x(t) dt}. \quad (4.6)$$

Prediction profile likelihood To obtain realistic confidence bounds for the response coefficients we perform a prediction profile likelihood estimation [73, 74]. The predicted response coefficients $\tilde{R} = F(D_{\text{pred}}, \theta)$ are considered as model outcomes F for a predicted experimental design $D_{\text{pred}} = (z_{\text{pred}}, t_{\text{pred}}, u_{\text{pred}})$. D_{pred} specifies a prediction observable z_{pred} at time point t_{pred} given the externally controlled stimulation u_{pred} . As defined in Eqn. 3.6 $z_{\text{pred}}(t_{\text{pred}}, \theta)$ comprises a model simulation that can be mapped to experimentally observ-

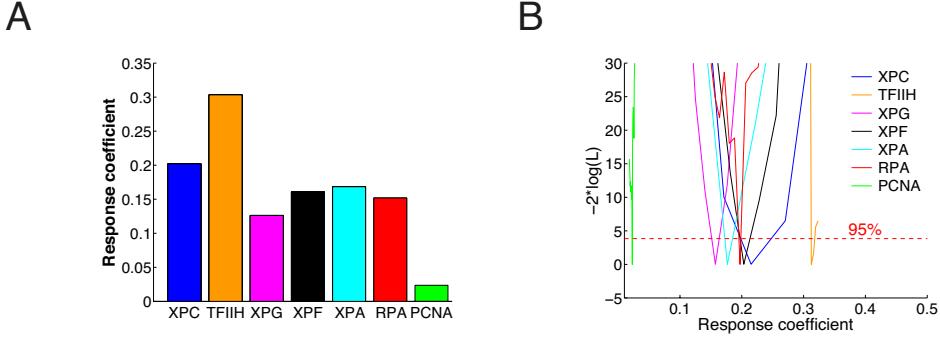


Figure 4.1.: Collective control of the repair rate. A) Response coefficients for seven repair factors (XPC, TFIIDH, XPG, XPA, XPF, RPA, PCNA) are small and uniformly distributed. B) Prediction profile likelihood indicate small prediction confidence bounds for each response coefficient. 95% threshold is given by the χ^2 distribution with one degree of freedom.

able quantities. Analogous to Eqn. 3.11, the prediction profile likelihood

$$PPL(\tilde{R}) = \max_{\theta^* \in \{\theta | F(D_{\text{pred}}, \theta) = \tilde{R}\}} L(z_k^\dagger | \theta^*, \tilde{r}) \quad (4.7)$$

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

$$\Delta X_{\theta_l}^2(\tilde{r}) = \min_{\{\theta, \tilde{r} \in \tilde{R}\}} (\chi^2(\theta, \tilde{r})) - \min_{\theta} (\chi^2(\theta)), \quad (4.8)$$

which describes the difference between the global χ^2 minimum and the best fit with r included into the objective function. Similar to Eqn. 3.13 we can determine prediction confidence bounds

$$PCI_{1-a}^{D_{\text{pred}}, z_k^\dagger} = \{\tilde{r} | \Delta X_{\theta_l}^2(\tilde{r}) \leq Q_{\chi^2}(1 - a, 1)\}, \quad (4.9)$$

which include the set of predictions $\tilde{R} = F(D_{\text{pred}}, \theta)$ for which $-2 \log(PPL)$ is below the threshold given by the χ^2 -distribution. The PPLs were computed within the d2d-framework [46] using the CVODE package [64] for the numerical integration of the ODEs together with the sensitivity equations (*cf.* Section 3.3). By applying the PPL analysis for each response coefficient we find that all of them are identifiable indicating how well the model predictions are determined by the data (*cf.* Figure 4.1B).

The moderate response predicted by the response coefficients also holds true for larger variations in the repair factor concentration (*cf.* Figure 4.2). The linear approximation (on which the response coefficients are based on) 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.

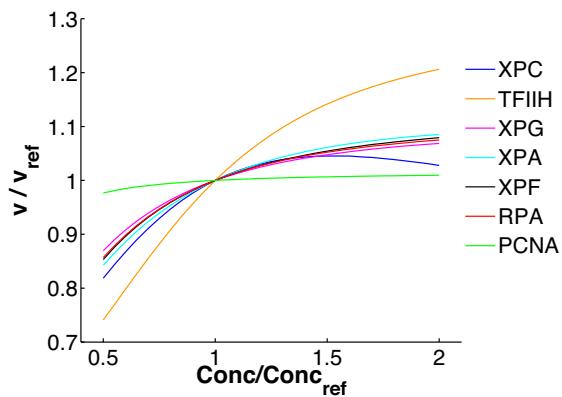


Figure 4.2.: Moderate repair rate response against natural NER factor expression variability. Repair rate ν as a function of concentration changes in individual repair factors. Simulated repair rates are normalised by the reference rates, which are predicted by the model when applying the experimentally measured NER factor concentrations Conc_{ref} .

4.2. Exploiting natural variability in protein expression to quantify rate control

To corroborate the model prediction of a distributed repair rate control (*cf.* Section 4.1), 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 despite the natural occurring variability in protein expression. By capturing the integrated nuclear fluorescence intensity of antibody-stained repair proteins we derive the expression values for XPC, TFIIH, XPA, XPF and RPA. (*cf.* Figure 2.2B).

4. Control analysis of the DNA repair rate

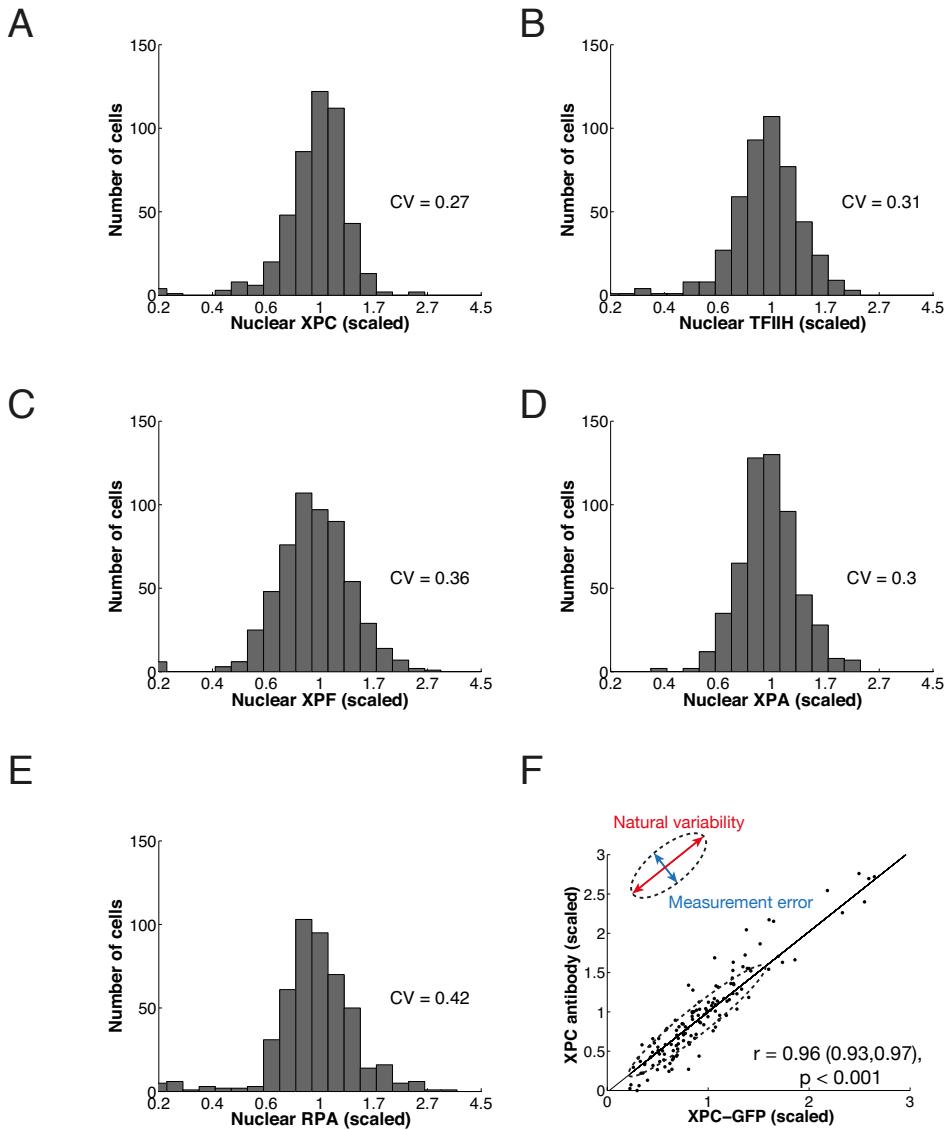


Figure 4.3.: Natural variability in nuclear NER factor expression is significantly larger than measurement error. A-E) Histograms of nuclear protein concentrations of the antibody stained NER factors XPC and TFIIH (n=470), XPF and XPA (n=565), RPA (n=487). F) Scatter plot of antibody stained XPC against XPC-eGFP stably expressed in XP-C XPC-eGFP cells. The dashed error-ellipse illustrates the proportion of natural variability and measurement noise. The black line represents the linear regression with correlation coefficient r and p-value. 95% confidence bounds of the correlation coefficients r were estimated by non-parametric bootstrap and are given in brackets (for further details cf. Section A.2). (Experiments by P. Verbruggen)

XPC	TFIILH	XPA	XPF	RPA
CV: 0.34 ± 0.05	0.33 ± 0.02	0.33 ± 0.03	0.4 ± 0.04	0.44

Table 4.1.: Mean and standard deviation of the CV in nuclear XPC, TFIILH, XPA and XPF expression. Distributions for nuclear protein expression were acquired in $n_b^{XPC} = 5$, $n_b^{TFIILH} = 3$, $n_b^{XPA} = 3$, $n_b^{XPF} = 3$ and $n_b^{RPA} = 2$ independent biological replicates. Within each measurement between $n = 250$ and $n = 572$ with an average of $n = 477$ cells were analysed.

For 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 ~0.37 (cf. Figure 4.3A-E). Table 4.1 gives the means and standard deviations from at least three biologically independent measurements of the expressed protein cell-to-cell variability.

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 human primary fibroblasts together with immunofluorescently labelled XPC in the same cell. Both quantities are strongly positively correlated (cf. Figure 4.3F) suggesting a large natural variability compared to much lower measurement noise. To quantitatively validate this observation, we performed a principal component analysis (PCA) [75]. 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 4.3F. Calculating the *CV* from the smaller component we determined a relative measurement error for the antibody labelling method of ~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 4.4A) sixty minutes after UV irradiation. Consequently, we conclude that the DNA lesions are not saturated, and thus, a higher NER factor concentration could potentially accelerate the repair rate.

4. Control analysis of the DNA repair rate

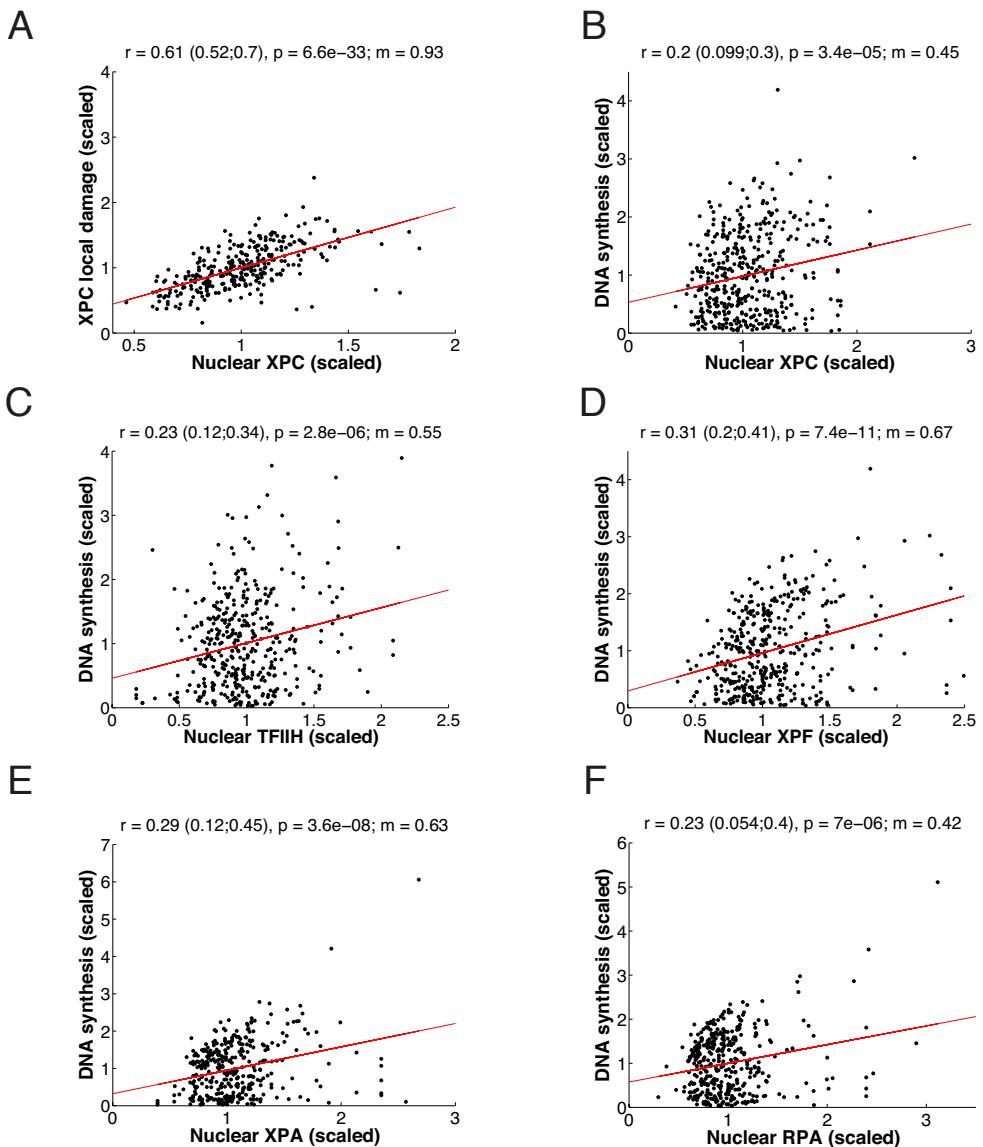


Figure 4.4.: Natural expression variability of the repair factors XPC, TFIIH, XPF, XPA and RPA has small effect on DNA repair synthesis. A) Scatter plot of accumulated XPC in local damage against the nuclear XPC concentration ($n=303$) 60 minutes post irradiation. B-F) DNA synthesis correlated with the endogenous concentration of XPC ($n=425$), TFIIH ($n=220$), XPF ($n=425$), XPA ($n=350$) and RPA ($n=383$) 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 (for further details cf. Section A.2). (Experiments by P. Verbruggen)

In fact, the nuclear protein concentrations for all five antibody-stained repair factors and the amount of incorporated EdU after one hour were weak, but still significantly correlated as indicated by the correlation coefficients and the corresponding small p-values (*cf.* Figure 4.4B-F). Remarkably, the dependency between protein concentration and EdU response, characterized by the slope of the regression line is very evenly distributed (*cf.* Figure 4.1A). 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.

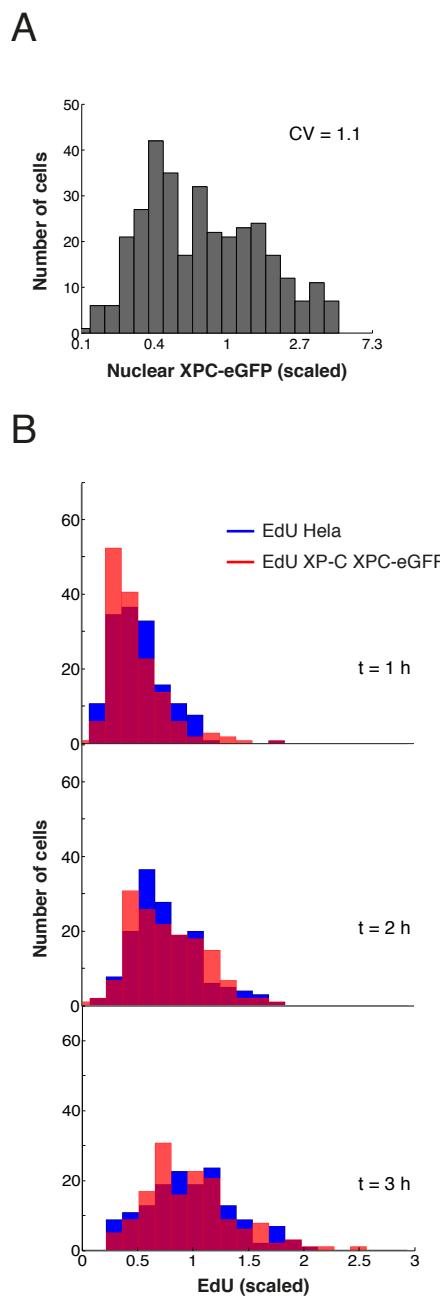
To reinforce the conclusion of a robustness DNA repair rate we compared the repair factor expression and its consequence on the repair variability in different cell lines. In contrast to indirectly labelled XPC in Hela cells [49] or in human primary fibroblasts (*cf.* Figure 4.3), the expression of stably transfected XPC-eGFP into XP-C patient cells is much broader distributed. Its *CV* has a value of ~ 1 , which is up to four times larger compared to XPC variability in Hela or fibroblasts cells (compare Figure 4.5A with Figure 4.3A and [49]).

In Hela cells and XPC-eGFP complemented XP-C cells we measured EdU incorporation after 1, 2 and 3 hours post UV-irradiation (*cf.* Figure 4.5). At all three time points the distributions of newly-incorporated DNA were nearly congruent suggesting that DNA repair follows the same kinetics regardless of the cell type and the corresponding NER factor expression variability. This result allows to generalize the concept of a robust repair rate for different cell types and their cellular environments.

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

Comparing the measured EdU incorporation as a function of NER factor concentration and the predicted response coefficients quantitatively (*cf.* Section 4.2), we noted that the measured repair responses towards changes in the nuclear protein concentration are consistently two to four-fold larger than the mathematically calculated values (compare the slopes m in Figure 4.4B-F with the response coefficients \tilde{R}_i in Figure 4.1A). This result suggests that the measured repair-rate response is not solely explained by the model-predicted response coefficients. It remains to be clarified, whether this discrepancy is due to a so far unknown additional NER component intrinsically contributing to the pathway control, or whether there is an external mechanism coordinating the NER factor expression and thereby superimposing the mathematically predicted response coefficients.

4. Control analysis of the DNA repair rate



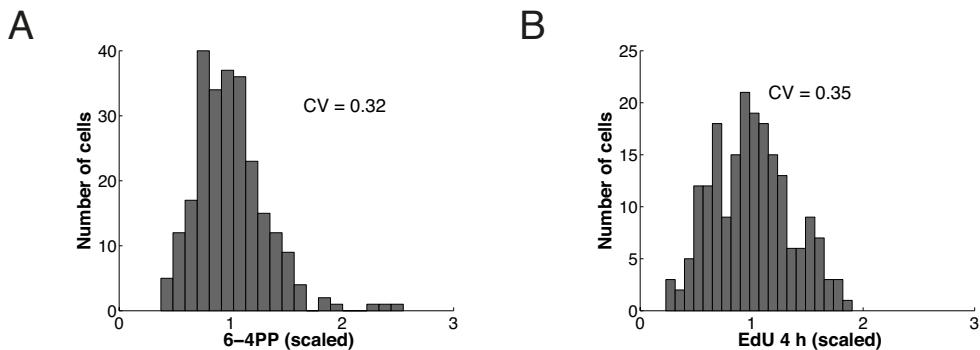


Figure 4.6.: Distributed amount of inflicted DNA damages contributes to repair rate variability. A) XP-C XPC-eGFP cells were locally UV-irradiated with an intensity of 100 J/m^2 . DNA damages were quantified by indirect immunofluorescence microscopy in $n=250$ cells from five experiments. B) XP-C XPC-eGFP 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. (Experiments by P. Verbruggen)

While searching for additional factors affecting the repair rate response we noticed the broad scatter of incorporated EdU, which appears to increase for larger protein concentrations (*cf.* Figure 4.4B-F). This variability indicates that, in addition to the repair proteins, there might be further sources of cell-to-cell heterogeneity, which could 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 to determine how the distribution of pore sizes would propagate (*cf.* Figure 2.1B). The distribution of DNA damage captured immediately after UV-irradiation very much resembles the variable EdU incorporation after 4 hours (*cf.* Figure 4.6), which is also confirmed by the similar *CVs*. These data suggest that the amount of DNA lesions contributes to the cell-to cell variability in the repair-rate.

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 evolved quantitative model introduced in Chapter 3. Thus, we simulated the model 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 *CVs* (*cf.* Figure 4.3 and Figure 4.6) and the previously measured mean values [6]. For PCNA and XPG we took the averaged *CV* ~ 0.33 derived from the measured distributions.

For XPC the simulated scatter plot for the accumulation of the nuclear repair factor at sites of local damage agrees remarkably well with the measured data (*cf.* Figure 4.7A, compare

4. Control analysis of the DNA repair rate

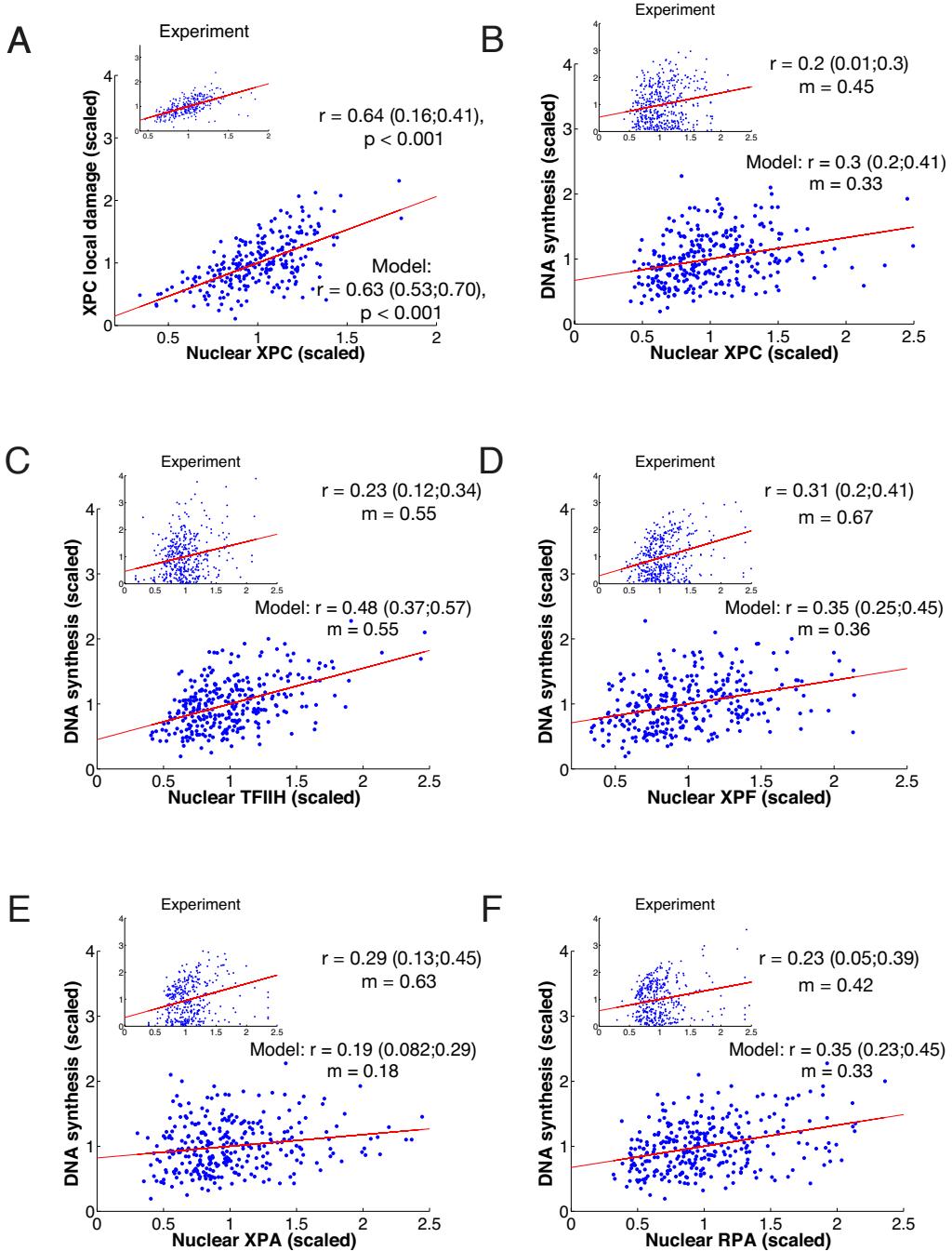


Figure 4.7.: Quantitative comparison between measured and simulated repair resynthesis in response to variable NER factor concentrations. A) Simulated scatter plot of locally accumulated XPC against the nuclear XPC concentration. For comparison the experimentally measured scatter plot is given in reduced 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 (only in A) and slope m . 95% confidence bounds of all correlation coefficients r were estimated by non-parametric bootstrap and are given in brackets (for further details cf. Section A.2). Nuclear protein expression values are scaled by dividing by the mean value. (Experiments by P. Verbruggen)

Repair protein	Measured m (cf. Figure 4.4)	Simulated m (cf. Figure 4.7)	\tilde{R}_i (cf. Figure 4.1)
XPC:	0.36 ± 0.21	0.33	0.20
TFIIFH:	0.45 ± 0.08	0.55	0.30
XPA:	0.75 ± 0.17	0.18	0.16
XPF:	0.62 ± 0.10	0.36	0.15
RPA:	0.37	0.33	0.15

Table 4.2.: Comparison of the measured and predicted repair responses caused by fluctuations of the NER factor concentrations. Changes in DNA repair synthesis in response to variations of the nuclear NER factor abundance quantified i) from experimental data (mean \pm SD of the regression slopes m from $n_{XPC} = 5$, $n_{TFIIFH} = 3$, $n_{XPA} = 3$, $n_{XPF} = 3$ and $n_{RPA} = 2$ independent biological replicates; within each measurement between $n = 250$ and $n = 572$ with an average of $n = 477$ cells were analysed), ii) from model simulation (nuclear NER factor concentrations and the amount of local damages were randomly drawn from log-normal distributions in accordance with the experimentally determined CVs (cf. Figure 4.3 and Figure 4.6)) and predicted iii) by the response coefficients \tilde{R}_i (cf. Section 4.1).

enlarged, simulated correlation with the smaller inlay, of the experimentally observed scatter plot). 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 concentration of the repair factors (cf. Figure 4.7B-F). 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 4.1), clearly marking the influence of the introduced distribution of DNA damages. In particular for XPA, XPF and RPA though, the simulated slopes are still significantly smaller than the measurement-based regression estimates (cf. Table 4.2). This indicates that we still miss out on an essential, perhaps intrinsic, factor regulating the repair-rate response.

To locate the lacking control, we first asked whether the natural variability, contributed by the NER factor expression variability and the distribution of inflicted lesions, can explain the distribution of the repair rate. Assuming that the repair rate ν is linearly dependent on the initial amount of inflicted lesions L and, as estimated in Section 2.4, also on the nuclear repair-factor concentrations C_i ($i = 1, \dots, N$), we can apply the general law of error propagation

$$\sigma(\nu) = \sqrt{\sum_{i=1}^N \left(\frac{\partial \nu}{\partial C_i} \sigma(C_i) \right)^2 + \left(\frac{\partial \nu}{\partial L} \sigma(L) \right)^2}. \quad (4.10)$$

Introducing the predicted response coefficients \tilde{R}_i we can estimate the overall repair rate variability CV_ν with

$$CV_\nu = \sqrt{CV_L^2 + \sum_{i=1}^N (\tilde{R}_i CV_i)^2}, \quad (4.11)$$

4. Control analysis of the DNA repair rate

where CV_i and CV_L denote the coefficients of variation of the distributions of the individual repair factors and initial amount of lesions. Obviously, the response coefficient for the initial amount of inflicted lesions is 1. Including the variability for the initial amount of inflicted lesions $CV_L \sim 0.32$ (*cf.* Figure 4.6A) and the average $CV_i \sim 0.33$ for each repair factor, we can determine the repair rate variability with $CV_v \sim 0.35$. This is around 20% less than the measured $CV_v \sim 0.45$ after one hour post UV-irradiation (*cf.* Figure 4.8).

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 4.8 we determined the CV_v 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 (*cf.* Figure 4.8 brown trajectory), which was responsible for about two thirds of the whole observed repair variability (*cf.* Figure 4.8 blue trajectory). Drawing all NER factors together with the initial amount of damages from their pre-determined distributions shrinks the gap between measured and simulated DNA repair variability (*cf.* Figure 4.8 blue (data) and red (simulation) trajectory). This result becomes more explicit, as seen, by comparing the strikingly matching temporal evolution of the measured and predicted distributions of repair synthesis (*cf.* Figure 4.9).

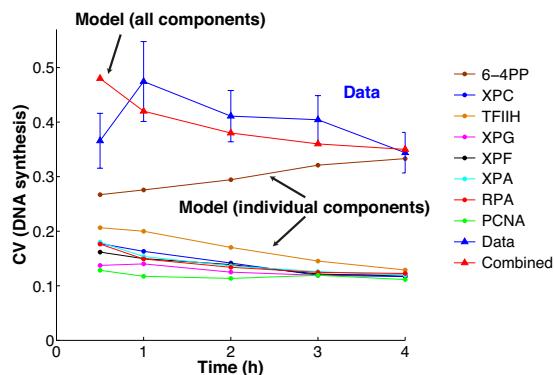


Figure 4.8.: Variability in repair synthesis results from the aggregate of initial damage variability and concentration fluctuations of each NER factor. Each time course represents the variability in repair synthesis produced by a single component (dots) or by all components together (triangles) at a 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.

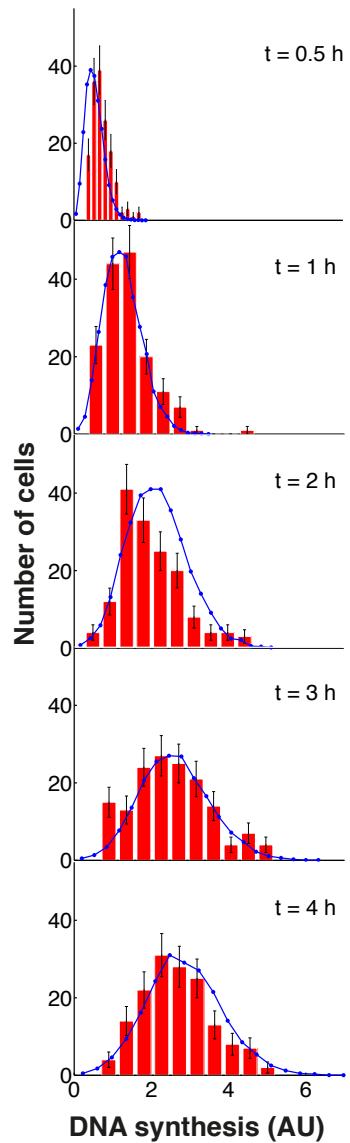


Figure 4.9.: Simulated repair rate distribution fits experimentally derived EdU incorporation. Temporal evolution of the repair rate variation measured in single cells after initial UV-irradiation (red histograms) and predicted from model simulations (blue lines). (Experiments by P. Verbruggen)

4. Control analysis of the DNA repair rate

To summarize, in this chapter we uncovered two sources of natural variability that are propagated throughout the repair process and consequently contributing to the emergent cell-to-cell variability of repair synthesis: (i) the initial amount of inflicted DNA lesions and (ii) the variation of the nuclear NER factor concentrations. However, the significant differences between measured and predicted repair rate responses (*cf.* Table 4.2) leave space for interpretation, whether there are additional factors involved in repair, but have not been explicitly accounted for, in the model.

5. 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 (*cf.* Chapter 2). 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 concentrations (*cf.* Section 4.1). 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 (*cf.* Table 4.2).

In this 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 [76]. This result suggests an additional control mechanism orchestrating NER factor expression on the pre-transcriptional level.

The microscopy experiments reported in this chapter have been planned and conducted by Paul Verbruggen (University of Amsterdam). Image segmentation was performed by Simon Eck (Division of Theoretical Bioinformatics, DKFZ). The flow cytometry analysis was done in close collaboration with Martin Teichert, who is part of the Division of Vascular Oncology and Metastasis led by Prof. Helmut Augustin.

5.1. Cross-correlation affects repair response

The finding that the measured responses in DNA synthesis is about two-fold higher than the model-predicted response coefficients (*cf.* Figure 4.1 and Figure 4.4) indicates that there is an additional factor regulating the repair response. Not yet considered is the poten-

tial interdependence of the NER factor expression, for example 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 the measured EdU responses (*cf.* Figure 4.4) and the model-predicted response coefficients (*cf.* Figure 4.1) to

$$\frac{C_i}{\nu} \frac{d\nu}{dC_i} = \frac{C_i}{\nu} \frac{\partial \nu}{\partial C_i} + \sum_{\substack{j=1 \\ j \neq i}} \left(\frac{C_j}{\nu} \frac{\partial \nu}{\partial C_i} \right) \left(\frac{C_i}{C_j} \frac{\partial C_j}{\partial C_i} \right), \quad (5.1)$$

where ν represents the rate of EdU incorporation and C_i the nuclear concentration of the i^{th} repair factor for $i = 1, \dots, N$ and $j = 1, \dots, N$. From Eqn. 5.1 we directly calculated the EdU incorporation for varying NER factor concentrations measured after one hour post UV-irradiation (*cf.* Figure 4.4)

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

By also taking into account the pairwise cross-correlations

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

we can derive the following system of equations

$$R_i + \sum_{\substack{j=1 \\ j \neq i}} B_{ij} R_j = A_i, \quad (5.2)$$

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

$$R_i = \frac{C_i}{\nu} \frac{\partial \nu}{\partial C_i}. \quad (5.3)$$

Assuming that all $B_{ii} = 1$ Eqn. 5.2 translates into the matrix notation

$$\vec{A} = B \vec{R}. \quad (5.4)$$

By inverting B we can directly solve Eqn. 5.4 for the desired response coefficients.

5. Co-expression of NER repair factors

	XPC	TFIILH	XPA	XPF	RPA
XPC	<u>X</u>	X	X	<u>X</u>	<u>X</u>
TFIILH	-	-		X	-
XPA			-	X	X
XPF				-	-
RPA					-

Table 5.1.: Immuno-fluorescence microscopy of NER factor co-expression. Crosses indicate pairwise measurement of indirectly antibody-labeled NER factor expression. Underlined crosses denote co-stainings involving directly labeled XPC.

5.2. 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 (*cf.* Figure 4.4). 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 5.1) using the primary and secondary antibodies listed in the Tables A3, A4 and A5.

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 [77, 78]. The cross-correlation analysis was performed in human diploid female fibroblasts, which were grown to confluence on cover-slips. Analogous to the description in Section 2.1 cells were UV-irradiated locally with a dose of 100 J/m². Cells were pre-incubated with serum-free medium containing 10 µM EdU and allowed to repair for 60 minutes in an incubator. After the subsequent direct or indirect antibody labelling of two selected repair factors, cells were additionally incubated with a DAPI solution, which visualizes the cell's DNA 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 2.2. Segmentation of the nucleus was performed on the DAPI signal, due to the large signal-to-noise ratio in this channel (*cf.* Figure 5.1A). 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 5.1B and C). For the same reason, segmentation of the locally UV-irradiated chromatin region was done with the area determined

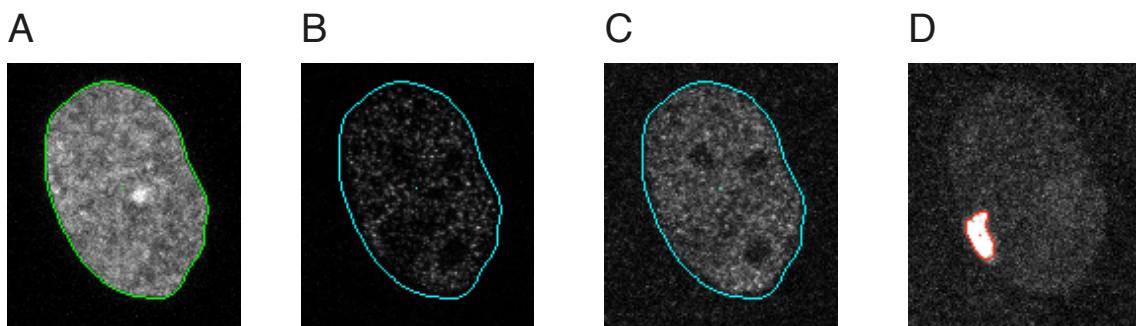


Figure 5.1.: 3D fluorescence microscopy imaging of NER factor co-expression. A) DAPI-stained DNA content measured in human primary fibroblast (Life technologies). DAPI segmentation at 461 nm (green contour line) determines nuclear dimension. B-C) Nuclear expression of indirectly immuno-stained repair 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 μ M EdU. Segmentation of the EdU signal is indicated by the red contour line. (Experiments by P. Verbruggen)

from the EdU signal and then projected onto the signal of accumulated protein (*cf.* Figure 5.1D).

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 5.2A), as described in section 2.2, 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.* Table 5.1) are strongly positively correlated with correlation coefficients between 0.66 and 0.87 (*cf.* Figure 5.2B-D and Figure A.3). 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 5.2B-D and 5.2E-F). This suggests that a dominating part of the positive 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.

5. Co-expression of NER repair factors

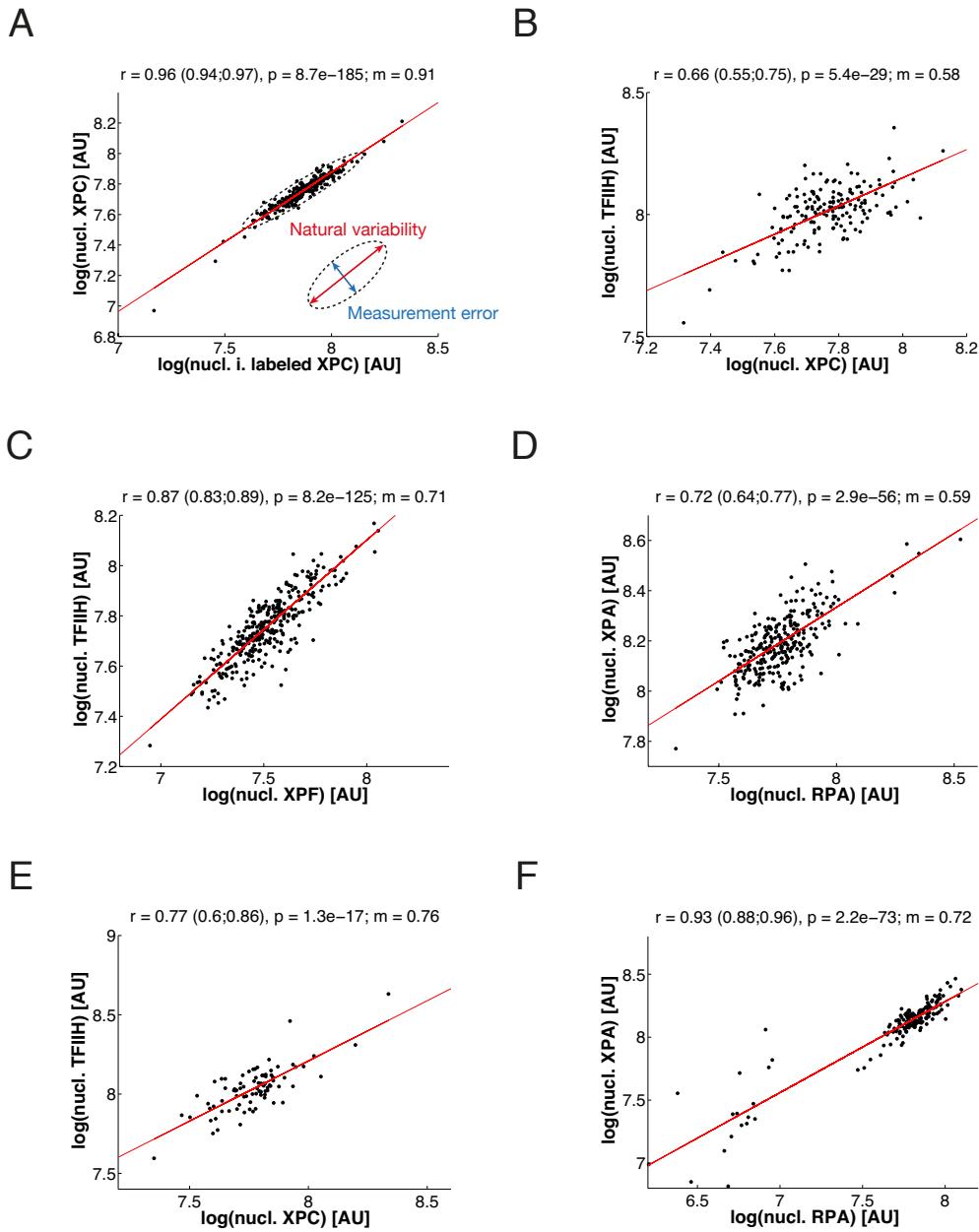


Figure 5.2.: NER factor cross-correlation is repair independent. A) Scatter plot of indirectly antibody-stained XPC against a directly labelled antibody recognising XPC ($n=336$) as determined by quantitative (immuno) fluorescence microscopy. B-D) Pairwise correlations of indirectly antibody-labelled TFIIH against XPC ($n=220$, B), TFIIH 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 TFIIH vs. XPC ($n=85$, E) and XPA vs. RPA ($n=170$, F) 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. (Experiments by P. Verbruggen)

5.3. Cell cycle independent cross-correlation of NER factor expression

To confirm the microscopy results 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 in this cell type the expression of the two repair factors also strongly correlates (*cf.* Figure 5.3A and B). Quantitatively, the correlation coefficients of the double-staining signals in both cell types (human brain pericytes and human primary fibroblasts) have the same order of magnitude. In particular, each value falls into the confidence interval of the other (*cf.* Figure 5.3B and Figure 5.2D). 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 = 0.74), it is still significantly positively correlated (*cf.* Figure 5.3C and D). As Ki67 increases during cell cycle we asked, to what extent 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 5.4A). Two distinct peaks denote the portion of cells traversing the G1 or the S 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 S and G2 phase in comparison to the G1 phase (*cf.* Figure 5.4B-D).

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 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.

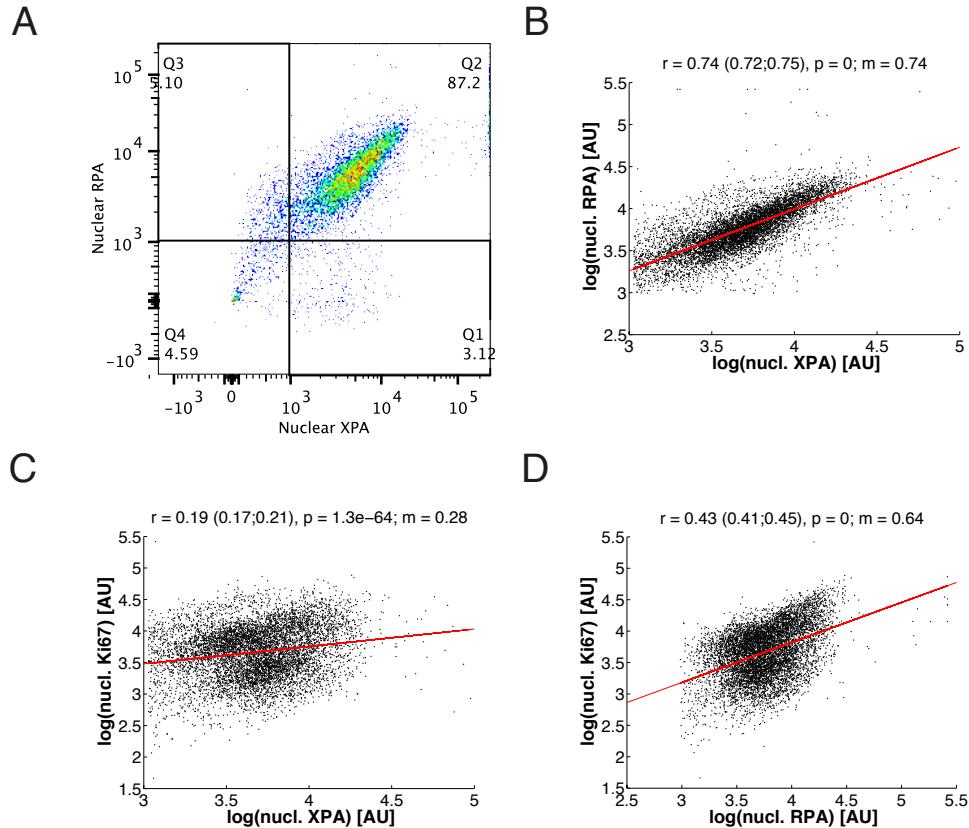


Figure 5.3.: Correlated expression of RPA and XPA in human brain pericytes. A) Selection of XPA and RPA positive human brain pericytes (Q2) as determined by flow cytometry ($n=8084$). B-D) Nuclear expression of indirectly antibody-labelled RPA against XPA (B), Ki67 against XPA (C) and Ki67 against RPA (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. (Experiments were performed in collaboration with M. Teichert)

5.4. 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 5.1 the measured cross-dependencies represent the entries of matrix B in Eqn. 5.4. In theory, the product of the inverse of B and A defines the vector of response coefficients. However, due to the strong multicollinearity the matrix B has variance inflation factors (VIFs) between 2.5 and 10.8 indicating that the variation of each protein is largely explained by a linear combination of the other protein expression values.

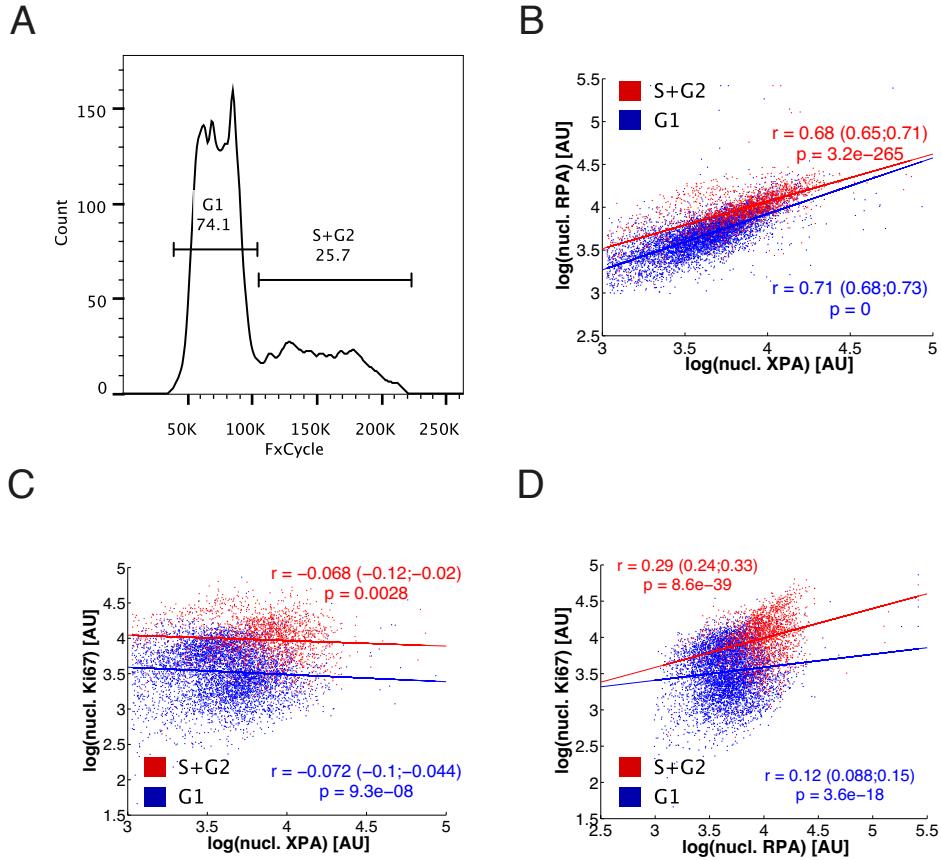


Figure 5.4.: NER factor cross-correlation is robust against cell cycle progression. A) Distribution of fluorescently labelled DNA proportional to the DNA content. Horizontal bars indicate the fractions of cells assigned to be in cell-cycle phase G1 (left peak, n=5469) and S+G2 (right peak, n=1945). B-D) Individual regression analysis of the nuclear expression values sorted according to the corresponding cell-cycle phase (G1 in blue; S+G2 in 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. (Experiments were performed in collaboration with M. Teichert)

VIFs measure how much of the predicted standard error of the response coefficient estimates is caused by the collinearity [79]. For instance, for a VIF of 5 the estimated error is about $\sqrt{5} \cong 2.3$ 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 \tilde{R} . Consider ν as a function of the concentration of the repair factors C_i for $i = 1, \dots, N$ multiplied with the initial amount of DNA lesions L :

$$\nu = Lf(C_1, \dots, C_N). \quad (5.5)$$

5. Co-expression of NER repair factors

Using, once more, the standard law for the propagation of uncertainty analogue to Eqn. 4.10

$$\sigma(\nu) = \sqrt{\sum_{i=1}^N \left[\left(\frac{\partial \nu}{\partial C_i} \sigma(C_i) \right)^2 + \left(\frac{\partial \nu}{\partial L} \sigma(L) \right)^2 + \sum_{\substack{j=1 \\ j \neq i}}^N \left(\frac{\partial \nu}{\partial C_i} \frac{\partial \nu}{\partial C_j} \text{cov}(C_i, C_j) \right)^2 \right]} \quad (5.6)$$

and introducing the response coefficients from Eqn. 5.3 we derive

$$CV_\nu^2 = CV_L^2 + \sum_{i=1}^N \left((\tilde{R}_i CV_i)^2 + \sum_{\substack{j=1 \\ j \neq i}}^N \tilde{R}_i \tilde{R}_j \frac{\text{cov}(C_i, C_j)}{\langle C_i \rangle \langle C_j \rangle} \right), \quad (5.7)$$

where $\langle C_i \rangle$ denotes the mean repair factor concentration. With the simplifying assumption that $\tilde{R} = \tilde{R}_i$ for all $i = 1, \dots, N$, which implies that all \tilde{R}_i are the same, we can solve Eqn. 5.7 for \tilde{R} and thereby derive an estimate for the average response coefficient:

$$R = \sqrt{\frac{CV_\nu^2 - CV_L^2}{\sum_{i=1}^N CV_i^2 + \sum_{\substack{j=1 \\ j \neq i}}^N \frac{\text{cov}(C_i, C_j)}{\langle C_i \rangle \langle C_j \rangle}}}. \quad (5.8)$$

On the right-hand side we measured most of the CV s and about a third of the NER factor covariations (cf. Table 4.1, Figure 5.2 and A.3). The missing entries were added by taking the average of the known elements. By inserting all measured and estimated quantities into Eqn. 5.8 we calculated an average response coefficient of $\tilde{R} = 0.16$. Remarkably, this result agrees exactly with the average response coefficients predicted by the model (cf. Figure 4.1) and thus, reconfirms the hypothesis of a small and collectively shared repair-rate control.

6. Discussion

Our DNA-repair analysis was made possible by the ability to explicitly observe the kinetics of resynthesised DNA during nucleotide-excision repair (NER). The quantification of this repair intermediate by following the incorporation of fluorescently labelled nucleotides on the single cell level shows that the repair process follows a slow, first-order kinetic (*cf.* Chapter 2). These new data supplement comprehensive fluorescent real-time measurements quantifying the dynamic behaviour of single repair proteins. Hence, we can observe the individual protein exchange kinetics at the chromatin fibre, which eventually lead to the formation of catalytic multi-protein machineries, concomitantly with the global repair process. Based on this information we were able to identify a detailed kinetic model of NER, which allows us to establish 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 (*cf.* Chapter 3). The most important implication of the model is that the repair rate is controlled by the repair factors in a collective way. In particular, the uniformly small response coefficients emphasize the robustness of the DNA repair process. This disproves the widely-held assumption that a single rate-limiting step controls the entire repair process. Our finding of a shared control has been corroborated experimentally by exploiting measurements of the natural variability of the nuclear NER factor concentrations (*cf.* Chapter 4). However quantitatively, the measured repair responses exceed the magnitude of the predicted response coefficients significantly.

While further pursuing the nature of the measured cell-to-cell variability and its influence on DNA repair, we observed that the substantial heterogeneity of the repair rate is generated primarily by the distribution of inflicted DNA lesions. This factor is complemented by the variability contributed from the nuclear concentration of each individual repair factor. The quantitative impact of these two parameters on the cell-to-cell variability of DNA repair is consistent with the robustness of the repair rate against concentration fluctuations of individual repair factors (*cf.* Chapter 4).

Interestingly, we also found evidence for the regulated NER factor expression indicated by strong cross-correlations of the protein concentrations. Considering these mutual interdependencies in our control analysis quantified their contribution to the repair-rate response and thus, explains the gap between measured and predicted DNA repair response (*cf.* Chapter 5). These results imply that the robustness of the repair rate is not only controlled by the pathway-intrinsic molecular interplay of chromatin-associated machineries but also extrinsi-

cally by a regulatory mechanism orchestrating NER factor expression.

In the following, we will elaborate a more detailed discussion about individual findings including: The DNA repair kinetics, NER model identifiability, model parametrization, repair-rate control and the co-regulation of NER factor expression.

Slow first order kinetics as a general systems property for chromatin associated processes

In our analysis of the NER kinetic we focused on the removal of one of the major UV-induced DNA lesions, the 6-4 pyrimidine-pyrimidone photoproducts (6-4PPs) (*cf.* Figure 2.5B), which are repaired significantly faster (~2-3 hours) compared to CPDs, which are still present after ~24 hours [6, 80, 81]. 6-4PPs are repaired by incision of the damaged DNA strand, and subsequent DNA resynthesis, 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, which requires the cells to be fixed. Hence, EdU accumulation cannot be measured continuously in one cell. Instead, we followed EdU incorporation stepwise by fixing cells after increasing time intervals, each starting immediately after UV-irradiation. In this way we were able to show that the 6-4PP repair fits to a first order kinetic with a half-life of ~1.2 hours. This long time scale is in contrast to the rapid exchange behaviour of early and late repair factors as reported by several live-cell imaging studies. [6, 18, 40–43, 48, 52].

In agreement with a simplified model simulating the assembly of a catalytic protein repair-complex [50], we find here that reversible exchange of NER factors is the key property that makes NER an apparent first-order process. In this description, the long time scale arises from the concurrence of stochastic assembly and reversible exchange of the constituent components, which leads to the formation of many non-functional preliminary stages before eventually the catalytic complex is completed. 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 [50]. As suggested by Luijsterburg *et al.* (2010) [6], the slow repair time might actually present an advantage for the specificity of the repair process, analogous to a kinetic proof-reading mechanism, where differential dissociation kinetics would prevent false-positive lesion detection.

The phenomenon of rapid protein exchange is widespread also in other chromatin associated processes such as replication, chromatin remodelling or transcription [20, 22, 82, 83]. Analogous to NER, these processes involve finding a specific reaction site before the catalytic machinery can assemble at the same location. In the particular case of DNA transcription, transient protein dynamics were experimentally demonstrated for transcription fac-

6. Discussion

tor binding and transcription machinery assembly [84]. The time scales for these reactions (several seconds to minutes) are comparable to our results for DNA repair. This encourages us to hypothesise that the slow time scales of transcriptional bursting in mammalian cells (tens of minutes to hours; [85, 86]) also arise through the reversible assembly of large macromolecular complexes.

A predictive kinetic model of DNA repair Augmenting the extensive kinetic data on 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 ODE-model of NER (*cf.* Chapter 3). To the best of our knowledge it is 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 experimental data.

On the technical side, to make the profile likelihood estimation for each parameter computationally feasible, we first needed to optimize the runtime of the parameter estimation algorithm. Therefore, we integrated the model into a dynamic modelling framework that applies a deterministic trust-region fitting algorithm implemented in MATLAB [70]. The accuracy and computational performance of the implementation benefit immensely from the simultaneous integration of the dynamic variables concomitantly with the derivatives of the objective function with respect to the model parameters [46, 87, 88]. Solving this extended system of ordinary differential equations with a CVODE solver implemented in C accelerated the speed for one deterministic optimization run by 200-fold compared to the previously applied stochastic Markov Chain Monte Carlo (MCMC) optimization algorithm [50]. The runtime spent for one fit on a 12-core processor with 3 GHz each is now reduced from two weeks to \sim 10 minutes. This considerable runtime improvement allowed us to sample the parameter space exhaustively using Latin hypercube sampling, a technique, which makes the deterministic optimization algorithm more robust against local optima [46].

In addition to fitting the parameters that determine the model dynamics, we estimated the parameters that characterize the measurement noise. This approach of explicitly including the measurement noise, has the advantage that it avoids an inaccurate ad-hoc estimate of the experimentally-determined error due to low numbers of replicates and thus gives a more exact assessment of the model parameters, without preprocessing the experimental data [46].

Imposed by the non-identifiability of several parameters in the model of Luijsterburg *et al.* [6], a number of simplifications concerning the model structure were required to achieve an identifiable model of the NER pathway. The two most significant modifications that we in-

troduced here are 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. Since we lack a direct readout for incised DNA, we additionally assumed a practically instantaneous incision of the DNA lesion once the pre-incision complex has been completely assembled. The missing information about DNA incision also limits the identifiability of the other catalytic rates (despite the rate of rechromatinisation), indicated by the existence of lower confidence bounds, only. This case of structural non-identifiability proved to be without consequence for the goodness of the model predictions (*cf.* Figure 4.1). Moreover, fast catalytic rate constants agree with a recent direct measurement in bacteria where DNA synthesis and ligation take seconds [89]. Our current model, whose structure is reconcilable with the experimental data, constitutes a concrete improvement over previous NER models [6, 57, 90]. We believe that, gaining further mechanistic understanding about system properties requires an intertwined advancement, where the addition of molecular detail in the model is consecutively balanced by appropriate quantitative measurements. Already at the present level of detail, we are confident 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 3.1).

Model parametrization and its implications The model fit (described in detail in Chapter 3) yields biochemically plausible estimates for the kinetic parameters of the individual assembly and dissociation kinetics *in vivo*, which account for both the long-term accumulation and for 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 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 [91]. As discussed in depth at Luijsterburg and coworkers (2010) [6] the strong reversibility is also beneficial for the specificity and regulation of the system, for example by preventing the trapping of NER proteins in incomplete (and thus unproductive) repair complexes. The reversibility advantage also applies for the trade-off between specificity and efficiency of the mechanism, which equally plays a role in chromatin remodelling and transcription [45, 92].

Apart from fast reversible assembly/dissociation, parameter estimation also revealed that the catalytic rate constants are fast, which indicates a tight coupling between lesion exci-

6. Discussion

sion and repair synthesis. This in turn implies that the lesions are repaired without much delay immediately after their recognition by XPC, which consequently prevents the accumulation of incised DNA. Our *in-vivo* finding stays in contrast to *in-vitro* experiments that have found a delay before repair synthesis [28, 93]. Our prediction that the repair system prevents single-stranded repair intermediates to accumulate corresponds with our expectation that an elevated abundance of incised DNA would trigger DNA degradation to avoid the risk of an inflammatory response or autoimmunity [94].

Collective control of the DNA repair rate A consequence of the NER model architecture, which describes the sequential organization of this chromatin-associated process by cycles of protein recruitment, is the absence of any rate-limiting factor for repair synthesis. Instead, the control of the repair speed is homogeneously distributed, with small contributions from each repair factor. This central result of our work agrees with the outcome of previous flux control analyses of various biochemical pathways that found no experimental support for the existence of a unique rate-limiting factor [72]. Moreover, response coefficients with values much smaller than 1 indicate that DNA repair is robust against natural fluctuations of the protein concentrations in the cell [95]. It is important to note that the degree of control as a kinetic property of a particular NER factor is independent of its position in the order of engagement in the repair process. Thus, the lesion recognition factor XPC, which binds first, has 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 [95, 96], DNA repair will be affected by 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 (*cf.* Figure 4.2). 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 [97–99]. 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 [97–101].

To analyse the effect of natural changes in NER factor expression on the repair rate experimentally, instead of changing the protein expression by gene knock-down or overexpression, which usually cause the partial or complete disruption of essential system functions [102], we exploited the natural variability of the nuclear protein concentration. Despite the weak correlation in the data, we saw a significantly positive dependency between protein expression for all five measured NER factors and DNA resynthesis. 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- β was significantly increased, when five out of six transcription factors necessary for transcriptional activation were overexpressed [103]. A similar result was obtained recently, which suggested that the expression of IL-2 is induced by the concerted interplay of four transcription factors together with FOXP3 [104]. These findings emphasize that also transcription is a collectively controlled process.

Co-regulation of the nuclear NER factor expression The question how robustness to perturbations of the cellular state emerges from the underlying biochemical networks, has attracted considerable interest. A range of regulatory motifs have been proposed as producers of a robust response (e.g. negative feedback, incoherent feed-forward loops and functional redundancy) [105–108]. However, most of these functional building blocks are not applicable for promoting robust DNA repair as the NER pathway shows robust regulation without any reported signalling events (as described in this work), nor should transcriptional regulation play a role on the considered time scale between damage infliction and repair initiation [8]. The individual repair reactions purely depend on the repair complex assembly time and thus are determined by the NER factor binding affinities (this work and [6]). In essence, robustness of the repair speed relies on the fact that minor delays or speed-ups caused by fluctuations of one repair factor during the assembly process of the full complex, are negligible on the global time scale. Therefore, randomness in many reversible steps is averaged by the law of large numbers. 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 cancer cells [109, 110], we observed this cross-correlated expression in single cells of various types under unperturbed, NER-independent conditions. Despite the overall increase in nuclear NER factor concentrations during S-phase and G2 phase, the cross-correlation persists independently of the cell cycle. Together, these results imply an additional regulative mechanism that indirectly controls the repair rate already on the transcriptional level. Since, as we show, NER control is distributed, we predict that this

6. Discussion

transcriptional control must act on many factors simultaneously to be effective.

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 [111, 112]. This co-localisation leads to co-expression, which in turn provides concentration robustness against variability of other pathway components [95]. So far such a transcriptional coupling mechanism is unknown for the repair proteins involved in NER. Therefore it would be 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 DNA is essential for the long-term survival of the cell since it is absolutely necessary for the fidelity of all chromatin-associated processes. It is therefore not surprising to find two molecular cooperating mechanisms that ensure robustness of the repair rate against natural variability. First, as a consequence of the dynamic design of the repair complex assembly, there exists a uniformly distributed control of the repair rate, where each factor has only a moderate effect on its own. This translates into a low system response in the event of NER factor fluctuations. Second, a yet unknown extrinsic mechanism regulates the coordinated NER factor expression, which limits the degree of protein variability, presumably, already pre-transcriptionally. Given that chromatin remodelling, transcription and translation have similar dynamic features as DNA repair, both mechanisms may function as a widespread mode of robust regulation of chromatin-associated processes.

7. 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. *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] Bartek, J., Bartkova, J. & Lukas, J. DNA damage signalling guards against activated oncogenes and tumour progression. *Oncogene* **26**, 7773–9 (2007).
- [8] 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).
- [9] Marteijn, J. A., Lans, H., Vermeulen, W. & Hoeijmakers, J. H. J. Understanding nucleotide excision repair and its roles in cancer and ageing. *Nature Reviews Molecular Cell Biology* **15**, 465–81 (2014).
- [10] Ciccia, A. & Elledge, S. J. The DNA damage response: making it safe to play with knives. *Molecular Cell* **40**, 179–204 (2010).
- [11] Lindahl, T. & Barnes, D. Repair of endogenous DNA damage. *Cold Spring Harbor Symposia on Quantitative Biology* **65**, 127–134 (2000).
- [12] Caldecott, K. W. Single-strand break repair and genetic disease. *Nature Reviews Genetics* **9**, 619–31 (2008).
- [13] Moldovan, G.-L. & D’Andrea, A. D. How the fanconi anemia pathway guards the genome. *Annual Review of Genetics* **43**, 223–49 (2009).
- [14] West, S. C. Molecular views of recombination proteins and their control. *Nature Reviews Molecular Cell Biology* **4**, 435–45 (2003).
- [15] Hoeijmakers, J. H. Genome maintenance mechanisms for preventing cancer. *Nature* **411**, 366–374 (2001).
- [16] Gillet, L. C. & Schärer, O. D. Molecular mechanisms of mammalian global genome nucleotide excision repair. *Chemical Reviews* **106**, 253–276 (2006).

- [17] Dinant, C., Luijsterburg, M. S., Höfer, T., von Bornstaedt, G., Vermeulen, W., Houtsmuller, A. B. & van Driel, R. Assembly of multiprotein complexes that control genome function. *The Journal of Cell Biology* **185**, 21–26 (2009).
- [18] 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).
- [19] Vermeulen, W. Dynamics of mammalian NER proteins. *DNA Repair* **10**, 760–71 (2011).
- [20] Stasevich, T. J. & McNally, J. G. Assembly of the transcription machinery: ordered and stable, random and dynamic, or both? *Chromosoma* **120**, 533–45 (2011).
- [21] Kuipers, M. A., Stasevich, T. J., Sasaki, T., Wilson, K. A., Hazelwood, K. L., McNally, J. G., Davidson, M. W. & Gilbert, D. M. Highly stable loading of Mcm proteins onto chromatin in living cells requires replication to unload. *The Journal of Cell Biology* **192**, 29–41 (2011).
- [22] 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).
- [23] Sugasawa, K., Okuda, Y., Saijo, M., Nishi, R., Matsuda, N., Chu, G., Mori, T., Iwai, S., Tanaka, K. & Hanaoka, F. UV-induced ubiquitylation of XPC protein mediated by UV-DDB-ubiquitin ligase complex. *Cell* **121**, 387–400 (2005).
- [24] Fousteri, M. & Mullenders, L. H. F. Transcription-coupled nucleotide excision repair in mammalian cells: molecular mechanisms and biological effects. *Cell Research* **18**, 73–84 (2008).
- [25] Sugasawa, K., Ng, J. M., Masutani, C., Iwai, S., van der Spek, P. J., Eker, A. P., Hanaoka, F., Bootsma, D. & Hoeijmakers, J. H. Xeroderma pigmentosum group C protein complex is the initiator of global genome nucleotide excision repair. *Molecular Cell* **2**, 223–232 (1998).
- [26] Hanawalt, P. C. & Spivak, G. Transcription-coupled DNA repair: two decades of progress and surprises. *Nature Reviews Molecular Cell Biology* **9**, 958–70 (2008).
- [27] 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. *Journal of Biological Chemistry* **275**, 9870–9875 (2000).
- [28] 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).
- [29] 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).
- [30] Compe, E. & Egly, J.-M. TFIIH: when transcription met DNA repair. *Nature Reviews Molecular Cell Biology* **13**, 343–54 (2012).
- [31] Evans, E., Fellows, J., Coffer, A. & Wood, R. D. Open complex formation around a lesion during nucleotide excision repair provides a structure for cleavage by human XPG protein. *EMBO Journal* **16**, 625–638 (1997).
- [32] 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 & Development* **12**, 2598–2609 (1998).

Bibliography

- [33] Wakasugi, M., Reardon, J. T. & Sancar, A. The non-catalytic function of {XPG} protein during dual incision in human nucleotide excision repair. *The Journal of Biological Chemistry* **272**, 16030–16034 (1997).
- [34] Park, C. J. & Choi, B. S. The protein shuffle. Sequential interactions among components of the human nucleotide excision repair pathway. *FEBS Journal* **273**, 1600–1608 (2006).
- [35] Camenisch, U., Dip, R., Schumacher, S. B., Schuler, B. & Naegeli, H. Recognition of helical kinks by xeroderma pigmentosum group A protein triggers DNA excision repair. *Nature Structural & Molecular Biology* **13**, 278–284 (2006).
- [36] 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).
- [37] 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. *Molecular Cell* **27**, 311–323 (2007).
- [38] Green, C. M. & Almouzni, G. Local action of the chromatin assembly factor CAF-1 at sites of nucleotide excision repair in vivo. *EMBO Journal* **22**, 5163–5174 (2003).
- [39] Polo, S. E., Roche, D. & Almouzni, G. New histone incorporation marks sites of UV repair in human cells. *Cell* **127**, 481–93 (2006).
- [40] 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).
- [41] 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).
- [42] 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).
- [43] Rademakers, S., Volker, M., Hoogstraten, D., Nigg, A. L., Mone, 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).
- [44] Gorski, S. A., Snyder, S. K., John, S., Grummt, I. & Misteli, T. Modulation of RNA polymerase assembly dynamics in transcriptional regulation. *Molecular Cell* **30**, 486–497 (2008).
- [45] 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).
- [46] 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).

- [47] Houtsmuller, A. & Vermeulen, W. Macromolecular dynamics in living cell nuclei revealed by fluorescence redistribution after photobleaching. *Histochemistry and Cell Biology* **115**, 13–21 (2001).
- [48] 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).
- [49] 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).
- [50] Terstiege, G. Mammalian DNA repair: a model for multi-protein complex assembly on chromatin. Ph.D. thesis, University of Heidelberg (2010).
- [51] 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 repairstome. *Molecular and Cellular Biology* **21**, 2281–2291 (2001).
- [52] Houtsmuller, A. B. Action of DNA repair endonuclease ERCC1/XPF in living cells. *Science* **284**, 958–961 (1999).
- [53] 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).
- [54] 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).
- [55] 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).
- [56] Mu, D., Hsu, D. S. & Sancar, A. Reaction mechanism of human DNA repair excision nuclease. *Journal of Biological Chemistry* **271**, 8285–8294 (1996).
- [57] 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).
- [58] 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).
- [59] 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).
- [60] Winkler, G. S., Sugashawa, 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).
- [61] 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).

Bibliography

- [62] 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).
- [63] 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).
- [64] 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).
- [65] Owen, A. B. A central limit theorem for Latin hypercube sampling. *Royal Statistical Society* **54**, 541–551 (2014).
- [66] 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).
- [67] 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).
- [68] Venzon, D. J. & Moolgavkar, S. H. A method for computing profile-likelihood-based confidence intervals. *Applied Statistics* **37**, 87–94 (1988).
- [69] Murphy, S. A. & van der Vaart, A. W. On profile likelihood. *Journal of the American Statistical Association* **95**, 449–485 (2000).
- [70] 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* **25**, 1923–9 (2009).
- [71] Hofmeyr, J.-H. S. & Cornish-Bowden, A. Quantitative assessment of regulation in metabolic systems. *European Journal of Biochemistry* **200**, 223–236 (1991).
- [72] Fell, D. A. Metabolic Control Analysis: a survey of its theoretical and experimental development. *Biochemical Journal* **286**, 313–330 (1992).
- [73] Kreutz, C., Raue, A. & Timmer, J. Likelihood based observability analysis and confidence intervals for predictions of dynamic models. *BMC Systems Biology* **6**, 120 (2012).
- [74] Hinkley, D. Predictive Likelihood. *The Annals of Statistics* **7**, 718–728 (1979).
- [75] Pearson, K. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine* **2**, 559–572 (1901).
- [76] Scholzen, T. & Gerdes, J. The Ki-67 protein: from the known and the unknown. *Journal of Cellular Physiology* **322**, 311–322 (2000).
- [77] Burry, R. W. Controls for immunocytochemistry: an update. *The Journal of Histochemistry and Cytochemistry* **59**, 6–12 (2011).
- [78] Giepmans, B. N. G., Adams, S. R., Ellisman, M. H. & Tsien, R. Y. The fluorescent toolbox for assessing protein location and function. *Science* **312**, 217–24 (2006).
- [79] Allison, P. D. Multiple regression: A primer. Pine Forge Press (1999).
- [80] Smerdon, M. J., Tlsty, T. D. & Lieberman, M. W. Distribution of ultraviolet-induced DNA repair synthesis in nuclease sensitive and resistant regions of human chromatin. *Biochemistry* **17**, 2377–2386 (1978).

- [81] 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 Journal* **14**, 360–367 (1995).
- [82] 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).
- [83] Erdel, F. & Rippe, K. Binding kinetics of human ISWI chromatin-remodelers to DNA repair sites elucidate their target location mechanism. *Nucleus* **2**, 105–12 (2011).
- [84] Hager, G. L., McNally, J. G. & Misteli, T. Transcription dynamics. *Molecular Cell* **35**, 741–53 (2009).
- [85] 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).
- [86] Suter, D. M., Molina, N., Gatfield, D., Schneider, K., Schibler, U. & Naef, F. Mammalian genes are transcribed with widely different bursting kinetics. *Science* **332**, 472–4 (2011).
- [87] Conn, A. R., Scheinberg, K. & Vicente, L. N. Introduction to derivative-free optimization, vol. 8. Siam (2009).
- [88] Ramachandran, R. & Barton, P. I. Effective parameter estimation within a multi-dimensional population balance model framework. *Chemical Engineering Science* **65**, 4884–4893 (2010).
- [89] 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).
- [90] 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).
- [91] Kirschner, M., Gerhart, J. & Mitchison, T. Molecular ‘Vitalism’. *Cell* **100**, 79–88 (2000).
- [92] Cook, P. R. A model for all genomes: the role of transcription factories. *Journal of Molecular Biology* **395**, 1–10 (2010).
- [93] 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).
- [94] Takeuchi, O. & Akira, S. Pattern recognition receptors and inflammation. *Cell* **140**, 805–20 (2010).
- [95] Blüthgen, N. & Legewie, S. Robustness of signal transduction pathways. *Cellular and Molecular Life Sciences* **70**, 2259–69 (2013).
- [96] Csete, M. E. & Doyle, J. C. Reverse engineering of biological complexity. *Science* **295**, 1664–9 (2002).
- [97] 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).
- [98] 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).

Bibliography

- [99] Renaud, E., Miccoli, L., Zactal, 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).
- [100] Stewart, D. J. Mechanisms of resistance to cisplatin and carboplatin. *Critical Reviews in Oncology/Hematology* **63**, 12–31 (2007).
- [101] 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).
- [102] Moriya, H., Shimizu-Yoshida, Y. & Kitano, H. In vivo robustness analysis of cell division cycle genes in *Saccharomyces cerevisiae*. *PLoS Genetics* **2**, e111 (2006).
- [103] Apostolou, E. & Thanos, D. Virus Infection Induces NF-kappaB-dependent interchromosomal associations mediating monoallelic IFN-beta gene expression. *Cell* **134**, 85–96 (2008).
- [104] 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).
- [105] Alon, U. Network motifs: theory and experimental approaches. *Nature Reviews Genetics* **8**, 450–61 (2007).
- [106] 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).
- [107] 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).
- [108] 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).
- [109] Damia, G., Guidi, G. & Incalci, M. D. 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).
- [110] 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).
- [111] Kollmann, M., Løvdok, L., Bartholomé, K., Timmer, J. & Sourjik, V. Design principles of a bacterial signalling network. *Nature* **438**, 504–7 (2005).
- [112] Løvdok, L., Bentele, K., Vladimirov, N., Müller, A., Pop, F. S., Lebiedz, D., Kollmann, M. & Sourjik, V. Role of translational coupling in robustness of bacterial chemotaxis pathway. *PLoS Biology* **7**, e1000171 (2009).
- [113] Kendall, M. G. & Stuart, A. The advanced theory of statistics. McMillan, London, 4th ed. (1979).
- [114] Fisher, R. Statistical methods for research workers. Hafner, 13th ed. (1958).
- [115] Efron, B. Bootstrap methods: another look at the jackknife. *Annual Statistics* **7**, 1–26 (1979).

8. Abbreviations

6-4PP 6-4 pyrimidine-pyrimidone photoproducts, first appearance on page p. 2

ANSI American national standards institute, p. 23

AraC cytosine- β -arabinofuranoside, p. 11

AU arbitrary units, p. 14

BER base-excision repair, p. 2

CAF1 chromatin assembly factor 1, p. 3

ChIP chromatin immunoprecipitation, p. 20

χ^2 chi-squares statistic, p. 22

CI confidence interval, p. 27

CPD cyclobutane-pyrimidine dimer, p. 2

CV coefficient of variation, p. 7

DAPI 4',6-diamidino-2-phenylindole fluorescent stain, p. 49

DDR DNA damage response, p. 1

DNA deoxyribonucleic acid, p. 1

\in element of

EdU 5-ethynyl-2'-deoxyuridine, p. 7

eGFP enhanced green fluorescent protein, p. 9

ERCC1 excision repair cross-complementation group 1 protein, p. 3

FLIP fluorescence loss in photobleaching, p. 7

FOXP3 forkhead-box-protein P3, p. 62

GFP green fluorescent protein, p. 3

GG-NER global genome NER, p. 3

GHz gigahertz

HU hydroxyurea, p. 11

h hours

HR homologous repair, p. 2

ICL inter-strand crosslink, p. 2

IL-2 interleukin-2, p. 62

J Joule

K_d dissociation constant, p. 29

l liter

LHS Latin hypercube sampling, p. 23

LigIII ligase 3, p. 3

log logarithm

LD local damage, p. 9

NER nucleotide excision DNA repair

NHEJ non-homologous end joining, p. 2

MCMC Markov chain Monte Carlo, p. 59

μ micro

m meter

min minutes

M Molar

mRNA messenger RNA, p. 62

ODE ordinary differential equation, p. 23

PCA principal component analysis, p. 36

PCI prediction confidence interval, p. 33

PCNA proliferating cell nuclear antigen, p. 3

pl picolitre

PL profile likelihood , p. 26

PLE profile likelihood estimation, p. 26

PPL prediction profile likelihood, p. 33

r correlation coefficient, p. 35

R² coefficient of determination, p. 14

ℝ set of real numbers

ROS reactive oxygen species, p. 2

ℝ_{≥0} set of non-negative real numbers

˜R response coefficient, p. 31

RNA ribonucleic acid, p. 2

RNA Pol I RNA polymerase I, p. 4

SD standard deviation, p. 14

TC-NER transcription coupled NER, p. 3

TFIIFH transcription initiation factor IIH, p. 3

UV ultraviolet, p. 2

UV-C ultraviolet-C light (100-280 nm), p. 7

VIF variance inflation factor, p. 53

W Watt

Abbreviations

XP *Xeroderma Pigmentosum*, p. 3

XPA-G *Xeroderma Pigmentosum*, group A-G protein, p. 3

XP-C patient XP cells with impaired XPC protein, p. 12

XRCC1 X-ray repair cross-complementing protein 1, p. 3

A. Appendix

A.1. Parameter estimation

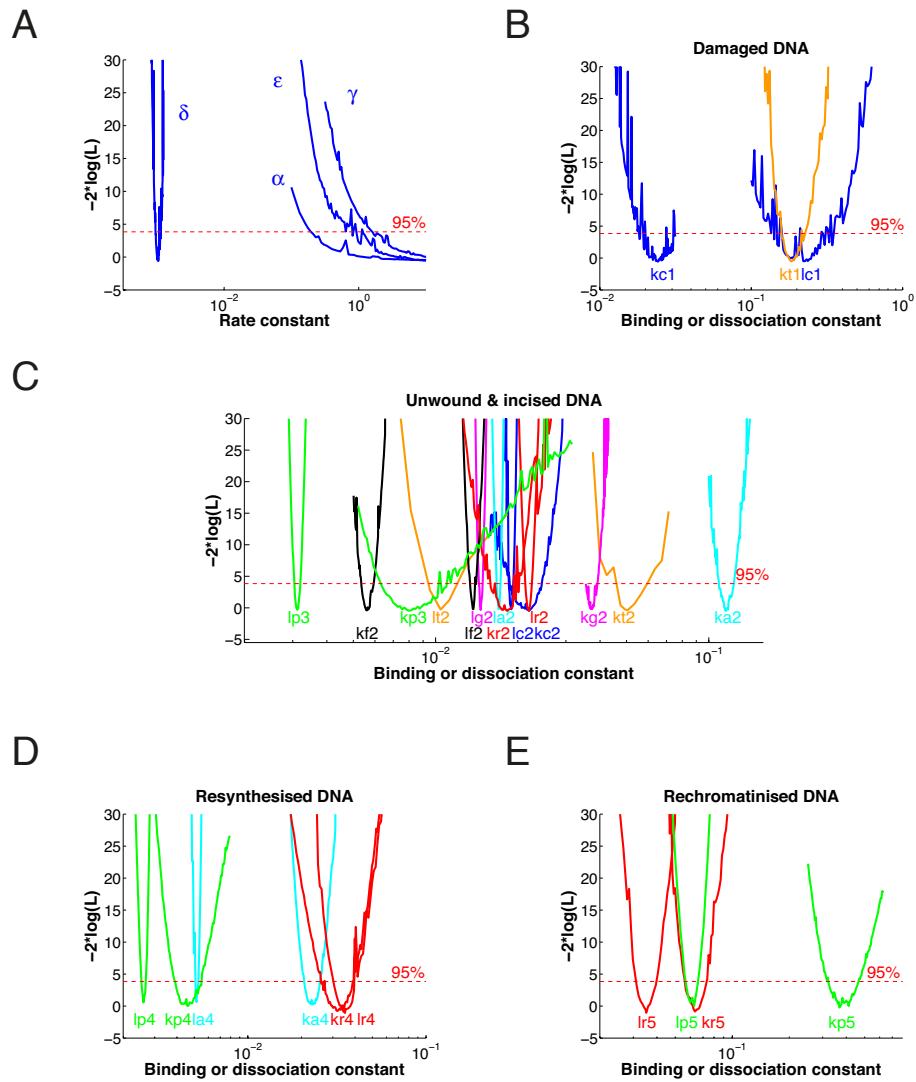


Figure A.1.: Identifiability analysis by profile likelihood estimation (PLE). PLE and 95% confidence limit (horizontal red line) of binding and dissociation parameters for A) all catalytic constants, B) damaged DNA, C) unwound and incised DNA, D) resynthesised DNA and E) rechromatinised DNA.

Value	XPC	TFIIL	XPG	XPF	XPA	RPA	PCNA
Concentration (μM)	0.140	0.360	0.440	1.110	0.170	1.110	1.110
Damaged DNA							
$k_{\text{on}}(\mu\text{M}^{-1}\text{s}^{-1})$	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
Unwound DNA							
$k_{\text{on}}(\mu\text{M}^{-1}\text{s}^{-1})$	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
Incised DNA							
$k_{\text{on}}(\mu\text{M}^{-1}\text{s}^{-1})$	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)
Resynthesised DNA							
$k_{\text{on}}(\mu\text{M}^{-1}\text{s}^{-1})$	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)
Rechromatinised DNA							
$k_{\text{on}}(\mu\text{M}^{-1}\text{s}^{-1})$	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.: Values of binding and dissociation rate constants. NA, not applicable. k_{on} , k_{off} and K_D ($k_{\text{off}}/k_{\text{on}}$) values are given for every repair protein and arranged in columns. Reference parameter set and 95% confidence intervals (in parentheses) are shown. Nuclear concentration (in μM) of NER factors are based on previously described data [6]

Catalytic rate	k_{cat} [s^{-1}]
Unwinding α	19.9 (>0.2)
Resynthesis γ	25.5 (>1.5)
Rechromatinization δ	0.001 (0.001;0.0011)
Reannealing ϵ	5.3 (>0.9)

Table A2.: Values of the catalytic rate constants. Reference parameter set and 95% confidence intervals (in parentheses) are shown. In case of practical non-identifiability only the lower confidence bound is given.

A.2. 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 H_0 states that the true correlation p_r between two variables equals zero $H_0 : p_r = 0$ (whereas the alternative hypothesis states $H_1 : p_r \neq 0$) and that for its observed value r the quantity

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

follows approximately a t -distribution for $N - 2$ degrees of freedom and a sample size N [113, 114]. 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 bootstrap [115]. 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.

A.3. Control analysis

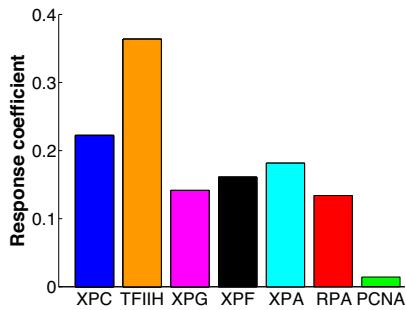


Figure A.2.: Uniformly distributed response coefficients for the rate of incision. Response coefficients for the control of the incision rate by the concentrations of the repair factors.

A.4. Cross-correlation

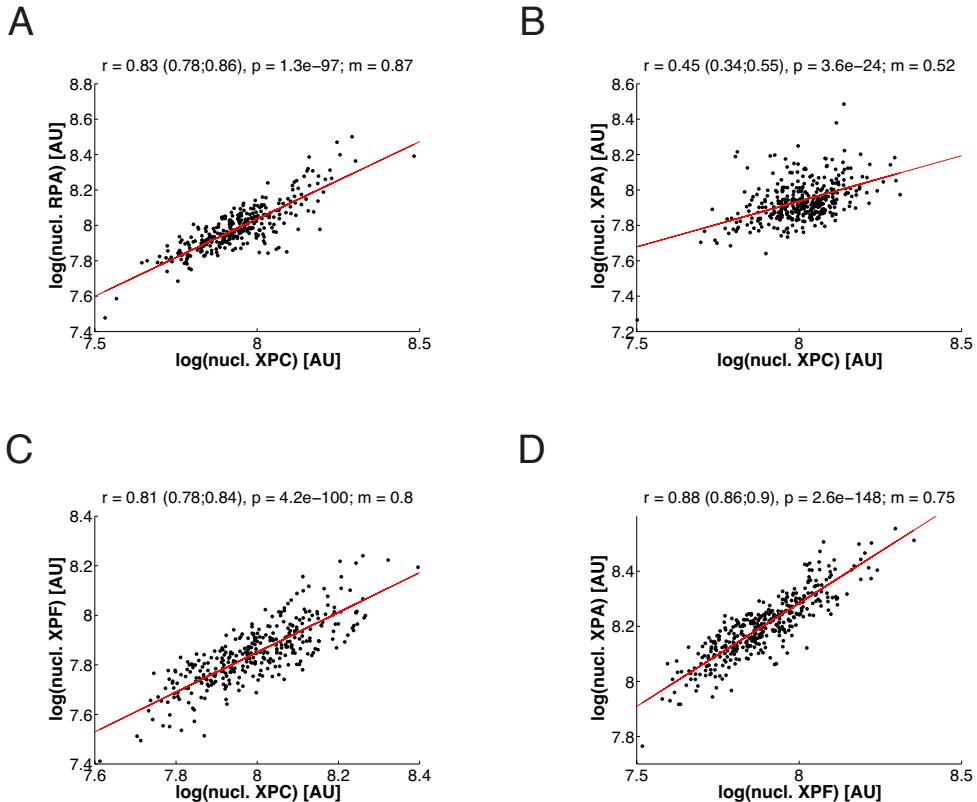


Figure A.3.: Nuclear expression of NER factors is correlated. Pairwise correlations of indirectly antibody-labelled RPA against XPC (n=383, A), XPA against XPC (n=464, B), XPF against XPC (n=425, C) and XPA against XPF (n=454 , D) in locally damaged cells. Expression values represent fluorescence intensities originating from the nucleus including the damaged region. 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.

Appendix

Species	Type	Protein	Manufacturer
Mouse	mAb	RPA70	Abcam
Mouse	mAb	XPC	Abcam
Mouse	mAb	XPF	Abcam
Rabbit	pAb	XPB	Abcam
Rabbit	pAb	XPA	Abcam

Table A3.: Primary antibodies. mAb = monoclonal antibody; pAb = polyclonal antibody

Species	Type	Fluorophore	Protein	Manufacturer
Donkey	IgG (H+L)	Alexa488	Mouse	Jackson Immunoresearch
Donkey	IgG (H+L)	Alexa647	Mouse	Jackson Immunoresearch
Donkey	IgG (H+L)	Alexa647	Rabbit	Jackson Immunoresearch

Table A4.: Secondary antibodies. IgG (H+L) = Immunoglobulin G (Heavy + Light chains)

Species	Type	Fluorophore	Protein	Manufacturer
Mouse	mAb	Alexa488	XPC	Abcam

Table A5.: Directly labelled antibody. mAb = monoclonal antibody