

INAUGURAL - DISSERTATION

zur
Erlangung der Doktorwürde
der
Naturwissenschaftlich-Mathematischen Gesamtfakultät
der
Ruprecht - Karls - Universität
Heidelberg

vorgelegt von

Diplom - Physikerin Gesa Terstiege geb. von Bornstaedt

aus: Bonn - Bad Godesberg

Tag der mündlichen Prüfung:

Mammalian DNA Repair: a Model for Multi-protein Complex Assembly on Chromatin

Gutachter: Prof. Dr. Karsten Rippe

Prof. Dr. Thomas Höfer

Contents

1	Introduction	1
1.1	Nucleotide Excision Repair	2
1.2	Fitting methods	4
1.2.1	Markov Chain Monte Carlo	4
1.2.2	Metropolis-Hastings Sampler	6
1.3	Research objectives	6
2	Nucleotide excision repair as a model system	9
2.1	In-vivo experiments of NER	9
2.1.1	Accumulation kinetics	10
2.1.2	Exchange of NER proteins measured by FLIP	12
2.2	Mathematical Model	13
2.2.1	Assembly strategies of NER	15
2.2.2	Model description	17
2.3	Fit of the model to the data	23
2.3.1	Results of the fit	24
2.3.2	Quality of the fit	27
2.4	Emerging properties of NER	30
2.4.1	Repair kinetics	30
2.4.2	Control coefficients	32
2.4.3	High capacity of NER	34
2.4.4	XPC kinetics in repair-deficient cells	36
3	Theoretical analysis of multiprotein complex assembly	39
3.1	Assembly and reaction times for multi-protein complexes	40
3.1.1	Irreversible binding of proteins	44
3.1.2	Reversible binding of proteins	44
3.1.3	Asymptotic behavior	45
3.2	Influence of cooperative binding on the reaction time	48

3.2.1	Mutual cooperativity	49
3.2.2	Next-neighbor cooperativity	50
3.3	Control of complex assembly	51
3.3.1	Contribution of individual off-rate constant	52
3.3.2	Control coefficients for identically binding proteins	54
3.4	Flux analysis	54
3.4.1	Random binding of two proteins	54
3.4.2	Random binding of three proteins	57
3.5	Optimal reversibility for finite protein concentrations	58
4	Specificity of recognition	61
4.1	Kinetic proofreading	62
4.2	Tradeoff between specificity and efficiency	65
5	Discussion	67
5.1	Efficiency	68
5.2	Specificity	69
5.3	Validation and limitation of the model	69
5.4	Conclusion: Multi-protein complex assembly on chromatin	70
5.5	Multi-protein complex assembly in other systems	71
6	Outlook: Possible continuations	73
6.1	Improving the quality of the model fit	73
6.2	Extending the NER model	74
6.3	Validating the model	75
	Bibliography	85
A	Appendix	87
A.1	Standard deviation σ for a sequential assembly	87
A.2	Reaction matrices for next-neighbor cooperativity	88
A.3	Explicit calculation of the control coefficients	94

Chapter 1

Introduction

To gain a better understanding of the assembly and functioning of a chromatin-associated protein machinery, we studied the mammalian nucleotide excision repair (NER) system, which removes UV-induced DNA damages and other DNA lesions from the genome in detail. On a more abstract level we focussed on the multi-protein complex assembly as such. To assess the importance of repair pathways, it is necessary to realize that the DNA is continuously damaged by heat, metabolic accidents, radiation of various sorts and exposure to environmental factors. The estimates of the number of DNA damage events in a single human cell ranges from $10^4 - 10^6$ per day [42]. There are several repair mechanisms to keep the mutation rate as low as 1 nucleotide change per 10^9 nucleotides each time the DNA is replicated [2]. A single repair pathway would not be able to cope with the large variety of DNA damages. The main repair pathways are mismatch repair, double strand break repair, base excision repair, direct reversal, non-homologous end joining and nucleotide excision repair, that is studied here. The mechanisms for DNA repair are highly conserved in evolution and the cells take a large effort to maintain DNA repair. In the genomes of bacteria and yeast several percent of the coding capacity is devoted to DNA repair [2]. The importance for humans can be deduced by looking at the severe inherited syndromes that are result of defects in one of the various DNA repair systems. For example people with xeroderma pigmentosum who have defects in nucleotide excision repair (NER) are prone to skin cancer, UV sensitivity and neurological abnormalities and there is variety of other syndromes caused by defects in DNA repair. The presence of DNA damages does not only result in unwanted mutations, but it restricts the correct functioning of transcription and replication as well.

1.1 Nucleotide Excision Repair

Nucleotide excision repair removes a variety of helix-distorting DNA lesions, including the UV-induced pyrimidine-pyrimidone (6-4) photoproducts (6-4 PPs) and cyclobutane pyrimidine dimers (CPDs). Placental mammals are fully dependent on NER to remove UV-damages from their genome. Mammalian NER follows a simple mechanism: i) damage recognition, ii) unwinding of the DNA, iii) excision of the damaged strand and iv) re-synthesis and ligation (Figure 1.1).

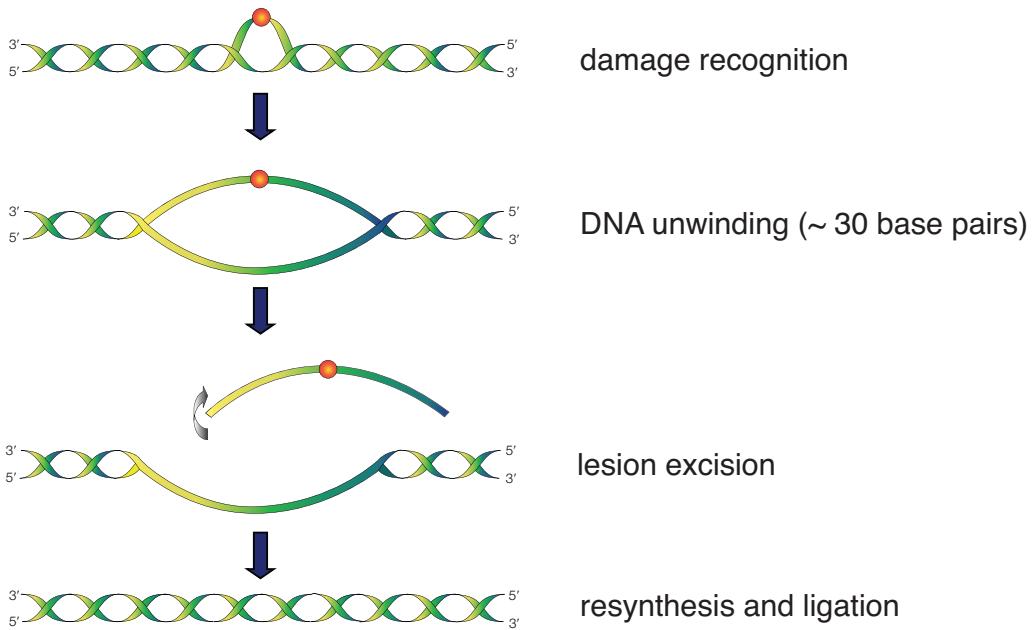


Figure 1.1: Basic mechanism of NER.

NER has two distinct sub-pathways: Global genomic NER (GG-NER) and transcription coupled (TC-NER). GG-NER removes lesions throughout the whole genome and TC-NER only from actively transcribed genes [48]. In TC-NER an RNA polymerase that is stalled by a DNA lesion recruits a small protein (CSB) that triggers the opening of the DNA helix and recruits the helicase TFIIH. After recognition TC-NER follows the same steps as GG-NER. TC-NER is much less abundant than GG-NER, since only a small fraction of the genome of every cell is transcribed in higher eukaryotes [42]. For that reason we will study only GG-NER in our model.

In global genome NER the DNA lesions are recognized by the XPC-HR23B complex [76, 83]. The bound XPC recruits subsequently the TFIIH complex that utilizes its helicase activity to unwind the DNA near the lesion [8]. The proteins RPA, XPA and the 3' endonuclease XPG stabilize the locally unwound DNA [23]. In this open confirmation

the single-strand binding protein RPA binds to the DNA strand opposite to the damage, XPA associates with the DNA lesion, and the endonucleases XPG and ERCC1/XPF excise about 30 nucleotides containing the DNA lesion [7, 15, 23, 60, 85]. DNA polymerase δ is subsequently loaded by PCNA to fill in the single-stranded gap in the DNA, which is sealed by LigIII-XRCCI [21, 31, 54]. In the end, CAF1 assembles new histones on the resynthesized DNA to restore the chromatin structure and to finalize the repair process. [28, 64].

In vivo NER requires the orchestrated action of about 30 polypeptides [16]. However, the proteins XPC, TFIIH, XPA, XPG, RPA and ERCC1/XPF are sufficient to perform dual incision and together with RF-C, PCNA, pol ε and DNA lig 1, required for repair synthesis, NER was successfully reconstituted in vitro [1]. A brief description of the essential proteins and their function will be given in the following.

XPC - damage recognition factor in GG-NER

The global genome NER (GG-NER) in mammalian cells is initiated by the damage recognition factor XPC. In agreement, there is no recruitment of TFIIH, XPA, XPG and ERCC1-XPF to UV-induced lesions in XP-C deficient cells *in vivo* [83]. Instead of binding to the lesion directly, XPC binds to the non-damaged strand opposite to it. This binding behavior allows to recognize a wide variety of lesion that are structurally unrelated [75].

TFIIH - DNA helicase

Transcription factor II H (TFIIH) is involved in various processes including RNA polymerase I and II as well as NER [19, 24, 36]. With ~ 700 kDa and 10 subunits is TFIIH the biggest factor involved in NER. In GG-NER XPC recruits TFIIH via direct protein-protein interactions [3] and TFIIH mediates thereafter the initial opening of the DNA (8 - 10 nucleotides). Afterwards RPA, XPA and XPG bind and lead together with TFIIH to the full open complex formation (~ 30 nucleotides) [22, 23, 57]. The TFIIH subunits XPB and XPD are ATP-dependent DNA helicases [69, 70, 78, 86]. Live-cell imaging of GFP-tagged XPB that is incorporated in functional TFIIH complexes shows that TFIIH shuttles between RNA pol I and II transcription and NER [33]. The subunit composition of TFIIH changes when binding to a DNA lesion compared to its involvement in transcription [5, 25]. In particular, TFIIH may also have a different affinity for damaged and undamaged DNA and thereby contribute to the damage recognition in NER [18, 82].

XPG - 3' endonuclease

The Xeroderma pigmentosum group G (XPG) protein is a 133 kDa endonuclease that

mediates the 3' incision during NER [59]. The 5' incision performed by ERCC1/XPF requires the presence of XPG, even if it is catalytically inactive [84].

XPA - damage verification factor

Xeroderma pigmentosum group A (XPA) protein (36 kDa) is a DNA binding zinc metalloprotein. The binding of GFP-tagged XPA *in vivo* is strictly dependent on XPC[66]. XPA deficient cells are unable to perform dual incision that suggests that XPA has an important role in the coordination of incision. Purified XPA has a higher affinity for damaged over undamaged DNA [37, 68] the affinity for damaged DNA is amplified by forming a complex with RPA [41]. Accordingly, XPA has the role of damage verification factor.

ERCC1/XPF - 5' endonuclease

The second endonuclease is a heterodimer consisting of ERCC1 (33 kDa) and XPF (103 kDa) that cleaves the DNA at the 5' side.

RPA- stabilizing factor

Replication protein A (RPA, 121kDa) has very high affinity for single-stranded DNA and is consequently part of various processes that contain single-stranded DNA, e.g. DNA replication, recombination and repair pathways. Bound RPA protects the single-stranded DNA from endonucleases and stabilizes its structure. A proper positioning and stimulation of the two endonucleases XPG and ERCC1/XPF in NER depends on RPA [14, 15], [29, 47] and is thus required for dual incision that removes the lesion [11]. Isolation of repair intermediates *in vitro* showed that RPA remains associated with the excised DNA during dual incision [67].

PCNA - part of the repair synthesis machinery

Proliferating Cell Nuclear Antigen (PCNA) is a sliding clamp encircling the DNA that acts as a processivity factor for DNA polymerase δ/ε [45, 62]. PCNA is important for both DNA synthesis and DNA repair [72].

1.2 Fitting methods

1.2.1 Markov Chain Monte Carlo

We used a Markov Chain Monte Carlo (MCMC) method to estimate the parameters of the model and their uncertainty. This method samples from a probability distribution by

generating a Markov chain that has the wanted distribution as equilibrium distribution.

Let D be a given set of data and let $P(D|x)$ be the probability of the data set given the values of the model parameters x . Assuming a prior distribution $P(x)$, Bayes' theorem yields that the posterior probability of the model is

$$P(x|D) = \frac{P(D|x)P(x)}{\int P(D|x)P(x)dx} \quad (1.1)$$

Even if the normalization constant $\int P(D|x)P(x)dx$ is unknown it is possible to sample from the distribution and to obtain relevant variables such as the mean and the variance. This sampling is not done randomly, but rather by a Markov chain.

A Markov chain with transition matrix T is a sequence Z_0, Z_1, \dots of random variables, such that

$$\begin{aligned} P(Z_n = x_n | Z_{n-1} = x_{n-1}, \dots, Z_0 = x_0) &= P(Z_n = x_n | Z_{n-1} = x_{n-1}) \\ &= T_{x_n, x_{n-1}} \end{aligned} \quad (1.2)$$

Let $\pi_x(t)$ be the probability to be in state x at time point n and the vector $\vec{\pi}$ comprises the probabilities for all states. It follows that

$$\pi(n) = T^n \pi(0). \quad (1.3)$$

A Markov chain converges to a stationary distribution π^s , where the probabilities of being in a particular state are not dependent of the initial conditions, if the chain is irreducible and aperiodic. Irreducibility guarantees a positive probability that the chain can reach any non-empty set from all starting points and aperiodicity excludes that the chain oscillates between different sets of states. The stationary distribution has the eigenvalue 1 and satisfies thus

$$\pi^s = T \pi^s \quad (1.4)$$

The stationary distribution is unique if the Markov chain fulfills the detailed balance equation

$$T_{i,j} \pi_j^s = T_{j,i} \pi_i^s. \quad (1.5)$$

Given that the chain already did m iterations and assuming that the sample points x_n with $n = m + 1, \dots, m + k$ are distributed according to the stationary distribution π^s , the first m iterations can be discarded. This phase is called "burn-in". The remaining k samples can be used for a statistical analysis. In particular it is possible to determine 90% confidence

intervals. The complicated part is however to determine this "burn-in" time.

1.2.2 Metropolis-Hastings Sampler

The Metropolis-Hastings Sampler constructs a Markov chain that converges to the wanted distribution. Each step samples a candidate point \vec{y} from a proposal distribution $q(\cdot | \vec{x}_t)$ depending on the current state. It can be any distribution, as long a sequence of steps generated by it can, in principle, reach any point in the region of interest. In our algorithm we use a standard normal distribution around the current point in the chain.

More precisely the procedure goes as follows.

The candidate point \vec{y} is accepted as the next state of the chain with the following probability

$$\alpha(\vec{x}_t, \vec{y}) = \min \left\{ 1, \frac{P(\vec{y}|D) q(\vec{x}_t|\vec{y})}{P(\vec{x}_t|D) q(\vec{y}|\vec{x}_t)} \right\} \quad (1.6)$$

If the point \vec{y} is not accepted the chain keeps the former value, i.e. $\vec{x}_{t+1} = \vec{x}_t$. The unknown normalization constant of the distribution $P(x|D)$ cancels out.

Algorithm

1. At $t = 0$ initialize the chain to \vec{x}_0
2. Generate a candidate point \vec{y} from $q(\cdot | \vec{x}_t)$
3. Generate a value γ from a uniform $(0, 1)$ distribution
4. If $\gamma \leq \alpha(\vec{x}_t, \vec{y})$ then set $\vec{x}_{t+1} = \vec{y}$, else set $\vec{x}_{t+1} = \vec{x}_t$
5. Set $t = t + 1$ and repeat steps 2 through 5

In this way we construct a sampling of a Markov chain that converges to the wanted distribution d . Hence for large t the algorithm automatically produces preferentially samples x_t for which $d(x_t)$ is large and thus the parameters are optimal. After subtracting the "burn-in" period, it is possible to determine 90% confidence intervals of the parameters. Moreover, it offers the option to analyze the correlations of the individual parameters.

1.3 Research objectives

Proteins usually act in complexes with other proteins, RNA or small molecules. Typical examples are chromatin-associated processes such as DNA replication, transcription and

repair. *In vitro* studies on NER have been essential in defining the core repair factors and their enzymatic action, but they can not elucidate the dynamic binding of NER factors to a chromatin substrate [1, 59, 67], [70, 74, 79]. Other chromatin associated processes have a similar background, such that many questions on the formation process and the regulatory function of the assembly remain open. The components of multi-protein complexes involved in DNA transcription, replication and repair exchange rapidly between chromatin-bound and freely diffusing states [20, 27, 35, 50]. However, up to now there is no consistent model that combines the ordered formation of chromatin-associated protein machineries with the rapid exchange of their individual components. These observations led to the following questions that we want to address with our mathematical model:

- How do the components of multi-protein machineries assemble and how can the ordered formation of the complexes be reconciled with the rapid exchange of their components? Does the complex assembly follow a sequential or a random strategy?
- How efficient is a random complex assembly compared to a sequential one? What limits the efficiency?
- Does the assembly strategy serve a regulatory function? How is the high specificity of target recognition achieved?

To better understand the assembly and functioning of protein complexes on chromatin, we started by studying the mammalian nucleotide excision repair (NER), which removes UV-induced DNA damages and other DNA lesions from the genome in chapter 2. NER shows the typical organization of a chromatin-associated process, starting with the recognition of a target site (e.g. a DNA lesion), followed by the assembly of a multi-protein complex and finally the enzymatic action of the machinery on the DNA substrate [17], [26, 31]. Guided by many *in-vivo* experiments, we developed a mathematical model of the NER process based on reversible protein binding and random complex assembly that could be fitted to the accumulation and dissociation kinetics of seven core NER proteins simultaneously. This model comprised the longterm accumulation and the fast exchange of the components. A further analysis of the emerging properties revealed that the random complex assembly from reversibly binding proteins is both efficient and ensures a high specificity.

The specific model of Nucleotide Excision Repair that was tightly connected to *in-vivo* experiments and revealed many properties of multi-protein complex assembly, however it also raised more general questions on the efficiency and regulatory function of the assembly process. In chapter 3 we show that a random complex assembly is more efficient

for a limited number of components and in chapter 4 we finally demonstrate that the reversible protein binding that enables a kinetic proofreading mechanism ensures the high specificity, observed in the target recognition of NER and other processes.

The modeling work presented here was done in close collaboration with the experimental group of Prof. Dr. Roel van Driel in Amsterdam and in particular most of the experiments that were used to analyze the process and to validate our model predictions were performed by Dr. Martijn Luijsterburg.

Chapter 2

Nucleotide excision repair as a model system

2.1 In-vivo experiments of NER

To motivate and to facilitate the understanding of our work, we will give a short overview of the experimental techniques used to investigate NER in general and especially of the live-cell imaging experiments that were the basis of the kinetic model presented here.

In these experiments NER is initiated by irradiating cells with ultraviolet light (UV-light). In local damage experiments only a small fraction of the nuclear volume ($\sim 1/7 - 1/10$) is irradiated. The local damage is created by UV-C (wavelength of 280nm - 100nm) irradiation at 100 Jm^{-2} through $5 \mu\text{m}$ pores of a polycarbonate mask [43, 52]. A scheme of the experimental setup is shown in Figure 2.1A and a merged picture of the pore mask and two cell nuclei in Figure 2.1B.

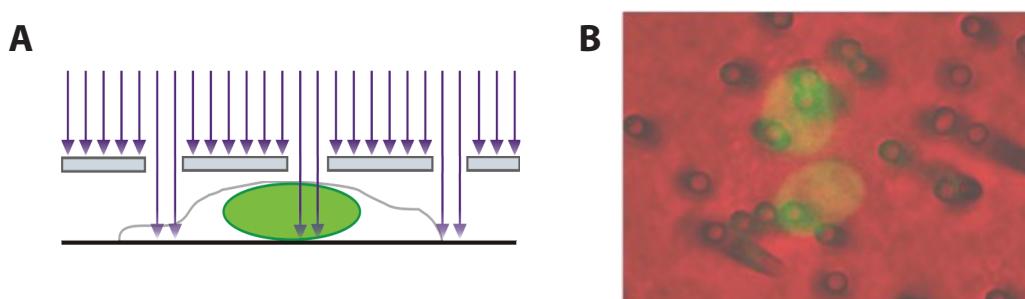


Figure 2.1: Experimental Setup. (A) Sideview of the experimental setup of a local damage experiment. (B) Merged picture of the pore mask (red) and two cell nuclei (green).

UV-C irradiation introduces several structurally related DNA lesions via photochemical

reactions that dimerize pyrimidine bases (thymine and cytosine). The most common ones are the 6-4 photoproducts (6-4PPs) and the more abundant cyclobutane pyrimidine dimers (CPDs) [4, 81]. The experiments reported here were done during the time span of 6-4PP repair, in which CPD repair can largely be neglected. There is hardly any CPD repair in the first 4 hours (Figure 2.2). The local irradiation of the cell nucleus generates approximately 60,000 DNA lesions (6-4 photoproducts) per irradiated area [53].

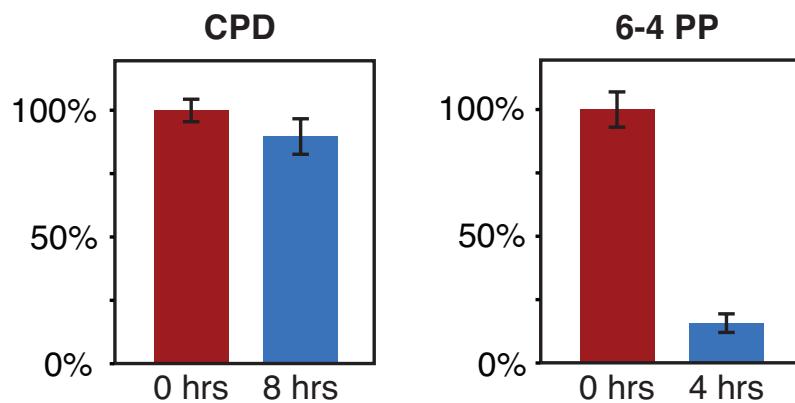


Figure 2.2: CPD vs. 6-4PP repair. There is almost no CPD repair within in the first 8 hours, but most of the 6-4PPs are already removed after 4 hours. The evaluation of the removal of CPDs and 6-4 PPs is done by quantitative immunostaining using specific antibodies, directly after UV irradiation (0 hrs) and 4 hrs (for 6-4 PP) or 8 hrs (for CPD) after UV-C irradiation. The value for each time-point is averaged over 50 to 70 cells.

To analyze the kinetics of the NER process in living cells, seven core NER proteins were tagged with green fluorescent protein (GFP or EGFP for the GFP-derivative "enhanced GFP") and the fusion proteins were expressed in NER-deficient cells or wild type cells at physiological levels. The GFP-tagged NER proteins complement the UV-sensitive phenotype of NER-deficient cells, demonstrating their functionality [21, 32, 33, 35, 66, 90].

2.1.1 Accumulation kinetics

Before the cell is irradiated there is a homogenous distribution of fluorescent labeled repair protein. After the irradiation the fluorescence in the damaged area increases (cf. Figure 2.3A). The start of the UV irradiation is defined as time point zero ($t = 0$). The quantification of the bound fraction of GFP-tagged proteins in the local damage was done as follows:

$$\text{Bound}(\%) = (I_{\text{spot}} - I_{\text{outspot}}) * \text{pixels}_{\text{spot}} / (I_{\text{nucleus}} - I_{\text{background}}) * \text{pixels}_{\text{nucleus}}, \quad (2.1)$$

where I_{spot} and I_{outspot} are the average pixel intensities inside the damaged spot and outside the spot, respectively. I_{nucleus} is the average pixel intensity of the whole nucleus, including the spot and $I_{\text{background}}$ is the average pixel intensity outside the cell. This procedure is necessary to correct for freely diffusing proteins in the nucleus that cannot be distinguished from the bound proteins. In a final step, the time courses are normalized with respect to the plateau level.

The time course for the accumulation kinetics were measured throughout the repair process for: (i) the lesion-recognition factor XPC, (ii) components of the pre-incision complex that excise the lesion (TFIILH, XPG, XPA, ERCC1/XPF), and (iii) proteins involved in the repair-synthesis (RPA, PCNA) throughout the repair process (Figure 2.3). The analysis of RPA and PCNA was done on a selection of cells that was not undergoing S-phase to assure that binding is not due to DNA replication. These repair synthesis proteins also accumulated at sites of DNA damage in quiescent cells, again confirming that the binding reflects engagement in DNA repair and not DNA replication. In order to compare the kinetics and accumulation levels of the different proteins, the curves have to be rescaled by the bound fraction and the nuclear concentration of the respective protein (cf. Table 2.1 for the nuclear concentrations). After the rescaling, we see that the pre-incision proteins XPA, XPG (the 3' endonuclease) and RPA accumulate to higher levels in the damaged area than the lesion-recognition protein XPC (Figure 2.3B and C), suggesting that the composition of NER complexes changes during the NER process.

Protein	Concentration	Bound fraction
XPC	0.140 μM	13%
TFIILH	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 concentration (in μM) of NER factors. The nuclear quantities of NER factors XPC, XPA and XPG are based on published data [3], while RPA and PCNA amounts are estimated to be 250,000 molecules per cell, and TFIILH and ERCC1/XPF were estimated at 65,000 and 50,000 molecules per cell, respectively, based on previous estimates [35, 52]. The nuclear concentrations are calculated assuming a nuclear volume of 0.3 pL. The bound fraction is determined by Equation (2.1) [44].

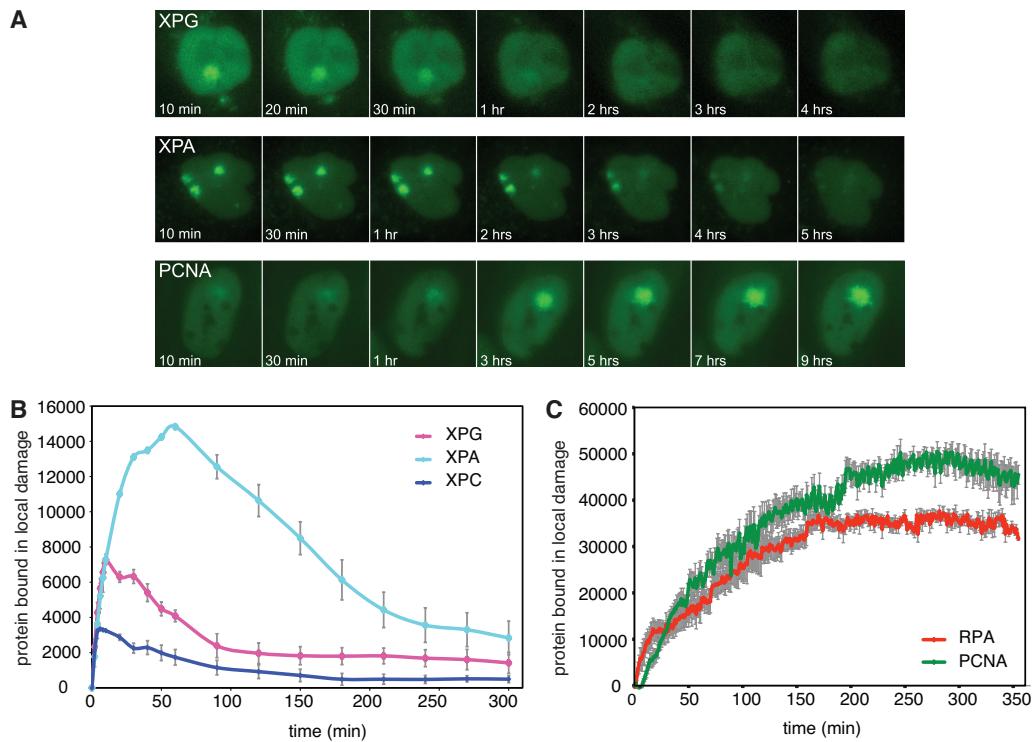


Figure 2.3: Long-lasting net accumulation at sites of DNA damage. (A) Cells expressing XPG-EGFP, EGFP-XPA and EGFP-PCNA shown at various times after local UV-C irradiation (100 J m^{-2} through $5 \mu\text{m}$ -diameter pores). (B) Quantification of bound XPC-EGFP, XPG-EGFP and EGFP-XPA following irradiation. (C) Quantification of bound EGFP-PCNA and RPA-EGFP.

2.1.2 Exchange of NER proteins measured by FLIP

The dwell times of the NER proteins at sites of DNA damage were measured by fluorescence loss in photobleaching (FLIP). When the accumulation of the repair proteins in the local damage reaches its plateau level, the experiments were started by continuously bleaching a stripe opposite to the lesion and simultaneously recording the loss of fluorescence in the local damage (cf. Figure 2.4A; the red rectangle indicates the bleaching area). When a GFP tagged repair protein dissociates from the sites of DNA damage, it can freely diffuse in the nucleus and will be bleached, when it enters the bleaching area (red rectangle). The experimental conditions are such that dissociating proteins have a high probability to be bleached before rebinding to a site of damage, i.e. after a few seconds there are no more fluorescent proteins outside the damaged area. Comparison of FLIP kinetics of proteins accumulated in the damaged area and of proteins outside the irradiated area at a similar distance from the bleaching area show that the FLIP kinetics for the latter were at least one order of magnitude faster. This observation implies that binding and not

diffusion is rate-limiting for the dwell time of the NER proteins in the damaged area.

All EGFP-tagged pre-incision proteins dissociate rapidly from repair complexes, with overall half-lives of 20s (RPA), 25s (XPC), 50s (TFIIH, XPG, ERCC1/XPF) and 80 s (XPA) (Figure 2.4A and B). The dissociation kinetics of the repair synthesis factor PCNA is strongly biphasic, with half-lives of 10s and 225s for the two components of the exponential decay (Figure 2.4A and C). Thus, all measured NER factors exchange rapidly between the freely diffusing and bound states.

To get a better insight into the composition of the repair complexes during repair, we check how a perturbed repair process affects the dwell times of NER proteins. Blocking NER before lesion excision, either by impaired unwinding in cells lacking functional XPB, XPA or XPG or by impaired excision in cells with compromised ERCC1 or XPF [23], leads to an about four-fold slower XPC dissociation than in wild-type cells. This result suggests that XPC binding is de-stabilized after the pre-incision complex is assembled and dual incision has occurred (Figure 2.4D). In order to determine the possible affinity change caused by repair synthesis, we look at experiments where the repair synthesis is blocked by the addition of drugs (hydroxyurea (HU) and cytosine- β -arabinofuranoside (AraC) [58]). We assume that this treatment has the biggest effect on the proteins that accumulate over an extended period of time. The affect on the dissociation of XPA and PCNA is shown in Figure 2.4C and E. XPA dissociation is about two-fold faster if DNA synthesis and ligation is inhibited, showing that XPA binds to repair-synthesis intermediates with high affinity. Dissociation of PCNA is slower in the presence of the inhibiting drugs, indicating its preferential binding to incised DNA [73]. However, the same treatment has no effect on the dissociation kinetics of XPC and ERCC1/XPF shown in Figure 2.4E. Thus, in contrast to the other pre-incision proteins, XPA binding becomes stabilized after repair synthesis. These changes in the FLIP kinetics show that the dwell times of NER proteins are altered as repair progresses. We conclude thus, that the affinity of the repair proteins depends on the different states of the DNA substrate.

2.2 Mathematical Model

The *in vivo* experiments show that the kinetics of all core NER proteins act on two very different time scales. The repair factors accumulate at sites of damage on a time scale of hours. In contrast to that, all core NER proteins exchange continuously and rapidly, on a sub-minute time scale between chromatin-bound and freely diffusing states. To solve this

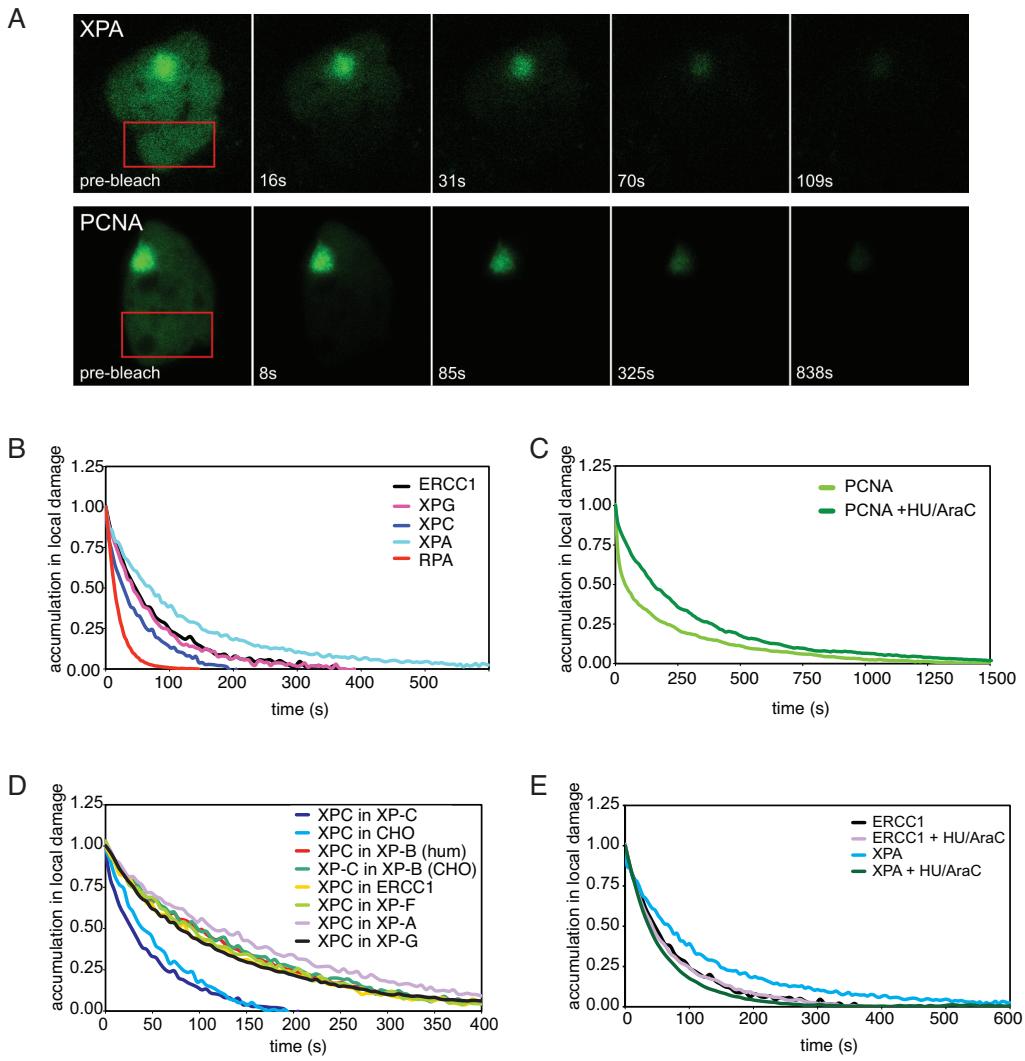


Figure 2.4: Rapid exchange of NER proteins at sites of DNA damage. (A) FLIP (fluorescence loss in photobleaching) measurements in cells expressing EGFP-XPA and EGFP-PCNA. 1h and 2h after local irradiation, respectively, the cells are continuously bleached in the undamaged region (red rectangle). Quantification of FLIP experiments in cells expressing (B): XPC-EGFP, XPG-EGFP, EGFP-XPA, RPA-EGFP, ERCC1-GFP, and (C): EGFP-PCNA. Quantification of FLIP experiments with perturbations of NER: (D) on XPC- mVenus expressed in various locally irradiated NER-deficient CHO and human cell lines and (E) in the absence or presence of HU and AraC on locally irradiated cells expressing EGFP- XPA or ERCC1-GFP.

paradox we propose a mathematical model where the repair proteins assemble stochastically and reversibly to form distinct complexes that catalyze the successive enzymatic reactions of the NER process, carrying out DNA unwinding, dual incision, and repair synthesis.

Previous biochemical and *in vivo* studies of NER have shown that the repair of a DNA lesion proceeds through a series of distinct repair intermediates [23, 56, 64, 72, 79]. In

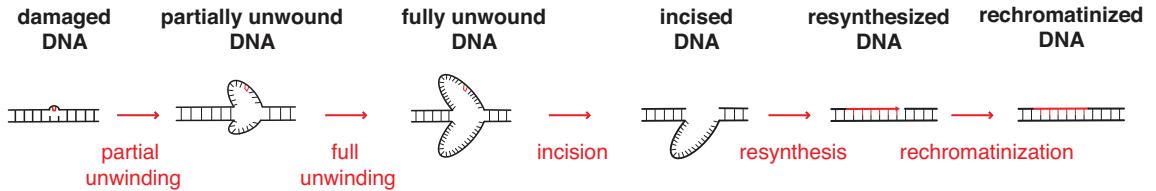


Figure 2.5: Repair intermediates of NER. The different states of the DNA substrate (repair intermediates) are interconverted by a series of enzymatic reactions (red arrows).

addition, we assume that the DNA adjacent to the lesion is unwound in at least two steps [23]. The scaffold of the model is thus given by the following six DNA repair intermediates: damaged, partially unwound, fully unwound, incised, resynthesized and rechromatinized DNA (Figure 2.5). For the formation of the multi-protein repair complexes on the repair intermediates, we assume a predominantly stochastic process in which proteins can associate and dissociate independently of each other and in any order as soon as the DNA becomes partially unwound. Whenever a catalytically active protein complex is fully assembled it can catalyze the transition to the subsequent repair intermediate, e.g., the pre-incision complex with the two endonucleases XPG and ERCC1/XPF excises a DNA fragment adjacent to the lesion. In that way, our model accounts for both the highly reversible protein binding and the irreversible enzymatic reactions that determine the sequence of NER.

2.2.1 Assembly strategies of NER

The sequence of enzymatic steps executing NER is well known. However, the assembly of the multi-protein complexes that catalyze the steps of repair is yet unclear. Previous models have suggested that the repair factors assemble into stable multi-protein complexes through a sequential assembly mechanism [51, 63, 83]. In this framework, the individual proteins stay in the stable complex bound to the DNA during the whole repair process and they are only released after the execution of the enzymatic steps. The other possible assembly strategy is a random complex assembly, where the repair factors continuously bind to and dissociate from repair complexes while the enzymatic reactions proceed. In this framework, the composition of the repair complexes changes all the time instead of building a single stable complex at the lesions containing all repair factors. Completely sequential and random assembly mechanisms are the extremes of a spectrum of potential assembly mechanisms. Any mixed assembly strategy could as well be realized. For illustration we show the possible binding pathways for a random assembly of the pre-incision complex on unwound DNA in Figure 2.6. A selected sequential pathway is shown in red for comparison. The marked contrast between the different assembly strategies raises the

question, whether experiments can exclude one or the other strategy.

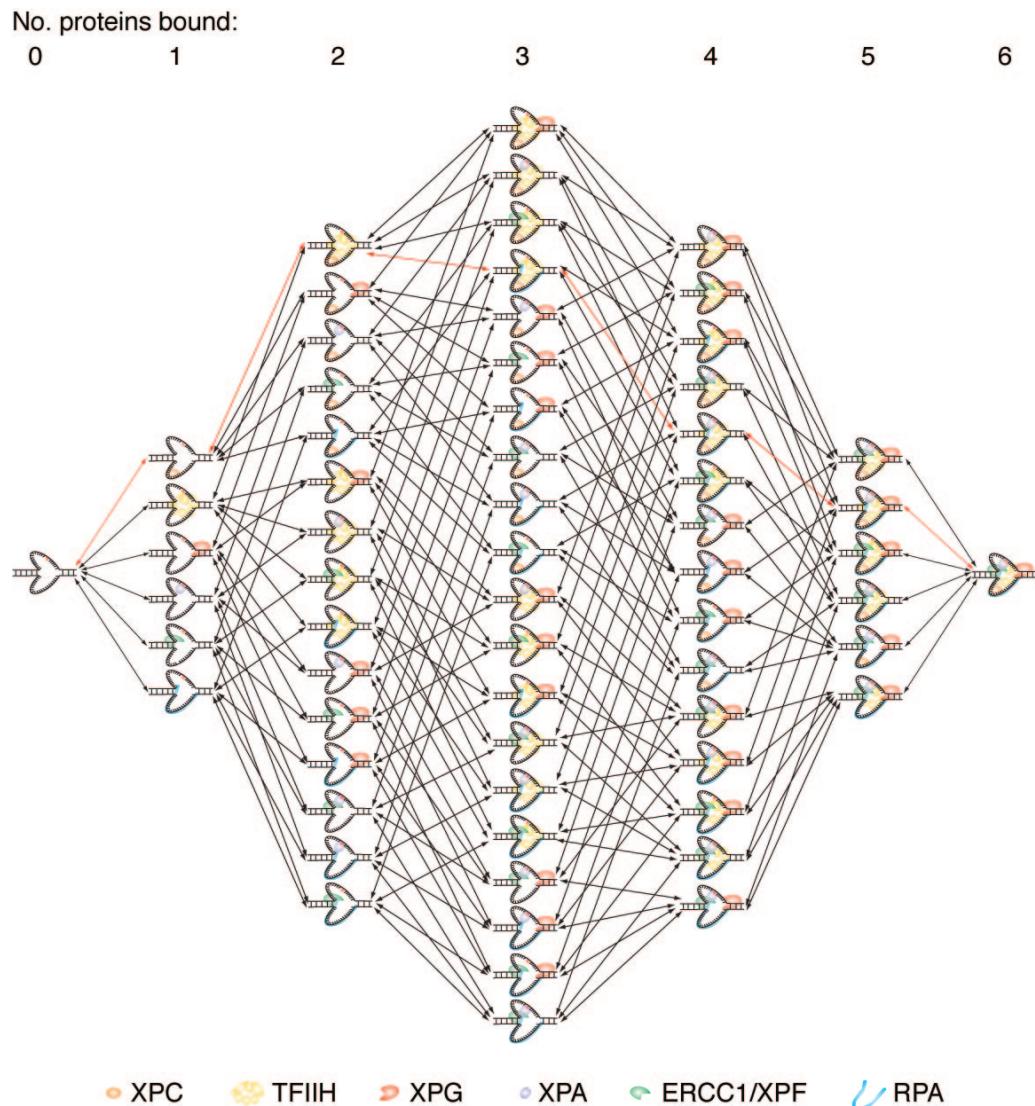


Figure 2.6: Possible assembly pathways for the pre-incision complex on unwound DNA. Random assembly can utilize all pathways shown, while sequential assembly will follow a unique pathway (e.g., the pathway indicated by the red arrows assuming ordered binding of XPC, TFIID, RPA, XPA, XPG, and ERCC1/XPF).

The accumulation experiments show that NER proteins are engaged in repair for several hours. However, the pre-incision proteins XPA, XPG (the 3' endonuclease) and RPA accumulate in the locally damaged area to higher levels compared to the lesion-recognition protein XPC (Figure 2.3). This observation argues against the recruitment of proteins into a stable pre-incision complex together with XPC at a 1:1 stoichiometry. In addition, the maximal accumulation of the proteins occurs at different time-points. These two findings indicate that the average molecular composition of the NER complexes changes during

the repair process. Nevertheless, these indications are not enough to exclude a stable binding of repair factors building long-lived complexes. The only way to assess, if the repair factors are rather rapidly and continuously exchanged while repair of a lesion is being carried out, is to look at the dwell times of the proteins. All measured NER factors exchange rapidly between freely diffusing and states bound to the chromatin. On average they are part of a repair complex for few tens of seconds only (Figure 2.4). Moreover, the experiments on the perturbed NER process show that the dwell times of NER proteins change as repair progresses, suggesting that the affinity of the repair proteins is primarily determined by the state of the DNA substrate.

A sequential assembly that is a strict binding order of the proteins to the repair intermediates would imply the stabilization of early-binding proteins by the subsequent proteins incorporated into the complex, resulting in long dwell times of early-binding proteins compared to short dwell times of late-binding proteins. Thus, the recruitment of proteins in a strict order is incompatible with the observed mutually independent and rapid dissociation of individual NER factors. In contrast to that, a random complex assembly of the repair proteins can account for both rapid exchange and slow net accumulation of NER proteins at sites of damage and resolve thereby the conflicting demands of the experiments.

2.2.2 Model description

The general structure of the model and the involved proteins are depicted in the model scheme shown in Figure 2.7. The scaffold of the model is given by randomly assembled protein complexes that catalyze the steps of repair. The experiments described in the previous section and other studies (cf. Table 2.2) allow us to deduce the composition of the enzymatically active multi-protein complexes that catalyze the transitions between the repair intermediates. In Table 2.2, we summarize all model assumptions and the experimental findings they are based on. For each repair intermediate we list the proteins that have affinity for it and indicate the proteins required to perform the subsequent enzymatic step, as well as known cooperative binding events. Nucleoplasmic diffusion of NER factors need not be considered explicitly, since it is much faster than the characteristic times for exchanging proteins at damage sites [66, 90].

The initiation of NER is of particular interest. Experiments show that the damage recognition factor XPC binds to DNA lesions, but none of the other NER proteins analyzed here accumulates in XPC deficient cells, suggesting that XPC is the first factor to bind and thus

the initiation factor of the NER process. This scenario with a single initiation factor is supported by other theoretical findings showing that the initial recognition step of the lesion should be sequential to ensure efficient initiation of NER at high amounts of DNA damage [63]. If several proteins started the repair process the protein pools would be used up in started but unfinished and yet catalytically inactive complexes, thereby slowing down the overall repair process significantly. Thus we assume that DNA lesions are recognized by XPC followed by the binding of the DNA helicase TFIIH that causes the unwinding of the DNA. As soon as the DNA is partially unwound, all pre-incision proteins can bind to, and dissociate from the repair intermediates in any order.

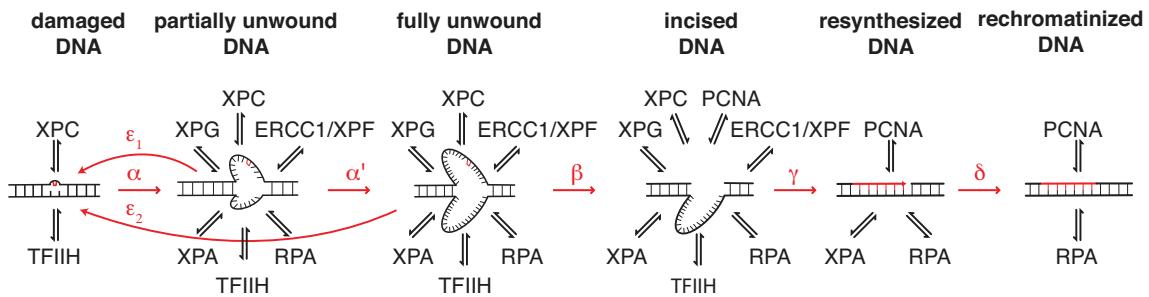


Figure 2.7: Kinetic model of NER The model distinguishes six DNA repair intermediates that are interconverted by enzymatic steps (red arrows: α , partial DNA unwinding; α' , full unwinding; β , dual incision; γ , resynthesis; δ , rechromatinization; ε_1 and ε_2 , reannealing of the unwound DNA). The indicated NER proteins can bind to the repair intermediates. The binding of TFIIH to the DNA lesion requires the prior binding of XPC and the binding of XPA and ERCC1/XPF is always cooperative.

The model scheme depicted in Figure 2.7 can be translated into a system of rate equations (Equation 2.2). The concentration of any possible assembly state, i.e., the different repair intermediate with possibly bound proteins is indicated by $y_{\pi}^R(t)$. The superscript index R refers to the repair intermediate that is defined by the modification of the DNA substrate (damaged DNA (I), partially unwound DNA (II), fully unwound DNA (III), incised DNA (IV), resynthesized DNA (V) and chromatinized DNA (VI), cf. Figure 2.7). Which proteins are bound to the given repair intermediate is denoted by the binary number π . The tuple π consists of 7 elements; one for each protein $p \in \{C, T, G, A, F, R, P\}$, where XPC (C), TFIIH (T), XPG (G), XPA (A), ERCC1/XPF (F), RPA (R) and PCNA (P). Each protein variable $\pi(p) \in \{1, 0\}$ indicates if the protein is bound ($\pi(p) = 1$) or not ($\pi(p) = 0$). The protein variable equals zero ($\pi(p) = 0$), if the protein cannot bind to the given repair intermediate (R).

Repair intermediate	Binding proteins	Catalyzed process Required proteins	Remarks	Ref.
(I) Damaged DNA with helical distortion	XPC,TFIIH (3 states)	Partial unwinding (reaction α) XPC and TFIIH	Initiation by binding of XPC and subsequent recruitment of TFIIH.	[23] [67] [89] [66] [83]
(II) Partially unwound DNA	XPC,TFIIH, XPG,XPA, ERCC1/XPF, RPA (48 states)	Full unwinding (reaction α') XPC, TFIIH, XPG, XPA, RPA	ERCC1/XPF only binds to repair complexes that contain XPA. If the DNA becomes devoid of any protein it will re-anneal to repair intermediate I (reaction ε_1).	[23] [83] [57]
(III) Fully unwound DNA	XPC,TFIIH, XPG, XPA, XPF-ERCC1, RPA (48 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_2). Dual incision requires the endonucleases XPG and ERCC1/XPF and is stimulated by TFIIH, XPA, RPA and possibly XPC.	[23] [59] [74] [87] [15]
(IV) Incised DNA	XPC,TFIIH, XPG, XPA, ERCC1/XPF, RPA, PCNA (96 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).	[23] [87]
(V) Resynthesized DNA	XPA, RPA, PCNA (9 states)	Rechromatinization (reaction δ) RPA and PCNA	Accumulation and FLIP data imply that XPA binds to repaired DNA while the pre-incision proteins do not (Figure 2.3 and 2.4).	[54] [73] [44]
(VI) Rechromatinized DNA	RPA, PCNA (4 states)		RPA and PCNA associate with rep. intermediate I, as levels of bound EGFP-PCNA and EGFP-RPA are high up to at least 4h after UV irradiation while other repair proteins are no longer bound.	[67] [44]

Table 2.2: Model assumptions

Since all proteins can bind in any order after the initial unwinding, each repair intermediate consists of 2^N possible states where N is the number of proteins that can bind to the given DNA substrate. However, there are two exceptions to that rule as already indicated in Table 2.2. Only XPC can recognize the DNA lesion and recruits TFIID upon binding, consequently the tuple with C=0 and T=1 is excluded in repair intermediate I (damaged DNA) and only $2^2 - 1 = 3$ states exist. The second exception is the cooperative binding of XPA and XPF/ERCC. During the whole NER process XPF/ERCC1 (F) can only bind if XPA (A) is already bound, thus any tuple with A=0 and F=1 must be excluded. This is consistent with data from intact cells [83]. In repair intermediate II (partially unwound DNA) and III (fully unwound DNA), 6 proteins can bind, which results in $2^6 - 2^4 = 48$ assembly states. In repair intermediate IV (incised DNA) the maximal number of 7 proteins can bind resulting in $2^7 - 2^5 = 96$ possible states. Repair intermediate V (resynthesized DNA) has $2^3 = 8$ states and repair intermediate VI (rechromatinized DNA) has only 4 states for the binding of PCNA and RPA.

The time development of these 207 states can be written in the form of the following differential equation. For each repair intermediate $R = \text{I}, \text{II}, \text{III}, \text{IV}, \text{V}$ and VI , one has to consider all allowed tuples π .

$$\frac{d}{dt} y_\pi^R = \sum_{p \text{ in } R} \eta \left((-1)^{\pi(p)} l_p^R y_\pi^R |_{\pi(p)=1} + (-1)^{1+\pi(p)} k_p^R C_p(t) y_\pi^R |_{\pi(p)=0} \right) + E(y_\pi^R) \quad (2.2)$$

The on-rate constant of protein p for a certain repair intermediate R is given by k_p^R and the corresponding off-rate constant by l_p^R . The time dependent concentration $C_p(t)$ will be given in equation (2.3).

The cooperativity factor η guarantees that no states appear that are excluded due to cooperative binding events.

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

Any possible enzymatic reaction is denoted as $E(y_\pi^R)$. If the state y_π^R has no in- or outgoing

enzymatic reaction: $E(y_\pi^R) = 0$. All enzymatic reactions $E(y_\pi^R)$ occurring in the model are given in the following.

For damaged DNA ($R = \text{I}$):

$$E(y_{00}^{\text{I}}) = \varepsilon_1 y_{000000}^{\text{II}} + \varepsilon_2 y_{000000}^{\text{III}} \quad \text{and} \quad E(y_{11}^{\text{I}}) = -\alpha y_{11}^{\text{I}}.$$

After the recruitment of TFIIH by XPC, TFIIH starts with its helicase activity (α) and the re-annealing reactions ($\varepsilon_{1,2}$) return the substrate to its original state of damaged DNA.

For partially unwound DNA ($R = \text{II}$):

$$\begin{aligned} E(y_{000000}^{\text{II}}) &= -\varepsilon_1 y_{000000}^{\text{II}}, & E(y_{110000}^{\text{II}}) &= \alpha y_{11}^{\text{I}}, \\ E(y_{111101}^{\text{II}}) &= -\alpha' y_{111101}^{\text{II}} \quad \text{and} \quad E(y_{111111}^{\text{II}}) &= -\alpha' y_{111111}^{\text{II}}. \end{aligned}$$

A two-step unwinding model requires XPC, RPA, XPA, XPG, but not ERCC1/XPF [23], thus we allow the reaction of the full unwinding (α') to be catalyzed by the complex XPC-TFIIH-XPG-XPA-RPA and this complex with ERCC1/XPF additionally bound. The re-annealing reaction (ε_1) will return the DNA substrate without any proteins bound to its original state.

For fully unwound DNA ($R = \text{III}$):

$$\begin{aligned} E(y_{000000}^{\text{III}}) &= -\varepsilon_2 y_{000000}^{\text{II}}, & E(y_{111101}^{\text{III}}) &= \alpha' y_{111101}^{\text{II}} \\ E(y_{011111}^{\text{III}}) &= -\beta y_{011111}^{\text{III}} \quad \text{and} \quad E(y_{111111}^{\text{III}}) &= -\beta y_{111111}^{\text{III}} + \alpha' y_{111111}^{\text{II}}. \end{aligned}$$

Dual incision of fully unwound DNA requires the endonucleases XPG and ERCC1/XPF, moreover, dual incision is stimulated by TFIIH, XPA, RPA and possibly XPC [23, 87]. Accordingly, we assume that two complexes are able to carry out incision in the model, these are TFIIH-RPA-XPG-XPA-ERCC1/XPF and this one with XPC additionally bound. The re-annealing reaction (ε_2) will return the DNA substrate to its original state, if it loses all stabilizing proteins.

For incised DNA ($R = \text{IV}$):

$$\begin{aligned} E(y_{111111}^{\text{IV}}) &= \beta y_{111111}^{\text{III}}, & E(y_{011111}^{\text{IV}}) &= \beta y_{011111}^{\text{III}} \quad \text{and} \\ E(y_{0001011}^{\text{IV}}) &= -\gamma y_{0001011}^{\text{IV}}. \end{aligned}$$

Repair synthesis (γ) is carried out by the complex XPA-RPA-PCNA assembled on DNA.

We assume that XPA is bound during repair synthesis, since the dissociation kinetics of XPA is severely affected by inhibition of repair synthesis, in contrast to the other pre-incision proteins that are not effected by this treatment (Figure 2.4E). This assumption is also supported by ChIP experiments with antibodies against the late NER factor XRCC1-Lig III, which pulled down RPA and XPA, but not XPC and TFIIH [54]. The requirement of RPA and PCNA to perform repair synthesis has been shown previously [73].

And finally for resynthesized ($R = V$) and rechromatinized DNA ($R = VI$):

$$E(y_{011}^V) = \gamma y_{0001011}^{IV} - \delta y_{011}^V \quad \text{and} \quad E(y_{011}^{VI}) = \delta y_{011}^V.$$

Only PCNA and RPA stay bound over an extended period of time.

The equation system given by Equation (2.2) is amended by 7 equations for the time development of the concentration $C_p(t)$ of the free proteins $p \in \{C, T, G, A, F, R, P\}$:

$$\frac{d}{dt} C_p = \sum_{R=I}^{VI} \sum_{\pi} \xi \left(\delta_{\pi(p)1} l_p^R y_{\pi}^R - \delta_{\pi(p)0} k_p^R C_p y_{\pi}^R \right) \quad (2.3)$$

The sum over π runs over any possible realization of the tuple π in the given repair intermediate R . The cooperativity factor ξ guarantees that only possible states appear.

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

The Kronecker delta

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

ensures that proteins only bind to complexes not containing the protein yet and only leave complexes containing them.

The accumulation curves of the core proteins XPC, TFIIH, XPG, XPA, ERCC1/XPF, RPA

and PCNA are the sum over the concentration of all states that contain the respective protein. The initial value is given by a vector that assumes an initial damage concentration of $y_{00}^I = 0.33 \mu M$ and all other states are zero. The time courses of the FLIP experiments are generated by summing over the same states, but the respective on-rate constants equal zero for the given protein, since we approximate that any protein that rebinds has lost its fluorescence before. The starting value is given by the corresponding value of the accumulation simulation at the time point, when the FLIP experiment was started, i.e. 600 s for XPC and ERCC1/XPF; 900 s for XPG and TFIIH; 2000 s for XPA and 7200 s for PCNA.

2.3 Fit of the model to the data

The mathematical model allows to derive all k_{on} and k_{off} values for the binding of the individual proteins to the different repair intermediates and k_{cat} values for the enzymatic reactions by fitting it to all available experimental data, i.e. the long-term accumulation and the FLIP kinetics of 7 NER proteins, as well as the FLIP kinetics of XPC, XPA and PCNA in perturbed NER processes (cf. section 2.1). To minimize the number of fitting parameters and to constrain the model further, we used the following experimental observations and simplifications:

- (1) Since RPA binds to long stretches of single-stranded DNA more strongly than to short stretches, we constrained the RPA affinity to fully unwound DNA (repair intermediate III) to be at least five times as large as to partially unwound DNA (repair intermediate II) [5] $\left(\frac{k_R^{III}}{k_R^{II}} > 5 \frac{k_R^{II}}{k_R^{III}} \right)$.
- (2) However, RPA should have the same affinity for fully unwound DNA (repair intermediate III) and incised DNA (repair intermediate IV), since the single-stranded binding partner stays the same.
- (3) All proteins except RPA (i.e., XPC, TFIIH, XPG, XPA and ERCC1/XPF) have the same affinity for partially unwound (repair intermediate II) and fully unwound DNA (repair intermediate III). For XPC this is implied by the FLIP data that show that the XPC dissociation rate changes only after formation of the full pre-incision complex (and likely after dual incision). For the other proteins we make the same assumption, because their binding does not appear to require large-scale DNA unwinding. These assumptions reduce the number of distinct binding and dissociation rate constants.

- (4) The binding of TFIIH is dependent on the binding of the damage recognition factor XPC to the DNA lesion (i.e. repair intermediate I), and the dimer XPF/ERCC1 can only bind if XPA is present, as indicated in the equations already [83].

2.3.1 Results of the fit

Our model fit reproduces faithfully all available experimental data simultaneously, including the net accumulation kinetics, the FLIP kinetics for normal NER and NER blocked at different stages (Figure 2.8). This result shows that our model can describe the comprehensive data set and thus reconciles the longterm accumulation and the fast exchange of components by employing a predominantly random and rapidly reversible assembly mechanism.

In order to determine a fit that faithfully reproduces the data and to judge the quality of the fitted parameters, we employed a Markov-Chain Monte Carlo (MCMC) method (cf. section 1.2.1 and [65]) for minimizing the sum of the least squares between the experimental data and the model simulations. This procedure yields the best-fit values and a distribution for each parameter that allows to deduce confidence intervals. Table 2.3 and 2.4 show the best-fit values and 90% confidence intervals for the 47 parameters (20 pairs of k_{on} and k_{off} values, five catalytic rate constants, k_{cat} , and two re-annealing rate constants). In addition, the dissociation constants $K_D = k_{\text{off}}/k_{\text{on}}$ were calculated and the characteristic times for the enzymatic reactions are given by $1/k_{\text{cat}}$.

Enzymatic rate	k_{cat} [s^{-1}]	confidence interval [s^{-1}]
Partial unwinding α	0.08	[0.06; 0.11]
Full unwinding α'	0.74	[0.59; 0.74]
Incision β	4.1	[3.8; 6.0]
Resynthesis γ	0.05	[0.04; 0.06]
Rechromatinization δ	0.012	[0.012; 0.013]
Reannealing ε_1	3.1	[2.5; 24.1]
Reannealing ε_2	11.0	[4.9; 11.1]

Table 2.3: Enzymatic rate constants

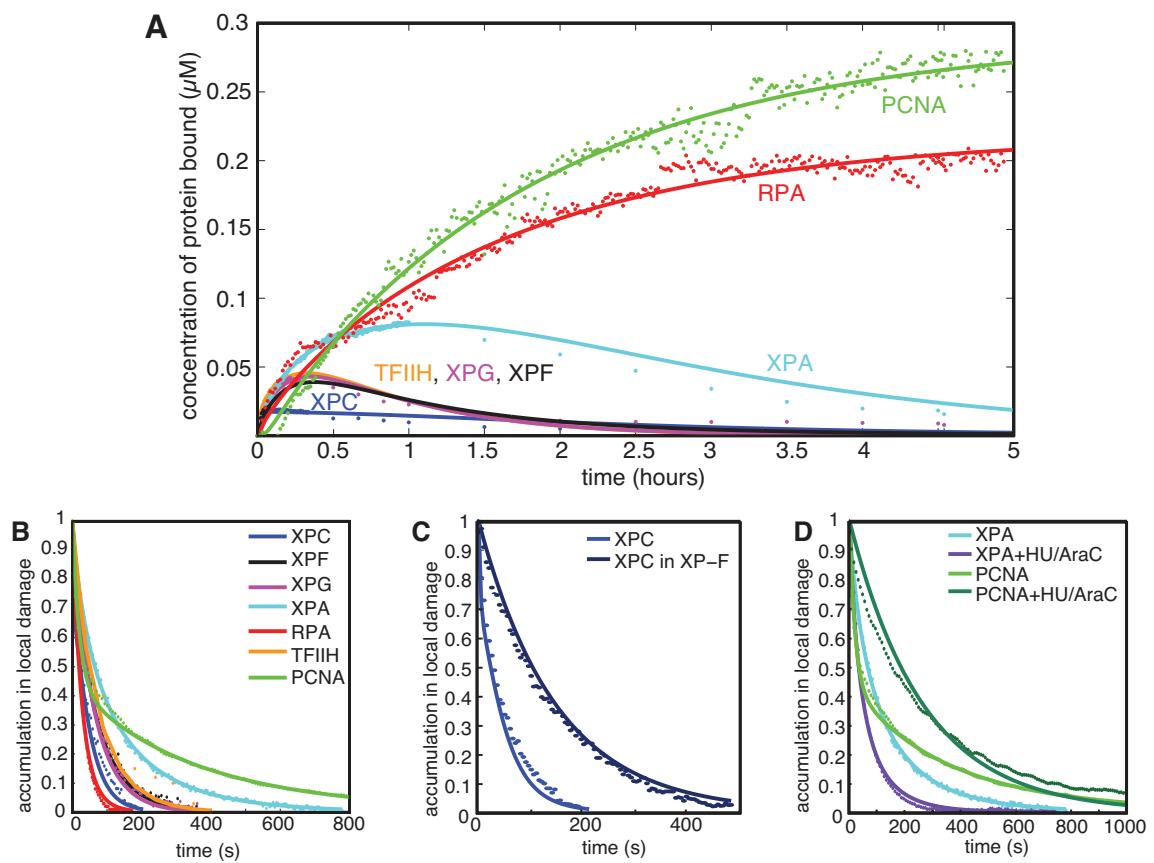


Figure 2.8: Model fit. Comparison of model simulations (lines) and experimental data (dots) showing: (A) net accumulation kinetics and (B) dissociation kinetics of core NER proteins; (C) dissociation kinetics of XPC in wild-type and XPF-deficient cells unable to perform damage excision; (D) dissociation kinetics of XPA and PCNA in the absence or presence of DNA synthesis/ligation inhibitors HU and AraC.

		XPC	TFIIE	XPG	XPA	XPF/ ERCC1	RPA	PCNA
Concentration	$[\mu M]$	0.140	0.360	0.440	1.110	0.170	1.110	1.110
damaged DNA	$k_{on} [\mu M^{-1} s^{-1}]$	0.008 [0.007;0.011]	1.6 [0.8;4.5]					
	$k_{off} [s^{-1}]$	0.061 [0.007;0.462]	0.05 [0.004;0.195]					
	$K_d [\mu M]$	7.8	0.033					
partially unwound DNA	$k_{on} [\mu M^{-1} s^{-1}]$	0.002 [0.001;0.003]	0.26 [0.11;0.27]	0.28 [0.19;0.31]	0.13 [0.12;0.16]	1.2 [1.1;1.6]	0.153 [0.107;0.221]	
$\tau = 35 \pm 41$ min	$k_{off} [s^{-1}]$	0.007 [0.006;0.008]	0.012 [0.009;0.016]	0.015 [0.012;0.015]	1.04 [0.75;1.30]	0.012 [0.011;0.014]	2.584 [1.736;3.577]	
fully unwound DNA	$K_d [\mu M]$	3.1	0.046	0.05	7.8	0.010	17	
$\tau = 41 \pm 36$ min	$k_{on} [\mu M^{-1} s^{-1}]$	0.002 [0.001;0.003]	0.26 [0.11;0.27]	0.28 [0.19;0.31]	0.13 [0.12;0.16]	1.2 [1.1;1.6]	0.006 [0.006;0.007]	
	$k_{off} [s^{-1}]$	0.007 [0.006;0.008]	0.012 [0.009;0.016]	0.015 [0.012;0.015]	1.04 [0.75;1.30]	0.012 [0.011;0.014]	0.021 [0.020;0.022]	
incised DNA	$K_d [\mu M]$	3.1	0.046	0.05	7.8	0.01	3.3	
$\tau = 41 \pm 36$ min	$k_{on} [\mu M^{-1} s^{-1}]$	0.220 [0.128;0.262]	0.0004 [0.0003;0.010]	0.001 [0.0004;0.007]	0.004 [0.004;0.005]	0.094 [0.066;0.111]	0.006 [0.006;0.007]	0.001 [0.001;0.002]
	$k_{off} [s^{-1}]$	0.396 [0.212;0.476]	0.0496 [0.039;0.065]	0.099 [0.043;0.114]	0.058 [0.046;0.067]	0.050 [0.040;0.101]	0.021 [0.020;0.022]	0.004 [0.004;0.004]
resynthesized DNA	$K_d [\mu M]$	1.8	137	89	14	0.5	3.3	2.8
$\tau = 2.0 \pm 0.7$ h	$k_{on} [\mu M^{-1} s^{-1}]$				0.054 [0.054;0.058]		0.080 [0.054;0.103]	0.010 [0.007;0.010]
	$k_{off} [s^{-1}]$				0.004 [0.004;0.005]		0.04 [0.03;0.05]	0.002 [0.002;0.002]
rechromatinized DNA	$K_d [\mu M]$				0.08 [0.061;0.074]		0.317 [0.248;0.336]	
$\tau = 2.2 \pm 0.7$ h	$k_{on} [\mu M^{-1} s^{-1}]$				0.041 [0.037;0.045]		0.052 [0.042;0.054]	
	$K_d [\mu M]$				0.61 [0.61		0.16	

Table 2.4: Binding and dissociation rate constants In addition, we indicated the characteristic time τ for each repair intermediate (cf. section 2.4.1 on repair kinetics).

2.3.2 Quality of the fit

However, we have to assess the quality of the 47 fitted parameters. The MCMC method allowed us, not only to determine a best fit, but also to derive confidence intervals and to look at correlations between the different parameters. These informations allow us to rate the quality of the parameter values. The confidence intervals give an estimate on how well the parameters are determined by the fit. The confidence intervals are considerably small for most of the parameters, with some exceptions, like the re-annealing rate constants ε_1 and ε_2 . However, the re-annealing of a ~ 30 nucleotide stretch of DNA is very fast, and we found that the precise values do not matter as long as the characteristic times for the re-annealing are in the subsecond range (as given in Table 2.3). In this case, the re-annealing of the DNA is limited by the dissociation rates of the proteins stabilizing the unwound state, as one can reasonably expect.

To analyze possible correlations between the parameters we looked at the correlation coefficient matrix. The correlation coefficient

$$\text{Cor}(X, Y) := \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)} \sqrt{\text{Var}(Y)}} = \frac{E[(X - E(X))(Y - E(Y))]}{\sqrt{\text{Var}(X)} \sqrt{\text{Var}(Y)}}, \quad (2.4)$$

where E is the expected value and Var its variance, is a dimensionless measure for the degree of the linear relationship between two parameters X and Y . The correlation coefficient lies always between -1 and +1. It is one for a complete linear relationship and zero, if there is no linear dependence. However, two parameters can still be correlated in a non-linear way, when the correlation coefficient is zero. Nevertheless, we assume that the linear dependence shows the leading order of possible correlations. The confidence intervals will be too big, if there are correlations between parameters. If two parameters have a strong linear correlation, it is only possible to determine the quotient of them accurately. In Figure 2.9 we show a graphical representation of the correlation matrix. The diagonal of the matrix is one by definition. The closer the value is to one, the stronger is the correlation between the parameters. For several parameters, like the on- and off-rate constant of XPC (k_C^{IV} , l_C^{IV}), XPG (k_G^{IV} , l_G^{IV}) and XPA (k_A^{IV} , l_A^{IV}) for incised DNA and the on- and off-rate constant of RPA (k_R^{VI} , l_R^{VI}) for rechromatinized DNA, we find high values on the secondary diagonal, thus we can only determine the dissociation constant $K_D = k_{\text{off}}/k_{\text{on}}$ accurately, but not the individual rates. For most of the correlations it is easy to see the underlying reason. The strong correlation of the off-rate constant of RPA for resynthesized and rechromatinized DNA are probably taking account for the requirement that many RPA proteins should be bound in the late phase of the NER process, however, the given data does not allow to differentiate if the RPA molecules are bound to the resyn-

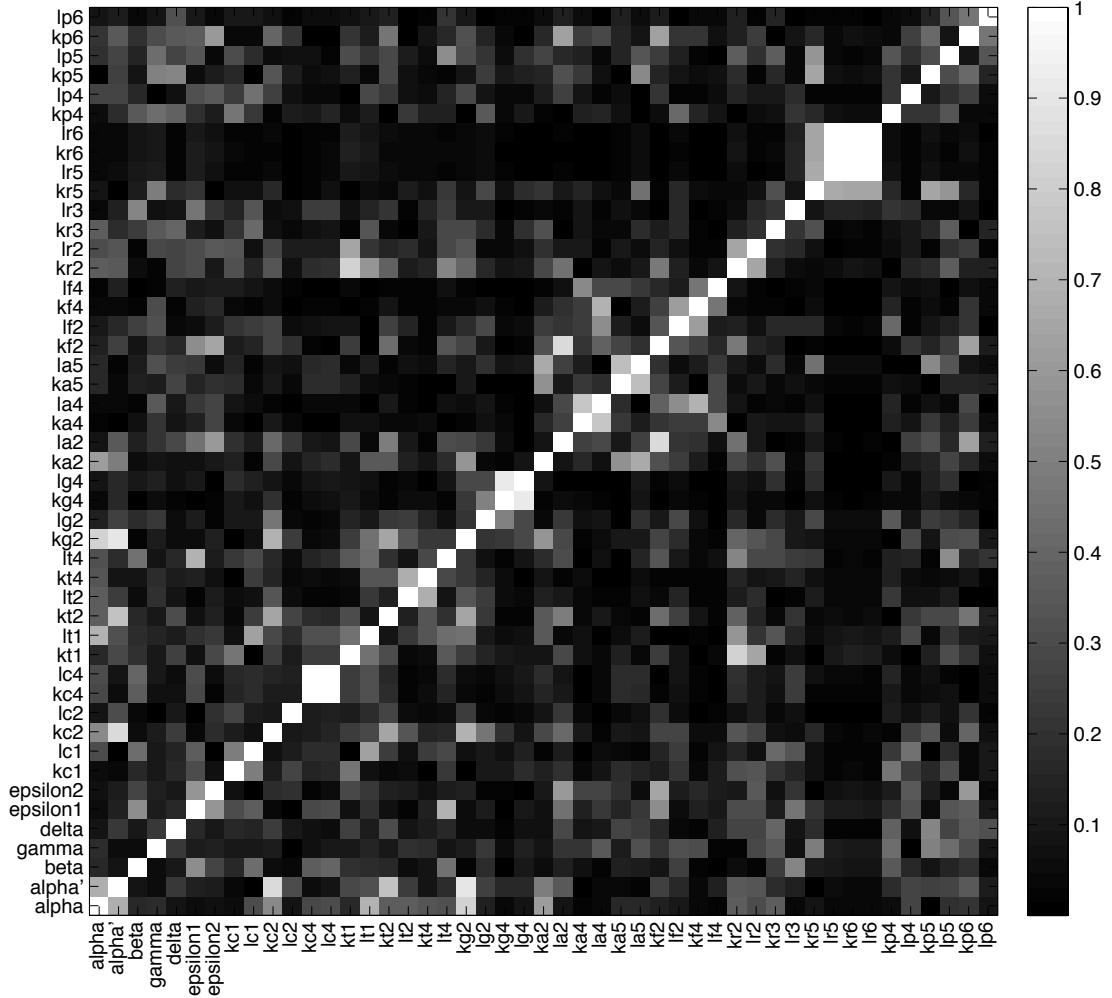


Figure 2.9: Correlation matrix. Graphical representation of the absolute values of the correlation matrix. The correlation of the diagonal is one by definition.

thesized or to the rechromatinized DNA. The correlation of the off-rate constant of XPA ($k_A^{II,III}$) and the on-rate constant of XPF ($k_F^{II,III}$) for unwound DNA is due to the strong cooperative binding of the two proteins. To summarize, our analysis shows that the parameters are not in all cases strictly determined, but they cannot be restricted significantly stronger, without further experimental data. However, it is important to note, that the fit shows that our model can explain the given experimental data well, without making any artificial constraints.

Moreover, all kinetic parameters extracted from the model fall in biochemically realistic ranges. The *in-vivo* affinities of the NER proteins (k_{on}/k_{off}) for the repair intermediates span a considerable range, from micromolar to nanomolar values for the dissociation equi-

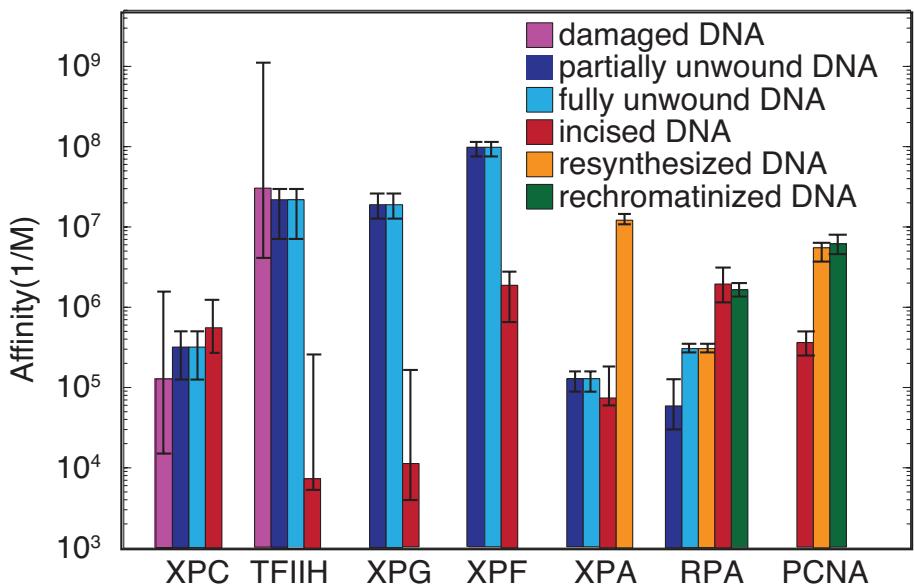


Figure 2.10: In-vivo affinities. Affinity of NER proteins for the individual repair intermediates ($K_a = \frac{k_{on}}{k_{off}}$).

librium constants (K_D) (Figure 2.10). The only exception is the rather low affinity of the damage-recognition factor XPC for the damaged but not yet unwound DNA ($K_D \sim 8\mu M$), whereas the XPC concentration is about $0.14 \mu M$. The binding of XPC is then stabilized by the recruitment of TFIID. The pre-incision factors XPG, TFIID and ERCC1/XPF bind with high affinity to unwound DNA, but have at least one order of magnitude lower affinity for incised DNA. In contrast, the affinity of XPA, PCNA and RPA increases upon repair synthesis, which explains why the bound levels of these proteins remain high for extended periods of time (Figure 2.8A). In general, the proteins have a high affinity while they perform their function and loose their affinity afterwards. Thus, the fitted kinetic parameters reflect our expectations from the experiments without implying any constraints on the parameters during the fit.

To draw the conclusion the mathematical modeling and the resulting analysis of the complete experimental data set of the kinetics of the core NER factors implies a predominantly random and rapidly reversible assembly mechanism of the repair complexes.

2.4 Emerging properties of NER

2.4.1 Repair kinetics

The model does not only offer the possibility to determine the kinetic parameters of NER, but allows to analyze the repair process in greater detail and to predict the repair kinetics. The model predicts an exponential degradation of DNA lesions (blue curve in Figure 2.11A), which is equivalent to the time development of the concentrations of all states in the repair intermediates preceding the excision of the lesion, that is damaged, partially and fully unwound DNA. To measure the degradation of DNA lesions experimentally, the cells are irradiated and fixed at various points of time. Afterwards they are stained for 6-4PPs and CPDs with antibodies. For each time point the value is averaged over approximately 50 cells, to minimize the error caused by statistical variations between the different cells. There is very little repair of CPDs during the first 8 hours after local UV irradiation (Figure 2.2) and most of the 6-4PPs are already removed within 1 hour (red curve in Figure 2.11A). The measured curve for the 6-4PP degradation shows a very similar behavior as the model prediction considering the large errors that are unavoidable for the staining with antibodies. The model predicts the repair kinetics even in more detail. The simulated time evolution of the individual DNA repair intermediates allows to determine the characteristic time and standard deviation of the different stages of the repair process (Figure 2.11B).

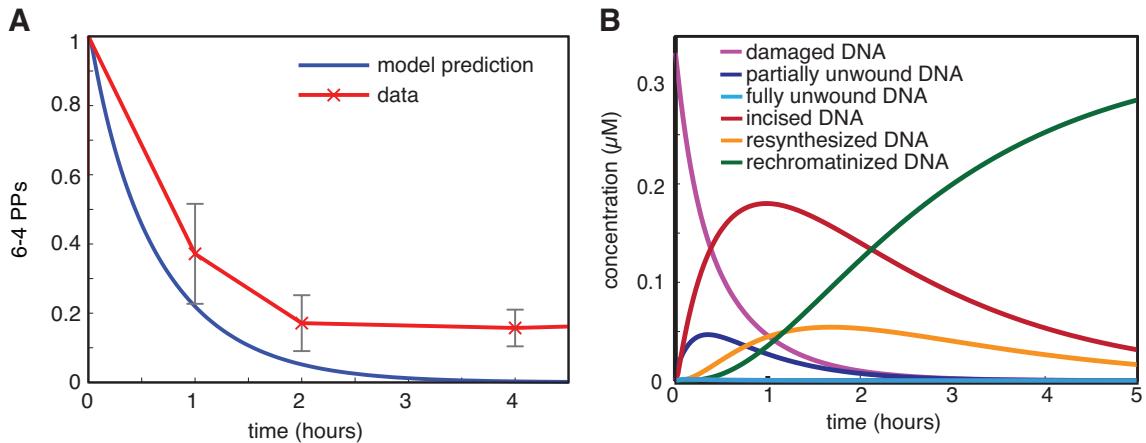


Figure 2.11: Repair kinetics. (A) Comparison of the predicted degradation of DNA lesions (blue) and the measured degradation of 6-4 PPs. (B) Simulated kinetics of all repair intermediates.

In the following we will calculate the characteristic time for the different repair intermediates and its statistical variation. The characteristic time $\tau(y_\pi^R)$ of the state $y_\pi^R(t)$ is defined

as the first moment of $y_\pi^R(t)$ divided by its zeroth moment.

With the k^{th} moment

$$\mu^{(k)} = \int_0^\infty t^k y_\pi^R(t) dt, \quad (2.5)$$

the characteristic time is

$$\tau = \frac{\mu^{(1)}}{\mu^{(0)}} \quad (2.6)$$

and the standard deviation is

$$\sigma = \sqrt{\frac{\mu^{(2)}}{\mu^{(0)}} - \tau^2}. \quad (2.7)$$

The characteristic time to partial unwinding ($R = I$), time to full unwinding ($R = II$) and time to incision ($R = III$) is given by

$$\tau_R = \frac{\int_0^\infty t \sum_{x=I}^R \sum_\pi y_\pi^x(t) dt}{\int_0^\infty \sum_{x=I}^R \sum_\pi y_\pi^x(t) dt}. \quad (2.8)$$

The first sum runs over all preceding repair intermediates and the repair intermediate itself. The second sum runs over all tuple π , i.e. the concentrations of all states within a given repair intermediate.

In order to guarantee the convergence of the integral, the time to re-synthesis has to be calculated by subtracting the sum of the concentration of all states in the subsequent repair intermediates (V+VI) from the initial concentration of damage at time point zero ($t = 0$) given by $y_{00}^I(0)$.

$$\begin{aligned} \tau_{\text{syn}} &= \frac{\int(y_{00}^I(0) - (\sum_\pi y_\pi^{IV}(t) + \sum_\pi y_\pi^V(t)) dt) \cdot t dt}{\int y_{00}^I(0) - (\sum_\pi y_\pi^{IV}(t) + \sum_\pi y_\pi^V(t)) dt} \\ \sigma_{\text{syn}} &= \sqrt{\frac{\int(y_{00}^I(0) - (\sum_\pi y_\pi^{IV}(t) + \sum_\pi y_\pi^V(t))) \cdot t^2 dt}{\int(y_{00}^I(0) - (\sum_\pi y_\pi^{IV}(t) + \sum_\pi y_\pi^V(t))) dt} - \tau_{\text{syn}}^2} \end{aligned}$$

In our model the initial value is given by $y_{00}^I(0) = 0.33 \mu M$. The time to chromatinize is calculated analogously by subtracting all states in repair intermediate ($R = VI$).

$$\tau_{\text{chrom}} = \frac{\int(y_{00}^I(0) - \sum_\pi y_\pi^V(t)) \cdot t dt}{\int y_{00}^I(0) - \sum_\pi y_\pi^V(t) dt}$$

$$\sigma_{\text{chrom}} = \sqrt{\frac{\int(y'_{00}(0) - \sum_\pi y_\pi^V(t)) \cdot t^2 dt}{\int(y'_{00}(0) - \sum_\pi y_\pi^V(t)) dt} - \tau_{\text{chrom}}^2}$$

repair intermediate	characteristic time τ_R
partially unwound DNA (II)	$\tau_{\text{part}} = 35 \pm 30 \text{ min}$
fully unwound DNA (III)	$\tau_{\text{full}} = 41 \pm 36 \text{ min}$
incised DNA (IV)	$\tau_{\text{inc}} = 41 \pm 36 \text{ min}$
resynthesized DNA (V)	$\tau_{\text{syn}} = 2.0 \pm 0.7 \text{ h}$
rechromatinized DNA (VI)	$\tau_{\text{chrom}} = 2.2 \pm 0.7 \text{ h}$

Table 2.5: Characteristic times for our reference parameter set.

The full list of characteristic times and standard deviations for our reference parameter set is shown in Table 2.5. The incision of the DNA is the first irreversible step in our modeling framework and is therefore a measure for the repair process as a whole. DNA lesions are excised on average 41 ± 36 min after UV irradiation, with large stochastic variation between the lesions. Damage recognition by XPC and subsequent partial unwinding of the DNA by TFIIH takes already ~ 35 min on average. In contrast to that ~ 6 minutes are on average enough to fully unwind the DNA and assemble the pre-incision complex containing XPA, XPG, ERCC1/XPF, RPA and TFIIH. Thus, the time to incision is mainly determined by slow damage recognition, while the functional pre-incision complex is rapidly formed by a random assembly mechanism. We conclude that the large number of different assembly pathways (cf. Figure 2.6) outweighs the large amount of partially assembled complexes and allows rapid complex assembly. To summarize, recognition of DNA lesions is time-consuming, while subsequent pre-incision complexes are formed rapidly by reversible and random binding of the individual components.

2.4.2 Control coefficients

To determine the control of each NER protein on the rates of incision and repair synthesis, we calculated the control coefficients

$$C_i^\tau = \frac{X_i}{\tau^{-1}} \frac{\partial \tau^{-1}}{\partial X_i}, \quad (2.9)$$

which yield the relative change in the repair rate, as measured by the inverse of the characteristic time τ^{-1} , as a result of a fractional change in the concentration of repair factor X_i . The time to incision (τ_{inc}) and to DNA re-synthesis (τ_{syn}) are controlled differently by the NER factors. Generally, various factors contribute to a different degree and there

is no single limiting factor for DNA repair, showing that the rate of NER is a systems property rather than being determined by a single protein (Figure 2.12). However, XPC has the dominant control on the rate of incision and the rate of repair synthesis is mainly controlled by RPA, XPA and PCNA.

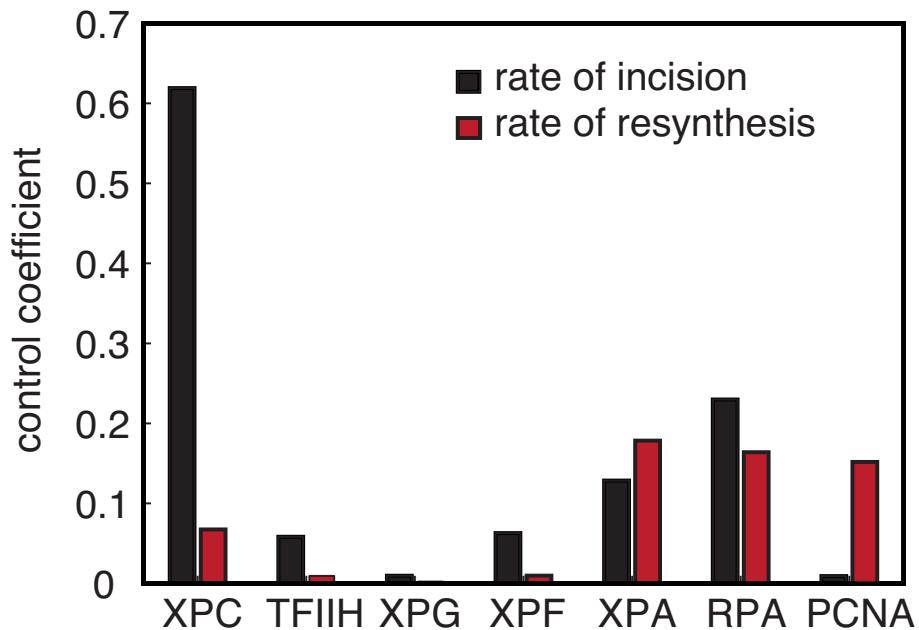


Figure 2.12: Control coefficients. Control of NER proteins on the rate of incision (black bars) and rate of DNA resynthesis (grey bars).

2.4.3 High capacity of NER

To quantify the rate and capacity of NER, we approximated the incision rate (ν) as a function of the amount of DNA lesions (D) by the Michaelis-Menten equation:

$$\nu(t) = \nu_{\max} \frac{D(t)}{K_M + D(t)}. \quad (2.10)$$

We approximate the time dependent change of the total damage by

$$\frac{dD}{dt} = -\nu(t)$$

and separation of variables

$$D(t)dt = -\frac{K_m + D(t)}{\nu_{\max}} dD$$

and subsequent integration yields a characteristic time for repair:

$$\tau = \frac{1}{D(0)} \int_0^{\infty} D(t)dt = \frac{1}{2} \frac{D(0)}{\nu_{\max}} + \frac{K_M}{\nu_{\max}} \quad (2.11)$$

According to this rate equation, the time to incision would rise as a linear function of initial damage $D(0)$.

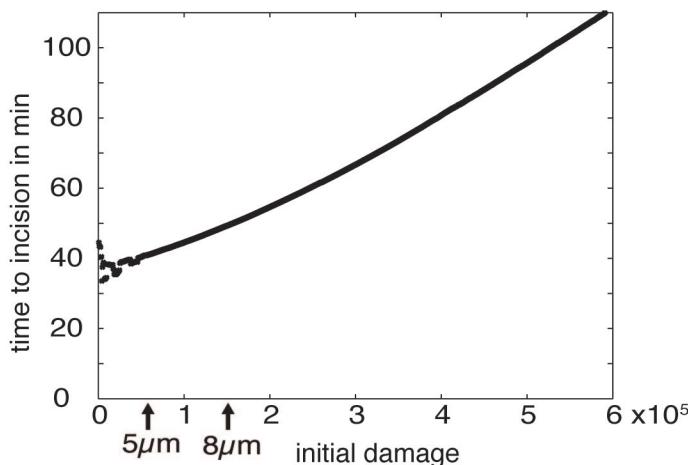


Figure 2.13: Capacity of NER. Computation of the average time to incision in the model versus the initial amount of DNA damage (6-4 PPs). The arrows indicate the amounts of damage estimated for the experimental setup with UV-permeable pores of $5 \mu\text{m}$ and $8 \mu\text{m}$ diameter.

Approximating the initial part of the τ_{inc} curve in Figure 2.13 for the characteristic time to incision vs. the initial amount of damage as predicted by the model by a straight line, yields this linear function.

The slope of the line is approximately:

$$\frac{1}{2} \frac{1}{\nu_{\max}} = \frac{5 \text{ min}}{60000 \text{ lesions}}$$

and the maximal rate of repair per cell nucleus is consequently:

$$\nu_{\max} = 6000 \frac{\text{lesions}}{\text{min}}$$

(where lesions refer to the rapidly repaired 6-4 PPs), which agrees with previous measurements [39, 88].

The interpolated intersection with the ordinate is approximately 36 min:

$$\frac{K_M}{\nu_{\max}} = 36 \text{ min}$$

and the estimate for the half-saturation constant is thus:

$$K_m = 36 \text{ min} \cdot 6000 \frac{\text{lesions}}{\text{min}} = 216000 \text{ lesions},$$

which is reached only upon intense UV irradiation of the nucleus [61]. The estimated half-saturation at $K_m = 216000$ lesions indicates that NER is not saturated under our experimental conditions ($\sim 60,000$ DNA lesions at $t = 0$).

To verify if NER is indeed unsaturated, we investigated how the NER system reacts on different amounts of DNA lesions. The corresponding experiment was done with a larger pore size of $8\mu\text{m}$ diameter for UV irradiation, as compared to the $5\mu\text{m}$ pores under standard conditions, and thus had a higher amount of DNA damage per nucleus. We tested the model by simulating the corresponding ~ 2.6 -fold increase in the initial amount of DNA lesions. The experimentally measured accumulation kinetics of XPG-EGFP for UV-C irradiation with 100 J m^{-2} and $8\mu\text{m}$ pore diameter (Figure 2.14A, red plus signs) matched the predicted curves generated by the model (Figure 2.14A, red line), thus confirming the validity of the model. In particular, the absolute magnitude of accumulation was correctly predicted. Moreover, we observed a similarly rapid decline of the XPG accumulation as for $5\mu\text{m}$ pore diameter (shown as a reference in Figure 2.14A). Since XPG preferentially dissociates after dual incision, this decline is indicative for the excision of the DNA lesions, implying that NER proceeds at unimpaired rate when the amount of initial damage is increased. In agreement with NER being far from saturation, we observed an essentially linear relationship between the maximal bound fraction of XPG and the number of DNA lesions, using $3, 5$ and $8\mu\text{m}$ pore diameters at 100 J m^{-2} UV-C irradiation (Figure 2.14B).

Thus, NER has a high capacity for processing DNA lesions in parallel.

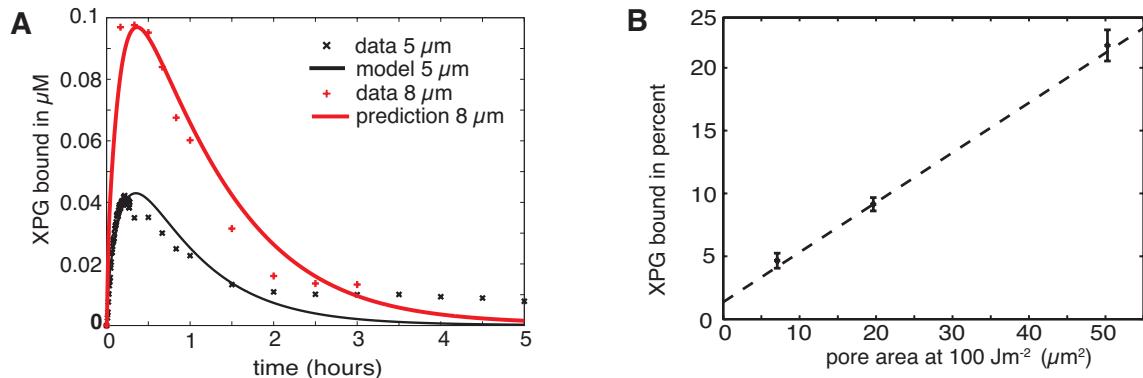


Figure 2.14: Capacity of NER. (A) The model correctly predicts the kinetics of XPG accumulation when the amount of initial DNA damage is increased ~ 2.6 fold (red +, experimental data for irradiation through $8 \mu m$ pores; red line, model simulation) as compared to reference conditions (x, experimental data; black line, model). (B) Maximally bound XPG-EGFP following local UV-C irradiation of differently-sized areas ($100 J m^{-2}$ through $3, 5$, and $8 \mu m$ pores).

2.4.4 XPC kinetics in repair-deficient cells

Since XPC has the highest control and global-genome NER is strictly dependent on damage recognition by XPC, we investigate if all available XPC proteins ($\sim 0.14 \mu M$) are bound to DNA damage under the standard experimental conditions of local damage experiments ($[6 - 4PP] \sim 0.33 \mu M$) when repair of the lesions is inhibited. In the corresponding experiment XPC-EGFP is expressed in repair-deficient XP-A cells and the binding kinetics of XPC is measured after localized UV irradiation (Figure 2.15A). The net accumulation of XPC-EGFP in XPA-deficient cells was slightly increased compared to its accumulation in repair-proficient cells and closely matched the amplitude predicted by the model (Figure 2.15B, red crosses and red line). The prediction and data for the slowly decreasing XPC kinetics in repair-proficient cells (Figure 2.15B, blue crosses and line) corresponds to the experiments under standard conditions. In agreement with our model prediction, XPC stays bound for a prolonged period of time in the repair-deficient cells. Remarkably, this plateau is at about 10% of the estimated amount of DNA damages (6-4PPs). This finding supports our prediction of a low XPC affinity and indicates that the unsaturated nature is, at least in part, due to the comparatively weak XPC binding.

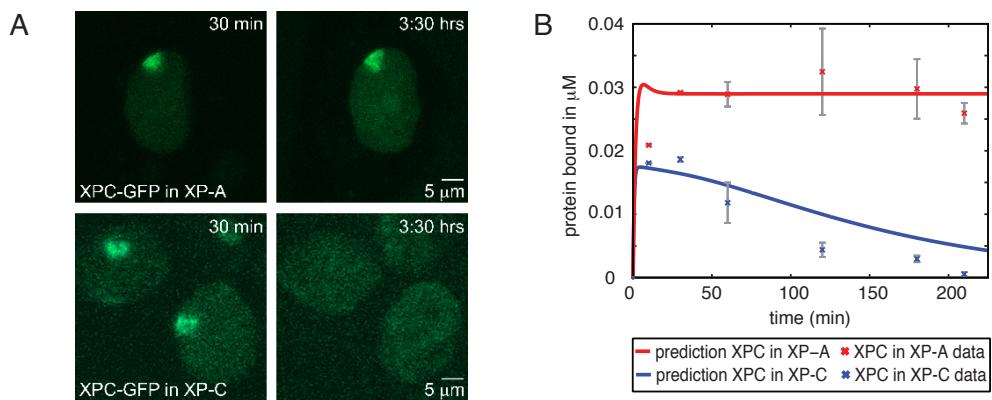


Figure 2.15: XPC kinetics in repair proficient and deficient cells. (A) XPC-GFP in XP-A and XP-C cells. (B) The blue curve predicts the kinetics of bound XPC-EGFP in XPC-deficient cells and the red curve predicts the prolonged accumulation of XPC-GFP in XPA-deficient cells. The measured data points are shown in red and blue respectively.

Chapter 3

Theoretical analysis of multiprotein complex assembly

To get a better understanding of the underlying concepts of the multi-protein complex assembly, we went back to a simplified model, that assumes the independent and reversible binding of N proteins and a subsequent enzymatic modification of the chromatin substrate that is catalyze by the full complex (Figure 3.1). The assembly scheme can be depicted as an N-dimensional hypercube, with 2^N different protein complexes forming the vertices including the empty DNA substrate and $2^{N-1}N$ protein binding reactions forming the edges.

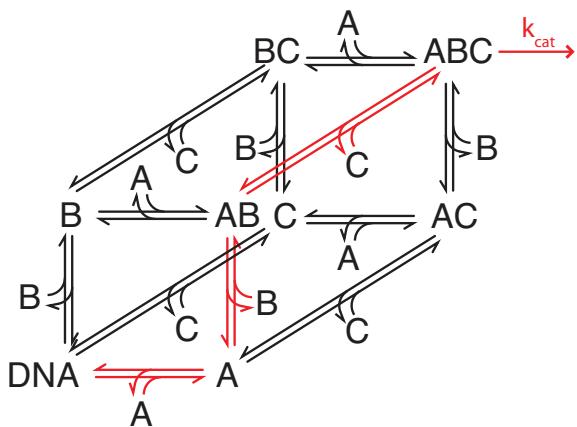


Figure 3.1: Complex assembly. The proteins A, B and C bind in a random order and the full complex ABC catalyzes the modification of the substrate (k_{cat}). Random assembly can utilize all pathways shown, while sequential assembly will follow a unique pathway for example the one indicated in red.

The concentration of each possible state is indicated by y_π . Which proteins are bound to the chromatin substrate is recorded in the binary The number π , that consists of p elements, where p is number of proteins. Each protein variable $\pi(p) \in \{1, 0\}$ reveals if the

protein is bound ($\pi(p) = 1$) or not ($\pi(p) = 0$). The time development is given by

$$\frac{d}{dt} y_\pi = \sum_p (-1)^{\pi(p)} l_p y_\pi |_{\pi(p)=1} + (-1)^{1+\pi(p)} k_p C_p(t) y_\pi |_{\pi(p)=0} + E(y_\pi). \quad (3.1)$$

The on-rate constant of protein p is given by k_p and the corresponding off-rate constant by l_p . The time dependent concentration C_p is determined by Equation (3.2) and the enzymatic rate is given by

$$E(y_\pi) = \begin{cases} -k_{\text{cat}} \hat{y} & \text{if } y_\pi = \hat{y} \\ 0 & \text{else} \end{cases},$$

where the concentration of the fully assembled complex is denoted by \hat{y} .

For each free protein the equation system is amended by an equation for the time development of the concentration C_p .

$$\frac{d}{dt} C_p = \sum_\pi (\delta_{p1} l_p y_\pi - \delta_{p0} k_p C_p y_\pi) \quad (3.2)$$

The sum runs over any possible realization of the tuple π , guaranteed by the Kronecker delta

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

3.1 Assembly and reaction times for multi-protein complexes

The rate of the reaction catalysed by the multiprotein complex is given by $v(t) = \alpha \hat{y}(t)$ where $\hat{y}(t)$ is the fraction of fully assembled complexes. The mean time at which the reaction occurs is

$$\tau = \frac{\mu_1}{\mu_0} \quad (3.3)$$

and the standard deviation

$$\sigma = \sqrt{\frac{\mu_2}{\mu_0} - \tau^2} \quad (3.4)$$

measures the fluctuations around the mean (Figure 3.2, cf. section 2.4.1).

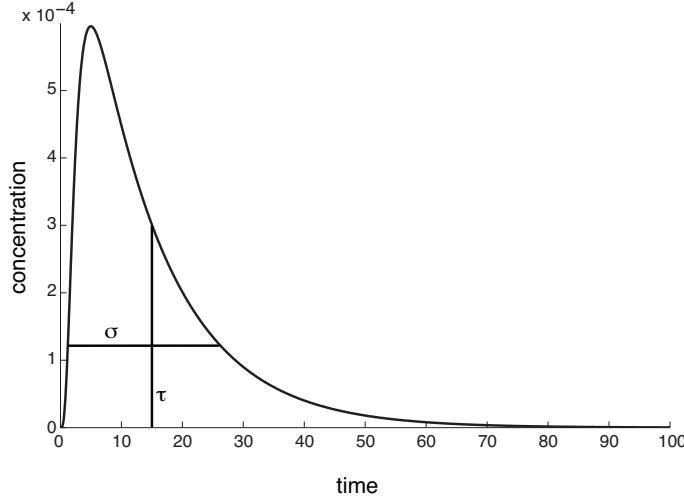


Figure 3.2: Characteristic time. Position of the characteristic time τ and the standard deviation σ .

Assuming that the free concentration of all proteins is constant, i.e., $\dot{C}_p(t) = 0$ in equation (3.2), which corresponds to a situation where only a small fraction of the proteins is bound to the chromatin at any time, and all binding and dissociation rate constants are equal ($k_i = k$ and $l_i = l$), the mean reaction time can be calculated explicitly. The time development of all possible assembly states (y_i) is linear under these conditions and can thus be written as matrix equation

$$\dot{\vec{y}} = S \vec{y}. \quad (3.5)$$

For a sequential assembly the tridiagonal matrix is given by

$$S_{i,j} = \begin{cases} -k & \text{for } j = i = 0 \\ -k - l & \text{for } j = i \text{ and } i \in \{1, \dots, N-1\} \\ -l - \alpha & \text{for } j = i = N \\ k & \text{for } j = i - 1 \text{ and } i \in \{1, \dots, N\} \\ l & \text{for } j = i + 1 \text{ and } i \in \{0, \dots, N-1\} \\ 0 & \text{else.} \end{cases} \quad (3.6)$$

The zeroth moment $\vec{\mu}_0$ for all components is given by

$$\begin{aligned}\vec{\mu}_0 &= \int \vec{y} dt \\ &= -S^{-1} \vec{y}(0),\end{aligned}\tag{3.7}$$

and the first moment by

$$\begin{aligned}\vec{\mu}_1 &= \int \vec{y} t dt \\ &= -S^{-1} \vec{\mu}_0\end{aligned}\tag{3.8}$$

since

$$\vec{\mu}_0 = \underbrace{\int \vec{y} dt}_{\text{Integration by parts}} = - \int \dot{\vec{y}} t dt = -S \int \vec{y} t dt$$

and the second moment analogously by

$$\begin{aligned}\vec{\mu}_2 &= \int \vec{y} t^2 dt \\ &= -2 S^{-1} \vec{\mu}_1\end{aligned}\tag{3.9}$$

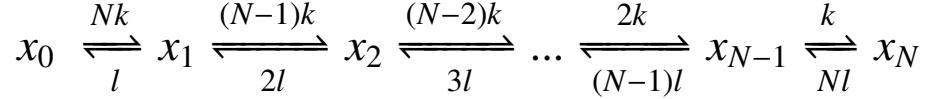
The characteristic time of assembly for the full complex $\hat{y} = y_N$ is thus

$$\tau_N = \frac{\int \hat{y} t dt}{\int \hat{y} dt} = \frac{(S^{-1} S^{-1} \vec{y}(0))_N}{(S^{-1} \vec{y}(0))_N}\tag{3.10}$$

and the standard deviation is

$$\sigma_N = \sqrt{\frac{\int \hat{y} t^2 dt}{\int \hat{y} dt} - \tau_N^2} = \sqrt{\frac{(-2S^{-1} S^{-1} S^{-1} \vec{y}(0))_N}{(S^{-1} \vec{y}(0))_N} - \tau_N^2}.\tag{3.11}$$

For a completely random binding scheme there are 2^N possible assembly states for a complex of N proteins. Since all proteins have the same binding rate constants, assembly states with the same number of proteins can be summarized in a single state x_i with $i = 0, \dots, N$ for zero proteins bound to the chromatin up to N proteins bound to it. This pseudo sequential assembly has the following reaction scheme:



The on- and off-rate constants are modified by weighting factors that account for the possible ways to bind a further protein or to release one. Accordingly, the random assembly can be mapped on a sequential assembly that is described by

$$\dot{\vec{x}} = R \vec{x}. \quad (3.12)$$

The tridiagonal matrix R for this pseudo sequential assembly is given by

$$R_{i,j} = \begin{cases} -(N-i)k - il & \text{for } j = i \text{ and } i \in \{0, \dots, N-1\} \\ -Nl - \alpha & \text{for } j = i = N \\ (N-i)k & \text{for } j = i-1 \text{ and } i \in \{1, \dots, N\} \\ il & \text{for } j = i+1 \text{ and } i \in \{0, \dots, N-1\} \\ 0 & \text{else} \end{cases} \quad (3.13)$$

The characteristic time of the random assembly and its standard deviation can thus be calculated with the formula for the sequential assembly (Equation 3.10) by interchanging the matrices S and R .

The characteristic time of a assembly for a complex of N proteins has the general form

$$\tau = \underbrace{\frac{1}{k} \sum_{i=0}^{N-1} A_i \left(\frac{l}{k}\right)^i}_{\text{first assembly}} + \underbrace{\frac{1}{\alpha} \sum_{i=0}^N B_i \left(\frac{l}{k}\right)^i}_{\text{reassembly and reaction}} \quad (3.14)$$

for both random and sequential assembly. In the limit of a very fast enzymatic reaction ($\alpha \rightarrow \infty$) only the first term on the right hand side of equation (3.14) remains. This term gives the time needed for the first assembly of the full complex. The second term that vanishes for $\alpha \rightarrow \infty$ is the additional time needed to carry out the reaction that also accounts for the possibility that complete complexes disassembles again and must reassemble before the reaction can take place. After some algebra, we find for random assembly

$$A_i^{\text{rand}} = \sum_{j=1}^{N-i} \frac{1}{i+j} \binom{N}{j-1}, \quad B_i^{\text{rand}} = \binom{N}{i}, \quad (3.15)$$

and for sequential assembly

$$A_i^{\text{seq}} = N - i, \quad B_i^{\text{seq}} = 1. \quad (3.16)$$

To find explicit formulas for the standard deviations is far more complicated, because they involve the second moment of the distribution. The general formula for the standard deviation of a sequential complex assembly can be found in Appendix A.1.

3.1.1 Irreversible binding of proteins

In the limit of irreversible protein binding ($l/k = 0$), the sequential binding time grows linearly with the number of proteins N while the average time for a random assembly increases only logarithmically (Figure 3.3A):

$$\tau_{\text{seq}} = \frac{1}{\alpha} + \frac{N}{k} \quad (3.17)$$

$$\tau_{\text{rand}} = \frac{1}{\alpha} + \frac{1}{k} \sum_{i=1}^N \frac{1}{i} \approx \frac{1}{\alpha} + \frac{1}{k} (\ln N + \gamma) \quad (3.18)$$

The approximation of the harmonic series is $\sum_{i=1}^N \frac{1}{i} \approx \ln N + \gamma$ with the Euler - Mascheroni constant $\gamma \sim 0.577$. Thus, the random assembly is always faster for irreversibly binding components. This result is due to the fact that there are $N!$ different assembly pathways that lead to the full complex, in contrast to the single pathway of a sequential assembly.

3.1.2 Reversible binding of proteins

For reversible binding of the proteins, random can still be faster than sequential assembly for a small number of components but eventually becomes much slower as the number of components grows (Figure 3.3B). For a balanced reversibility, where the binding and release of the proteins happen on the same time-scale ($K = k_{\text{off}}/k_{\text{on}} = 1$), the equations simplify to

$$\tau_{\text{seq}} = \frac{1}{k} \frac{N(N+1)}{2} + \frac{1}{\alpha} (1 + N) \quad (3.19)$$

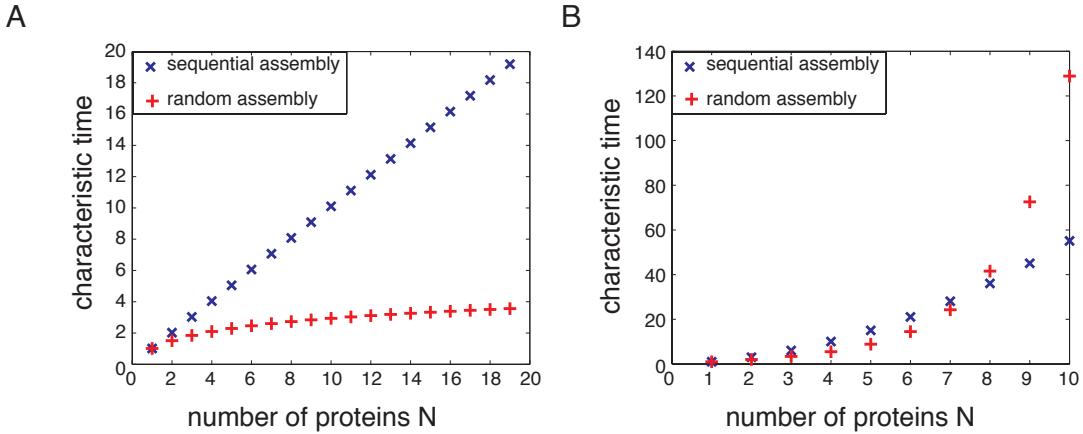


Figure 3.3: Mean reaction time. The characteristic time to assemble a complex vs. the number of proteins N for (A) irreversible binding ($k_{\text{off}} = 0$) and for (B) reversible binding ($k_{\text{off}} = k_{\text{on}}$). The catalytic rate constant is very large ($\alpha \rightarrow \infty$).

$$\tau_{\text{rand}} = \frac{1}{k} \sum_{i=0}^{N-1} \sum_{j=1}^{N-i} \frac{1}{i+j} \binom{N}{j-1} + \frac{1}{\alpha} 2^N. \quad (3.20)$$

With these equation it is easy to deduce that for $N \geq 8$ random assembly takes always longer, because

$$(1+N) \leq 2^N \quad \forall N \quad (3.21)$$

and

$$\frac{N(N+1)}{2} > \sum_{i=0}^{N-1} \sum_{j=1}^{N-i} \frac{1}{i+j} \binom{N}{j-1} \quad \text{for } N < 8 \quad (3.22)$$

$$\frac{N(N+1)}{2} < \sum_{i=0}^{N-1} \sum_{j=1}^{N-i} \frac{1}{i+j} \binom{N}{j-1} \quad \text{for } N \geq 8. \quad (3.23)$$

If the on-rate constant k becomes bigger than the off-rate constant l this relationship moves into the direction of irreversible binding and random complex assembly is faster even for a larger number of proteins.

3.1.3 Asymptotic behavior

This behavior gets even more pronounced, when the proteins bind even more reversible. Equation (3.14) and (3.15) imply that the average time for random assembly grows exponentially $\tau_{\text{rand}} \sim e^N$ with the number of components for reversible protein binding (cf. Figure 3.5). The average reaction time for a sequential assembly depends on the degree

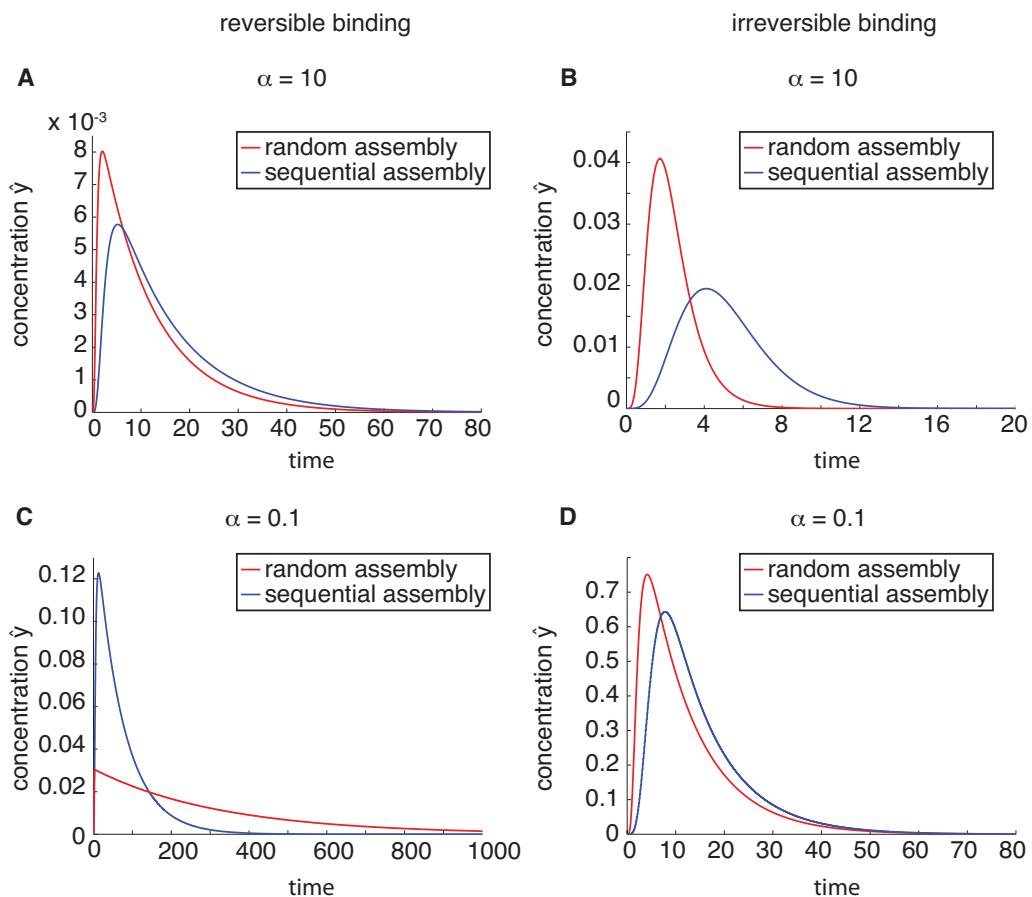


Figure 3.4: Kinetics of the full complex. Comparison of the characteristic time courses of the full complex (\hat{y}) for sequential (blue) and random (red) assembly. The left side shows irreversible ($k_{on} = 1, k_{off} = 0$) and the right side reversible ($k_{on} = 1, k_{off} = 1$) binding. The first row has a large enzymatic rate α and the second row a small one. The number of proteins is five.

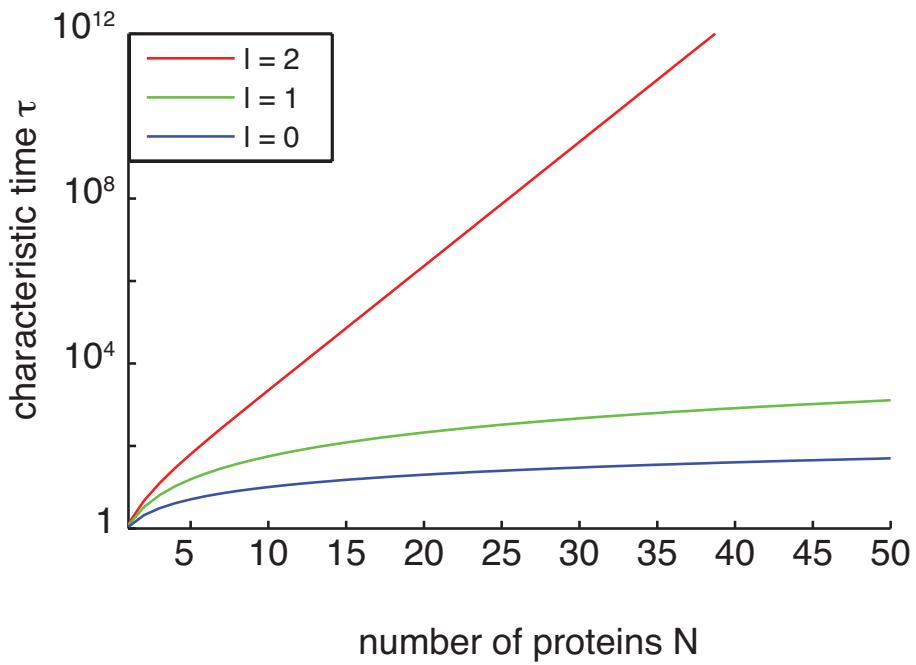


Figure 3.5: Asymptotic behavior. The characteristic time of random assembly for irreversible (blue), balanced reversible (green) and strongly reversible (red) binding. (The reference parameters are $k=1$, $\alpha=10$)

of reversibility. It grows linearly for $l < k$, quadratic for the special case of balanced reversibility ($k = l$) and exponentially when the reversibility gets more pronounced (Table 3.1 and Figure 3.5). Thus, the random assembly of a large number of components can become unfeasible slow. In contrast to that sequential assembly remains comparatively efficient provided that the reversibility of protein binding is not too strong.

Degree of reversibility		Sequential (τ_{seq})	Random (τ_{rand})
Irreversible binding	$l = 0$	N	$\ln N$
Weak reversibility	$0 < l < k$	N	e^N
Balanced reversibility	$l = k$	N^2	e^N
Strong reversibility	$l > k$	e^N	e^N

Table 3.1: Asymptotic reaction times. Dependence of the reaction time τ on the number of components N of the enzymatically active complex (to leading order).

These asymptotic results for large N do not cover the practically important case that the number of components is moderate. In that case we find that random assembly of an enzymatically active complex can drive the reaction at least as rapidly as sequential assembly. In particular, if the enzymatic reaction is rather fast, so that the fully assembled complex will, in the majority of cases, catalyze the reaction before disassembling again. However,

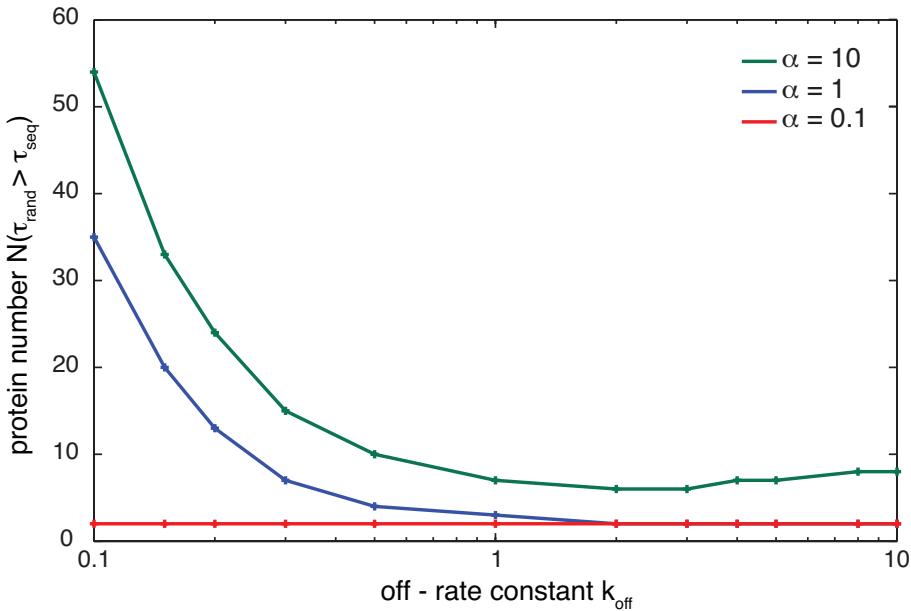


Figure 3.6: Efficiency range of random assembly. Protein number $N(\tau_{\text{rand}} > \tau_{\text{seq}})$ where the random assembly takes for the first time longer than the respective sequential assembly vs. the off-rate constant k_{off} for different enzymatic rates α .

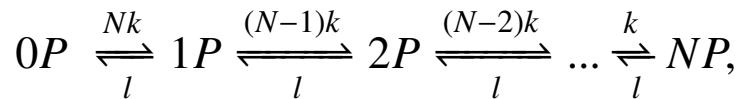
if the enzymatic reaction is rather slow sequential assembly will be faster, because it offers less opportunities to disassemble the complex again (Figure 3.6). To summarize, a random strategy of complex assembly is very efficient for a quasi-irreversible binding of proteins and for a reversible assembly of a moderate number of proteins ($N \leq 8$) to catalyze a fast reaction. In these cases, a sequential strategy will be less efficient, however, a sequential mechanism will be able to assemble a larger number of reversibly binding proteins in a reasonable time.

3.2 Influence of cooperative binding on the reaction time

A binding is called cooperative, if the affinity of a ligand depends on the amount of ligands that are already bound. In this section, we assume that the proteins in our model do not only have a binding site for the chromatin, but also for the other proteins. The aim of our analysis is to determine the effect of cooperatively binding components on the characteristic time of assembly. In the following we will investigate the effect of mutual cooperativity, where the affinity of all proteins is stabilized upon binding of a new factor and next-neighbor cooperativity, where only the affinity of the adjacent neighbors is modified.

3.2.1 Mutual cooperativity

For the mutual cooperativity, we take the reaction scheme of a random complex assembly as basis and modify the off-rate constants to account for the mutual stabilization of the proteins. We assume that the off-rate constant of each protein released from a complex of N proteins P is $\frac{1}{N}$ smaller than the respective off-rate constant in a non-cooperative random complex assembly. This assumption results in the following reaction scheme:



where $0P$, $1P$ and NP indicates zero proteins, 1 protein and N proteins, respectively.

The matrix R^{cop} in the time development of the assembly states

$$\dot{\vec{x}} = R^{\text{cop}} \vec{x} \quad (3.24)$$

(cf. Equation 3.13) is given by:

$$R_{i,j}^{\text{cop}} = \begin{cases} -(N-i)k - l & \text{for } j = i \text{ and } i \in \{0, \dots, N-1\} \\ -l - \alpha & \text{for } j = i = N \\ (N-i)k & \text{for } j = i-1 \text{ and } i \in \{1, \dots, N\} \\ l & \text{for } j = i+1 \text{ and } i \in \{0, \dots, N-1\} \\ 0 & \text{else.} \end{cases} \quad (3.25)$$

Analogously to the random assembly without cooperativity, it is possible to calculate the characteristic time analytically as follows.

$$\tau_{\text{cop}} = \frac{1}{\alpha} \sum_{i=0}^N \frac{1}{i!} \left(\frac{l}{k}\right)^i + \frac{1}{k} \sum_{j=1}^N \frac{1}{j!} \sum_{i=0}^{j-1} (j-1-i)! \left(\frac{l}{k}\right)^i \quad (3.26)$$

The characteristic time versus the number of assembling proteins for different degrees of reversibility are shown in Figure 3.7. The characteristic time increases for a more reversible binding of the proteins like before, however, the comparison with assembly of independently binding proteins shows that the cooperative binding shortens the characteristic time of reaction. The cooperative binding stabilizes the complexes with each protein that binds, thus, the probability of a disassembly is smaller than before. This probability is minimized, i.e. zero, for an irreversible binding scheme. If the binding becomes irreversible the cooperative and uncooperative random assembly become obviously identical.

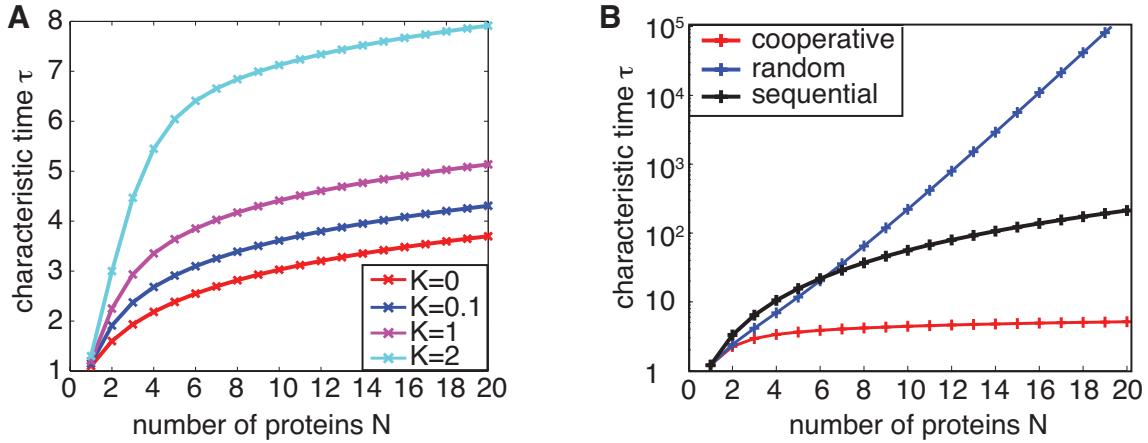


Figure 3.7: Stabilization through cooperative binding. (A) The characteristic time τ_{cop} for a cooperative binding for different equilibrium rate constants K . (B) Cooperative binding compared to a completely random and sequential binding scheme for a balanced reversibility ($l = k$). The reference parameter set is $k = 1$ and $\alpha = 10$.

3.2.2 Next-neighbor cooperativity

As a more realistic model we assume that the proteins bind on a ring and only neighboring proteins stabilize each others binding. This is equivalent to the binding of proteins on a chain with cyclic boundary conditions. The structure of the model is similar to the random assembly (cf. Equation (3.12) and (3.13)). For the random assembly all states with the same number of proteins bound were summarized in a single state. For the cooperative ring all states with the same symmetry and the same number of proteins bound are summarized in a single state. Two states have the same symmetry, if one state can be transferred into the other state by reflection or rotation of the ring. The binding rate constants are modified with respective weighting factors. If two states with the same symmetry are summarized in a single state the respective binding rate constant is $2k$. The off-rate constants have the corresponding weighting factors. Moreover, we assume that the off-rate constants l depend on the number of neighbors. If the released protein has no direct neighbor the off-rate constant equals l , for a single neighbor l' and for the maximum of two direct neighbors l'' . As an example we show the binding scheme for a ring (hexagon) that offers six binding spots for six proteins in Figure 3.8. The characteristic time of the final state with all proteins bound (reaction time) can be calculated by constructing a matrix that includes all the binding processes similar to Equation (3.5) and using Equation (3.10) to calculate the reaction time. The matrices of the binding processes M_n where n is the number of binding spots can be found in Appendix A.2 for $n = 2, \dots, 8$.

The product of the equilibrium rate constants $K = l/k$ should be the same for each shortest

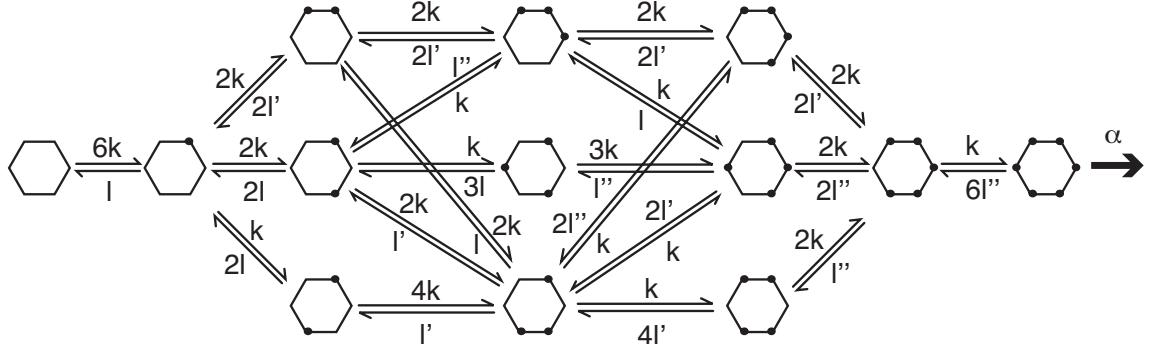


Figure 3.8: Binding scheme for next-neighbor cooperativity. Binding of 6 proteins on a hexagon that offers 6 binding spots. The binding rate constants are indicated for each reaction.

pathway through the scheme shown in Figure 3.8 and comparable schemes for different protein numbers. That implies the following equation:

$$l = \frac{l'^2}{l''} \quad (3.27)$$

We introduce a factor κ to tune the influence of the cooperativity in the following way

$$l' = \kappa l \quad \text{and} \quad l'' = \kappa^2 l \quad (3.28)$$

that obeys Equation (3.27). If the binding is not cooperative ($\kappa = 1$), the characteristic time equals the one for the random complex assembly (blue curve in Figure 3.9). For an increasing cooperativity that corresponds to a decreasing cooperativity constant (κ) the curves are flatten.

3.3 Control of complex assembly

To determine the control of the different rate constants (k_{on} , k_{off} and k_{cat}) on the mean reaction time, we calculated the control coefficients

$$C_p = \frac{1/p}{\tau} \frac{d\tau}{d(1/p)} \quad \text{for } p = k, l, \alpha \quad (3.29)$$

that quantify how a change in the rate affects the reaction time.

The sum of the control coefficients for the binding rate constant k , the off-rate constant l and the enzymatic rate constant α equals one ($C_k + C_l + C_\alpha = 1$). Since τ is a homogenous

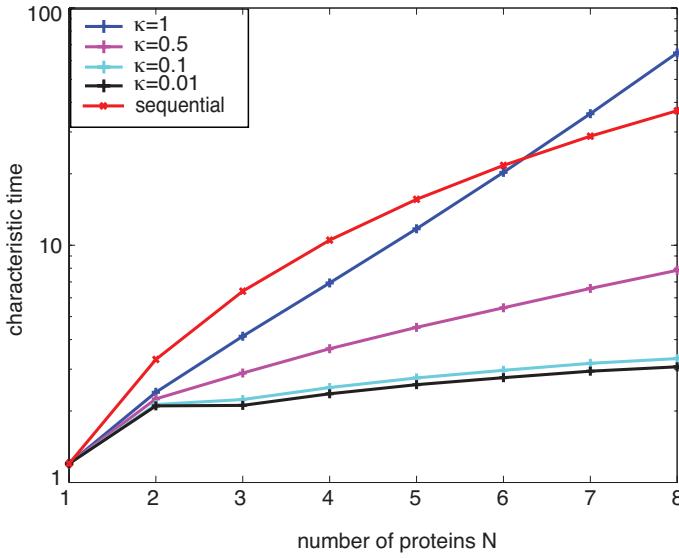


Figure 3.9: Characteristic time of reaction for next-neighbor cooperativity. The blue curve corresponds to the regular random assembly. The other curves show the increasing influence of cooperativity. The reference parameter set is $k = l = 1, \alpha = 10$. The curve for a sequential assembly is shown in red for comparison.

function of degree one, i.e.

$$\tau\left(\frac{k}{r}, \frac{l}{r}, \frac{\alpha}{r}\right) = r \tau(k, l, \alpha)$$

and Euler's theorem yields:

$$\frac{1}{k} \frac{d\tau}{d(1/k)} + \frac{1}{l} \frac{d\tau}{d(1/l)} + \frac{1}{\alpha} \frac{d\tau}{d(1/\alpha)} = \tau(k, l, \alpha) \quad (3.30)$$

3.3.1 Contribution of individual off-rate constant

We assume that each protein has a different off-rate constants l_i . Since the function τ is homogeneous we know that

$$\tau(k, l_1, \dots, l_n, \alpha) = \frac{1}{k} \frac{d\tau}{d(1/k)} + \sum_{i=1}^n \frac{1}{l_i} \frac{d\tau}{d(1/l_i)} + \frac{1}{\alpha} \frac{d\tau}{d(1/\alpha)}. \quad (3.31)$$

The reaction time τ must be the same for Equation 3.30 and 3.31.

Random complex assembly

In a random complex assembly each protein assumes each position in the chain of binding once, consequently each protein contributes in the same way and the derivates $\frac{d\tau}{d(1/l_i)}$ are identical for all proteins.

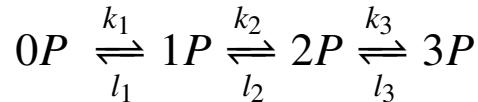
If all l_i for $i \in \{1, \dots, n\}$ are the same, we can conclude that:

$$\begin{aligned} &\implies \frac{1}{l} \frac{d\tau}{d(1/l)} = \sum_{i=1}^n \frac{1}{l_i} \frac{d\tau}{d(1/l_i)} = n \frac{1}{l_i} \frac{d\tau}{d(1/l_i)} \\ &\Leftrightarrow C_{l_i} = \frac{1}{n} C_l. \end{aligned} \quad (3.32)$$

Each off-rate constant l_i contributes thus a factor $\frac{1}{n}$ to the total control coefficient of the off-rate constant (C_l).

Sequential complex assembly

In contrast to the random complex assembly the proteins have a defined binding order in a sequential binding scheme. For that reason we conjecture that the derivates $\frac{d\tau}{d(1/l_i)}$ differ for the different proteins. As an example we will calculate the control coefficients for the following sequential binding of 3 proteins.



The time development of the states is given by

$$\dot{\vec{y}} = \begin{pmatrix} -k_1 & l_1 & 0 & 0 \\ l_1 & -k_1 - k_2 & l_2 & 0 \\ 0 & k_2 & -l_2 - k_3 & l_3 \\ 0 & 0 & k_3 & -l_3 - \alpha \end{pmatrix} \vec{y} \quad (3.33)$$

and the characteristic time can be calculated with the help of Equation 3.14 and 3.16. Assuming that all on-rate constants (k) and all off-rate constants (l) are identical, the control coefficients C_{l_i} are:

$$C_{l_1} = \frac{-l_1(\alpha(k_3 + l_2) + l_2 l_3)}{k_1 k_2 k_3 + \alpha(k_3(k_2 + l_1) + l_1 l_2 + k_1(k_2 + k_3 + l_2)) + l_1 l_2 l_3 + k_1(k_2 + l_2)l_3} \quad (3.34)$$

$$C_{l_2} = \frac{-(k_1 + l_1)l_2(\alpha + l_3)}{k_1 k_2 k_3 + \alpha(k_3(k_2 + l_1) + l_1 l_2 + k_1(k_2 + k_3 + l_2)) + l_1 l_2 l_3 + k_1(k_2 + l_2)l_3} \quad (3.35)$$

$$C_{l_3} = \frac{-(l_1 l_2 + k_1(k_2 + l_2))l_3}{k_1 k_2 k_3 + \alpha(k_3(k_2 + l_1) + l_1 l_2 + k_1(k_2 + k_3 + l_2)) + l_1 l_2 l_3 + k_1(k_2 + l_2)l_3}. \quad (3.36)$$

The denominator is the same for the three control coefficients.

The contribution of the individual control coefficients depends on the position in the chain and the occupation of the state.

to be continued ...

to be done ...

Figure 3.10: Comparision of control coefficients. The control coefficients C_{l_1} , C_{l_2} and C_{l_3} vs. the enzymatic rate α for (A) weak reversibility (B) balanced reversibility (C) strong reversibility.

3.3.2 Control coefficients for identically binding proteins

Given that all proteins have the same binding rate constants, we can use the general formulas (3.14) for random and sequential binding and (3.26) for the mutal cooperativity and deduce analytic expression for the control coefficients (cf. Appendix 3.2)

The on-rate constant k_{on} and the enzymatic rate α have a positive control and the off-rate constant k_{off} has a negative control on the characteristic time to assembly. The exact formulas can be found in the Appendix. The on-rate constant has the strongest control of the three rates. Small on-rate constants k and big off-rate constants l have the highest control. It is possible to deduce several relations between the control coefficients, as follows.

$$C_k > C_\alpha \text{ and } |C_l| > C_\alpha$$

$$|C_k^{\text{seq}}| > |C_l^{\text{seq}}|$$

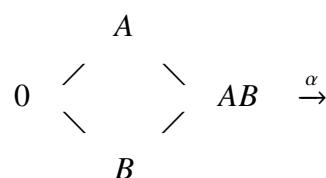
$$|C_k^{\text{rand}}| > |C_l^{\text{rand}}|$$

$$C_\alpha^{\text{rand}} > C_\alpha^{\text{seq}}$$

Thus, the on-rate constant k has a stronger control than the off-rate constant l and the enzymatic rate α and there is a stronger control for more reversible binding.

3.4 Flux analysis

3.4.1 Random binding of two proteins



Control coefficient	$\lim_{\frac{1}{k} \rightarrow 0}$	$\lim_{\frac{1}{k} \rightarrow \infty}$	$\lim_{\frac{1}{l} \rightarrow 0}$	$\lim_{\frac{1}{l} \rightarrow \infty}$
C_k^s	0	N	N	$\frac{\alpha N}{k+\alpha N}$
C_l^s	0	$\frac{\alpha}{\alpha+l} - N$	-N	0
C_α^s	1	$\frac{l}{\alpha+l}$	1	$\frac{k}{k+\alpha N}$
C_k^r	0	N	N	$\frac{\sum_{i=1}^N \frac{1}{k(1-i+N)}}{\frac{1}{\alpha} + \sum_{i=1}^N \frac{1}{k(1-i+N)}}$
C_l^r	0	$\frac{\alpha}{\alpha+lN} - N$	-N	0
C_α^r	1	$\frac{lN}{\alpha+lN}$	1	$\frac{1}{1+\alpha \sum_{i=1}^N \frac{1}{k(1-i+N)}}$
C_k^{coop}	0	N	N	$\frac{\sum_{i=1}^N \frac{1}{ki}}{\frac{1}{\alpha} + \sum_{i=1}^N \frac{1}{ki}}$
C_l^{coop}	0	$\frac{\alpha}{\alpha+l} - N$	-N	0
C_α^{coop}	1	$\frac{l}{\alpha+l}$	1	$\frac{\frac{1}{\alpha}}{\frac{1}{\alpha} + \sum_{i=1}^N \frac{1}{ki}}$

Table 3.2: Limits of the control coefficients.

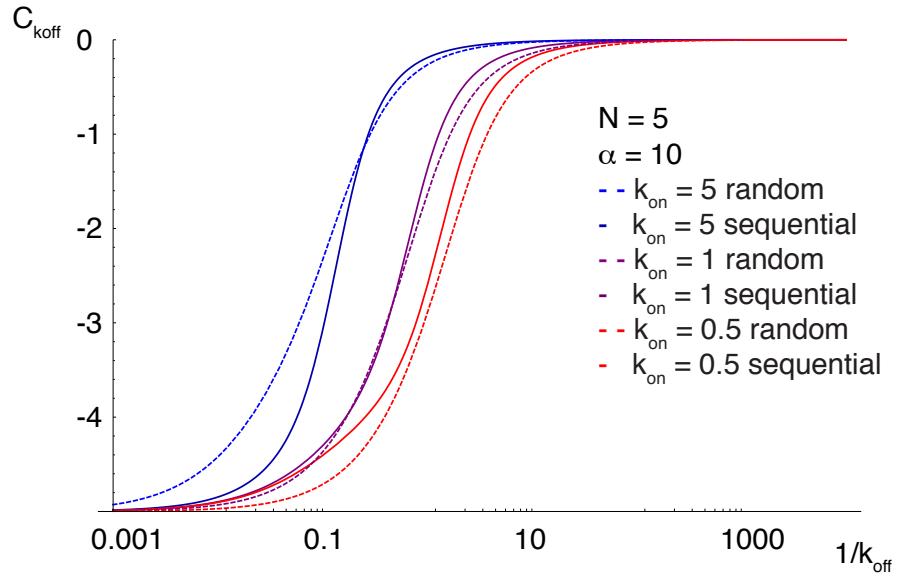


Figure 3.11: Control coefficient $C_{k_{\text{off}}} = C_l$ for 5 proteins and the reference parameter set: $\alpha = 10$ and $k = 5, 1$ and 0.5 .

Assume:

$$\dot{\vec{x}} = A \vec{x} \quad \text{with} \quad \vec{x}(0) = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad (3.37)$$

Flux through the upper way, i.e. protein A binds first:

$$v_{A \rightarrow AB} = -k_B x_A + l_B x_{AB} \quad (3.38)$$

Flux through the lower way, i.e. protein B binds first:

$$v_{B \rightarrow AB} = -k_A x_B + l_A x_{AB} \quad (3.39)$$

Fraction of the upper and lower flux, respectively:

$$\frac{\int v_{A \rightarrow AB} dt}{\int v_{total} dt} = \frac{k_A + l_B}{k_A + k_B + l_A + l_B} \quad (3.40)$$

and

$$\frac{\int v_{B \rightarrow AB} dt}{\int v_{total} dt} = \frac{k_B + l_A}{k_A + k_B + l_A + l_B} \quad (3.41)$$

since $\int v_{total} dt = -1$.

Ratio of the flux through the two pathways:

$$\frac{\int v_{A \rightarrow AB} dt}{\int v_{B \rightarrow AB} dt} = \frac{\int -k_B x_A + l_B x_{AB} dt}{\int -k_A x_B + l_A x_{AB} dt} = \frac{k_A + l_B}{k_B + l_A} \quad (3.42)$$

The flux depends on the binding rate constants only and is independent of the enzymatic step α . The ratio of the fluxes for a stationary system is the same.

Occupancy of the states

The ratio r of the occupancy of the two states:

$$r = \frac{\int x_A dt}{\int x_B dt} = \frac{k_A}{k_B} \left[\frac{\alpha (k_A + l_B) + l_B (k_A + k_B + l_A + l_B)}{\alpha (k_B + l_A) + l_A (k_A + k_B + l_A + l_B)} \right] \quad (3.43)$$

Limits for slow and fast enzymatic steps:

$$\lim_{\alpha \rightarrow 0} r = \frac{k_A l_B}{k_B l_A} \quad (3.44)$$

$$\lim_{\alpha \rightarrow \infty} r = \frac{k_A (k_A + l_B)}{k_B (k_B + l_A)} \quad (3.45)$$

3.4.2 Random binding of three proteins

There are 6 shortest pathways through the binding cube corresponding to the following fluxes:

$$\begin{aligned}
 v_{A \rightarrow AB} &= -k_B x_A + l_B x_{AB} \\
 v_{A \rightarrow AC} &= -k_C x_A + l_C x_{AC} \\
 v_{B \rightarrow AB} &= -k_A x_B + l_A x_{AB} \\
 v_{B \rightarrow BC} &= -k_C x_B + l_C x_{BC} \\
 v_{C \rightarrow AC} &= -k_A x_C + l_A x_{AC} \\
 v_{C \rightarrow BC} &= -k_B x_C + l_B x_{BC}
 \end{aligned}$$

The relative flux $f_{A \rightarrow AC}$ for initial binding of protein A, followed by the binding of C and finally B compared to the total flux through the cube is:

$$\begin{aligned}
 f_{A \rightarrow AC} &= \frac{\int v_{A \rightarrow AC} dt}{\int v_{total} dt} \\
 &= \frac{k_a^2(k_c + l_b) + l_b l_c (k_b + k_c + l_b + l_c) + k_a (k_b l_b + (k_c + l_b)(k_c + l_a + l_b) + (k_c + 2l_b)l_c)}{(k_a + k_c + l_a + l_c)(k_b + k_c + l_b + l_c)(k_a + k_b + k_c + l_a + l_b + l_c)}
 \end{aligned}$$

Irreversible binding of the proteins

The relative fluxes for irreversible binding schemes ($k_{off} = 0$) are:

$$f_{A \rightarrow AB} = \frac{k_A k_B}{(k_B + k_C)(k_A + k_B + k_C)} \quad (3.46)$$

$$f_{A \rightarrow AC} = \frac{k_A k_C}{(k_B + k_C)(k_A + k_B + k_C)} \quad (3.47)$$

$$f_{B \rightarrow AB} = \frac{k_A k_B}{(k_A + k_C)(k_A + k_B + k_C)} \quad (3.48)$$

$$f_{B \rightarrow BC} = \frac{k_B k_C}{(k_A + k_C)(k_A + k_B + k_C)} \quad (3.49)$$

$$f_{C \rightarrow AC} = \frac{k_A k_C}{(k_A + k_B)(k_A + k_B + k_C)} \quad (3.50)$$

$$f_{C \rightarrow BC} = \frac{k_B k_C}{(k_A + k_B)(k_A + k_B + k_C)} \quad (3.51)$$

Equal rate constants

Assuming that all rates are equal ($k = k_A = k_B = k_C = l_B = l_C$) except l_A it is only important, if the protein A binds first, second or last. Consequently, $f_{A \rightarrow AC} = f_{A \rightarrow AB}$, $f_{B \rightarrow AB} = f_{C \rightarrow AC}$ and $f_{B \rightarrow BC} = f_{C \rightarrow BC}$.

$$\begin{aligned} f_{A \rightarrow AC} &= \frac{\int v_{A \rightarrow AC} dt}{\int v_{total} dt} \\ &= -\frac{1}{2}k \left(-\frac{2}{3k + l_A} + \frac{1}{5k + l_A} \right) \end{aligned} \quad (3.52)$$

(3.53)

Limits:

$$\lim_{l_A \rightarrow 0} f_{A \rightarrow AC} = \frac{7}{30} \quad (3.54)$$

$$\lim_{l_A \rightarrow 0} f_{C \rightarrow AC} = \frac{6}{30} \quad (3.55)$$

$$\lim_{l_A \rightarrow 0} f_{C \rightarrow BC} = \frac{2}{30} \quad (3.56)$$

3.5 Optimal reversibility for finite protein concentrations

So far we assumed that the concentration of the binding proteins were much larger than the concentration of the binding substrate, granting that only a small fraction of the proteins is bound at any time. However, this assumption is generally not fulfilled. In the following we investigate the effects of finite concentrations, in particular the effect on the reaction time. Since the computational efforts are highly increased for non-linear differential equations, we will focus on a simplified model of random binding of two proteins A and B (Figure 3.12). The interesting effects can already be observed for such a simple model. The scheme can be translated into the following set of differential equations:

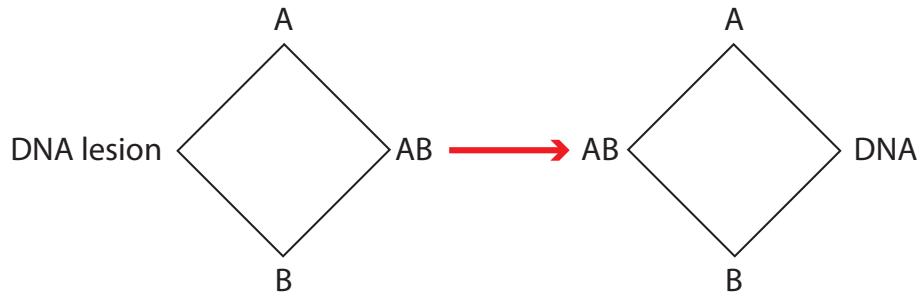


Figure 3.12: Modelscheme of a simplified model

$$\begin{aligned}
 \dot{y}_1 &= -(k_A C_A + k_B C_B) y_1 + l_A y_2 + l_B y_3 \\
 \dot{y}_2 &= k_A C_A y_1 - l_A y_2 - k_B C_B y_2 + l_B y_4 \\
 \dot{y}_3 &= k_B C_B y_1 - l_B y_3 - k_A C_A y_3 + l_A y_4 \\
 \dot{y}_4 &= k_A C_A y_3 + k_B C_B y_2 - (l_A + l_B) y_4 - \alpha y_4 \\
 \dot{y}_5 &= \alpha y_4 - l'_A y_5 - l'_B y_5 + k'_B C_B y_6 + k'_A C_A y_7 \\
 \dot{y}_6 &= l'_B y_5 - l'_A y_6 + k'_A C_A y_8 - k'_B C_B y_6 \\
 \dot{y}_7 &= l'_A y_5 - l'_B y_7 - k'_A C_A y_7 + k'_B C_B y_8 \\
 \dot{y}_8 &= l'_A y_6 + l'_B y_7 - k'_A C_A y_8 - k'_B C_B y_8 \\
 \dot{C}_A &= l_A (y_2 + y_4) + l'_A (y_5 + y_6) - k_A C_A (y_1 + y_3) - k'_A C_A (y_7 + y_8) \\
 \dot{C}_B &= l_B (y_3 + y_4) + l'_B (y_5 + y_7) - k_B C_B (y_1 + y_2) - k'_B C_B (y_6 + y_8)
 \end{aligned} \tag{3.57}$$

Numerical simulations of the system show that the characteristic time of reaction τ depends on the off-rate constant l in such a way that we observe a minimum for an intermediate off-rate constant (Figure 3.13A). The minimum gets more pronounced for a smaller concentration, however, we still observe a minimum if the concentration of each protein coincides with the concentration of the binding substrate. For increasing off-rate constants it takes a considerable amount of time to assemble the complex, since it becomes more and more improbable to bind a protein long enough to catalyze the enzymatic step. The concentration of the proteins goes immediately down and increases slowly which is equivalent to the immediate binding and the slow decrease of the concentration of the complex (Figure 3.13B and C). Only a small percentage of the proteins is bound even at the maximal accumulation of the proteins. For very small off-rate constants the effect is very different. The probability that the proteins stay bound is very high, thus the proteins accumulate to a very high level and decrease fast afterwards. At an intermediate level they rise again very slowly. This second maximum in the accumulation is explained by the fact that the proteins get trapped in states with a single protein bound. Only if a rare

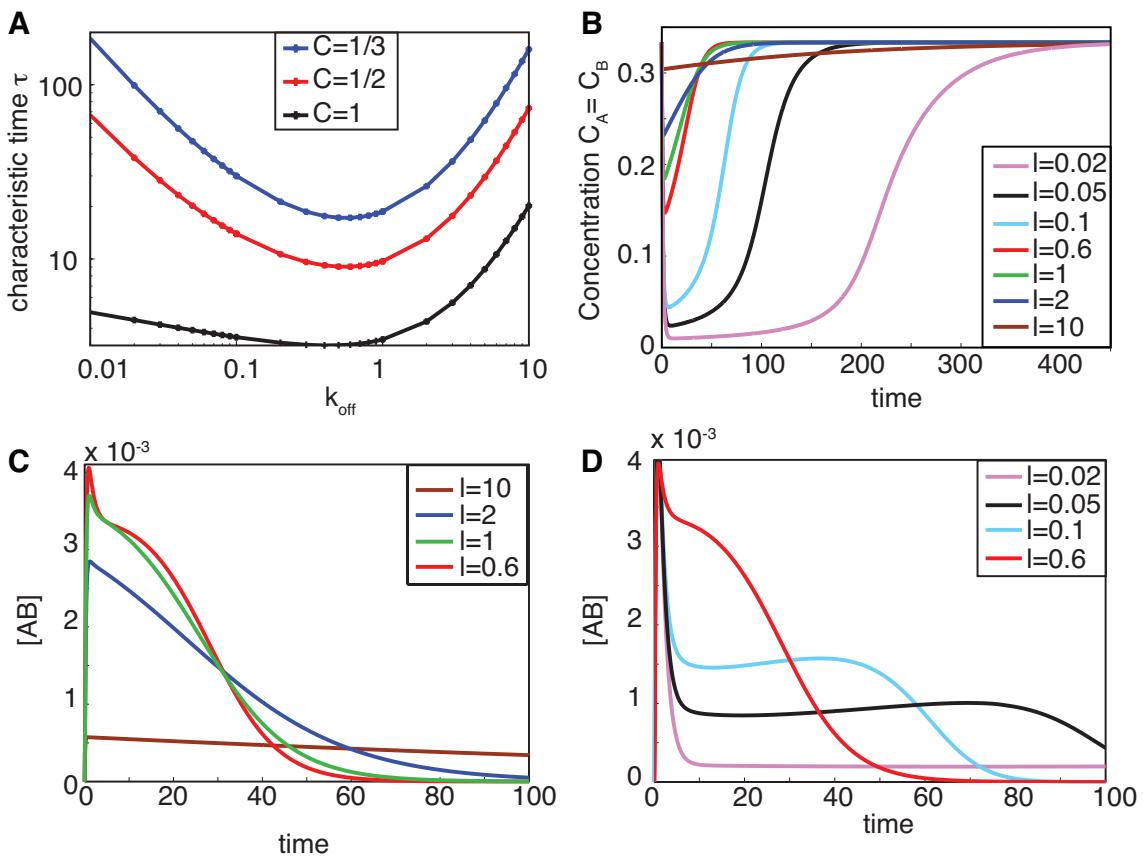


Figure 3.13: Optimal reversibility. (A) Minimum in the characteristic time for an intermediate reversibility for different concentrations of the proteins A and B. (B) Time development of the concentration ($C_A = C_B$) for different off-rate constants l . (C) and (D) Time development of the state AB that catalyzes α for different off-rate constants l . The concentration of the two proteins A and B is each 1/3 of the amount of damaged DNA substrate. Reference parameter set: $l=1, k=1, \alpha = 10$ and the concentration of the binding substrate is 1.

unbinding of one of the proteins occurs there is a probability that it assembles a complex with one of the other bound proteins. Thus, the accumulation of the protein complex AB that catalyzes the enzymatic step α consists of two parts, the immediate complex assembly when enough proteins are available and a second delayed part that has to wait for the release of the proteins trapped in states with a single protein bound. The time course for the minimal reaction time shows an intermediate behavior. The complex accumulates to a high level and decreases considerably fast, but there is no distinctive second maximum that delays the reaction time.

Chapter 4

Specificity of recognition

The NER machinery must recognize DNA lesions with high specificity to avoid erroneous incisions of undamaged DNA, which is potentially mutagenic. High specificity is a general requirement of recognition processes on the DNA. The affinity of the damage recognition factor XPC for unspecific binding to DNA is only a 100-fold lower than for DNA lesions [30, 32]. This low affinity difference raises the question how efficient NER would recognize damages if the specificity would be solely based on that affinity difference of XPC. To investigate that question, we assume for the moment that every binding event of XPC triggers the full NER process independent of the actual presence of a DNA lesion.

Measurements indicate that a DNA-bound XPC molecule occupies 20 to 30 base pairs of DNA [49, 76]. Given that a diploid human cell has $6.4 \cdot 10^9$ base pairs, we estimate approximately $2.1 \cdot 10^8 - 3.2 \cdot 10^8$ unspecific binding sites (B) for XPC, and hence the NER complex. This translates to a nuclear concentration of 1.21 mM - 1.77 mM (assuming a nuclear volume of 0.3 μL). The 100-fold lower affinity of XPC for unspecific binding corresponds to $K_D = 0.78$ mM using our model parameters.

On average the concentration of XPC bound to undamaged sites ($[XPC-B]$) will be

$$[XPC-B] = \frac{[XPC]_{\text{free}} [B]_{\text{free}}}{K_D} \quad (4.1)$$

and

$$[XPC]_{\text{total}} = [XPC]_{\text{free}} + [XPC-B] \quad (4.2)$$

for the total amount of XPC and

$$[B]_{\text{total}} = [B]_{\text{free}} + [\text{XPC}\text{-}B] \quad (4.3)$$

for the total amount of binding sites.

Our analysis indicates that between 61 – 69% of the XPC molecules (15000-18000) are non-specifically bound at any given time. This number agrees with recent measurements on XPC-GFP in vivo where 50% of the XPC pool was shown to be transiently bound to chromatin at all times [32]. From our model we estimate that it takes between 6 – 10 minutes to incision if XPC is already bound, consequently there should be $\sim 10^4 - 10^5$ incisions per hour at undamaged sites of the genome. Obviously, much higher damage specificity is required to prevent accidental incisions on undamaged DNA.

4.1 Kinetic proofreading

The model provides naturally a very efficient way of improving the specificity by several orders of magnitude. The reversible binding of the proteins combined with the possibility of DNA re-annealing, if it loses all stabilizing proteins, allows to abort processes that were started on undamaged DNA. In kinetic proofreading an enzyme-substrate complex is taken through a series of high-energy intermediates, at the expense of metabolic energy, before the final committing reaction step can take place. The stability of the complex is tested in each of the intermediate states and thus leading to a more faithful discrimination between the true substrate and close analogues than could be achieved by a single binding step [34]. Thus, we suggest that kinetic proofreading guarantees the high discriminative power of the NER system.

The specificity of the kinetic proofreading can be amplified by testing the binding stability of several proteins. The affinity of the damage recognition factor XPC for DNA depends on the distorted helical structure [46] and contributes obviously to the recognition process. However, there are two more factors suggested to have an affinity difference depending on the presence of the UV- lesions. XPA preferentially binds to kinks in the helical DNA structure that are induced by DNA lesions and therefore can also contribute to the discrimination between damaged and non-damaged DNA [6]. Moreover, the subunit composition of TFIIH is different when binding to a DNA lesion compared to its engagement in transcription [9, 10, 25], so that it may also bind with different affinities to damaged and non-damaged DNA [18, 82]. Therefore, we analyzed the potential contribution of these

three factors to damage recognition (Figure 4.1) in the following.

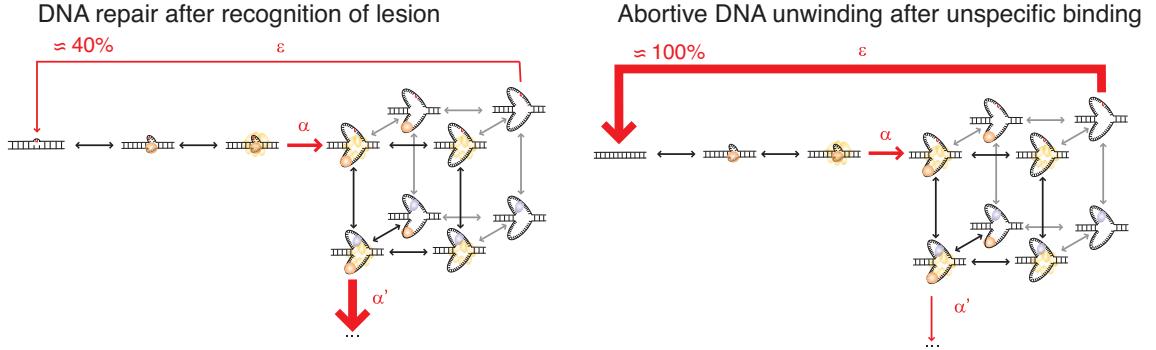


Figure 4.1: Kinetic proofreading mechanism. Damage recognition, DNA unwinding and kinetic proofreading by XPC, TFIID and XPA.

Let s be the specificity ratio for the individual factors, such that the stability of binding to undamaged DNA is s times smaller than the stability of the binding to damaged DNA: $k_{\text{off}}^{\text{unspecific}} = s k_{\text{off}}^{\text{specific}}$. A measure of erroneous dual incisions (as the committing step for NER) at undamaged DNA is given by the error fraction f , defined as the ratio of the characteristic times τ to dual incision:

$$f = \frac{\tau^{\text{specific}}}{\tau^{\text{unspecific}}} . \quad (4.4)$$

In the following we will calculate the minimal error fraction achievable by the kinetic proofreading scheme shown in Figure 4.1. This scheme considers only three proteins: XPC (C), TFIID (T) and XPA (A) and we assume that the concentrations of the free proteins C_P is constant $\frac{d}{dt}C_P = 0 \forall P \in \{C, T, A\}$, because only a small fraction of the proteins is bound at any time. The time development of all possible assembly states y_i can thus be written as matrix equation

$$\dot{\vec{y}} = A \vec{y}, \quad (4.5)$$

with $\vec{y}^T = \{y_0^I, y_C^I, y_T^I, y_{CT}^I, y_C^{II}, y_T^{II}, y_{CT}^{II}, y_0^{II}, y_{CTA}^{II}, y_{CA}^{II}, y_{TA}^{II}, y_A^{II}\}$. The matrix A for the submodule shown in Figure 4.1 is given by

$$A = \left(\begin{array}{cccccccccc} -k_C & l_C & 0 & 0 & 0 & 0 & \varepsilon & 0 & 0 & 0 & 0 \\ k_C & -l_C - k_T & l_T & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & k_T & -l_T - \alpha & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \alpha & -l_C - l_T & k_T & k_C & 0 & l_A & 0 & 0 & 0 \\ 0 & 0 & 0 & l_T & -l_C - k_T & 0 & k_C & 0 & l_A & 0 & 0 \\ A = & 0 & 0 & 0 & l_C & 0 & -k_C - l_T & k_T & 0 & l_A & 0 \\ 0 & 0 & 0 & 0 & l_C & l_T & -k_C - k_T & 0 & 0 & 0 & l_A \\ 0 & 0 & 0 & k_A & 0 & 0 & 0 & -l_C - l_T & k_T & k_C & 0 \\ 0 & 0 & 0 & 0 & k_A & 0 & 0 & l_T & -l_C - k_T & 0 & k_C \\ 0 & 0 & 0 & 0 & 0 & k_A & 0 & l_C & 0 & -k_C - l_T & k_T \\ 0 & 0 & 0 & 0 & 0 & 0 & k_A & 0 & l_C & l_T & -k_C - k_T \end{array} \right)$$

The on-rate constants are indicated by k_P and the off-rate constants by l_P , where $P \in \{\text{C}, \text{T}, \text{A}\}$ refers to the protein.

The characteristic time of state y_{CTA}^{II} that catalyzes the final committing step is taken as a measure for the time needed for kinetic proofreading. After the final committing step NER hast to proceed. In general, the zeroth (μ_0) and first (μ_1) moment as it is needed for the characteristic times can be calculated for each component using the matrix equation (cf. section 3.1 on assembly times). The characteristic time of full complex y_{CTA}^{II} is given by

$$\tau_8 = \frac{(-A^{-1} \mu_0)_8}{(\mu_0)_8}. \quad (4.6)$$

The error fraction is the ratio of the characteristic time for the specific and unspecific binding (cf. Equation (4.4)).

The time to incision calculated for the unspecific binding to the DNA ($\tau^{\text{unspecific}}$) is calculated the same way as for the specific binding, except that all off-rate constants k_{off} are substituted by $s_p^R \cdot k_{\text{off}}$. The specificity ratios s_p^R have a superscript index indicating the repair intermediate, which is I for damaged DNA and II for partially unwound DNA. The subscript index specifies the corresponding protein XPC (C), TFIIH (T) or XPA (A).

This yields an involved algebraic expression that can be further analyzed in limiting cases. For pronounced reversibility of protein binding ($k_{\text{off}} \rightarrow \infty$) we find for the minimal error fraction achievable by the kinetic proofreading scheme including XPC, TFIIH and XPA:

$$\frac{1}{f} = \frac{1}{6} s_C^I s_T^I (s_C^{\text{II}} + s_T^{\text{II}})(s_C^{\text{II}} + s_T^{\text{II}} + s_A^{\text{II}}). \quad (4.7)$$

If we assume all specificity ratios to be equal $s_C = s_T = s_A = 100$ for both repair intermediates, the formula yields $f_{\min} = 10^{-8}$. When only XPC and XPA have discriminating functions, f_{\min} is still $\sim 3 \cdot 10^{-6}$. In the absence of kinetic proofreading, if no re-annealing is permitted ($\varepsilon = 0$), the minimal error fraction f_{\min} would increase to 10^{-4} .

On the one hand the error fraction could be even smaller for our model of the NER system, since it includes two unwinding steps, but on the other hand f_{\min} represents the limit of extremely unstable protein complexes. Numerical simulations of the full model yield $f = 7 \cdot 10^{-9}$ when $s_C = s_A = s_T = 100$. When the kinetic proofreading is disabled artificially, i.e., $\varepsilon_{1,2} = 0$ the simulations yield $f = 1.8 \cdot 10^{-4}$, thus kinetic proofreading alone increases the specificity by more than five orders of magnitude. In the model simulations only $\sim 40\%$ of the DNA unwinding events around a true lesion are aborted. However, almost 100% re-anneal if the unspecific binding of XPC was followed by unwinding.

4.2 Tradeoff between specificity and efficiency

Since XPC has the strongest control on the rate of incision (Figure 2.12), we investigated the influence of the reversible XPC binding on the specificity and the speed of repair in detail. We determined numerically the error fraction (Equation 4.4) for the full model assuming a 100fold affinity difference between specific and unspecific binding for XPC, XPA and TFIIH. The specificity measured by the inverse error fraction increases as the reversibility of the XPC binding increases (Figure 4.1A). Pronounced stably bound proteins suppress kinetic proofreading by stabilizing the unwound DNA and thereby suppressing the possibility of reannealing. For example a 100fold increased dwell time of XPC increases the error fraction by six orders of magnitude $f \sim 10^{-2}$ (cf. Figure 4.2). However, a too low affinity of XPC for the DNA lesion leads to a very slow and thereby inefficient repair process (Figure 4.2B). Thus, specificity and efficiency of the NER system cannot be maximized simultaneously, and the kinetic design of the NER system must realize a tradeoff between these two objectives. The model predicts that a comparatively low XPC affinity together with a readily reversible XPC binding and other repair proteins, results in high specificity and efficiency.

The simplified model for the random assembly of two proteins with finite concentrations predicts an optimal reversibility that minimizes the reaction time (see Section 3.5). This effect can be found in the full model of NER, too. The time to incision is minimal for an intermediate value of the dissociation rate constant (k_{off}), while a more stable, quasi-irreversible XPC binding causes slower repair (Figure 4.2B). If the XPC binding is too strong, NER processes are started on almost every lesion and the proteins are trapped incomplete and hence inactive NER complexes, leading to a depletion of the free XPC pool.

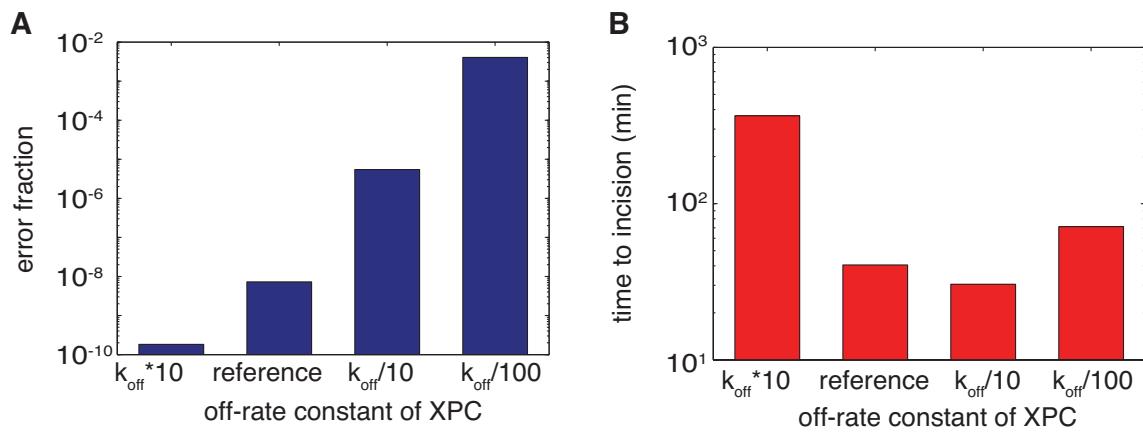


Figure 4.2: Specificity vs. efficiency. (A) Error fraction in the model for different dissociation rates of XPC. The affinity ratio of damaged and undamaged DNA is always 100. (B) Average time to incision for different dissociation rates of XPC.

Chapter 5

Discussion

Using NER as a model system, we developed a experimentally-based kinetic framework to investigate the assembly of multi-protein complexes on the chromatin in nuclei of living cells. During the modeling process we iterated several times between modeling and experiments. NER follows the general organization of chromatin associated processes and serves thus as a paradigm for such processes. From the specific NER model that was tightly connected to the experiments we moved to a more theoretical approach of modeling and analyzing the formation and functioning of multi-protein machineries acting on chromatin that allows to infer more general concepts. In the future it would be interesting to return to the experiments and find experimental validation of our theoretical results for specific systems.

Our results of the NER model reveal that after the recognition of a lesion and subsequent DNA unwinding, NER protein complexes assemble predominantly random and readily reversible. When protein complexes with the appropriate enzymatic activity are formed, they catalyze the essential repair steps: DNA unwinding, lesion excision, and repair synthesis. The accurate recognition of the lesion is a time-consuming process that limits the rate of repair. Our model suggests a proofreading mechanism for achieving high specificity in lesion recognition that is based on reversible DNA unwinding which is enabled and supported by reversible protein binding. The theoretical analysis of complex formation shows that random assembly is indeed fast for a limited number of components. However, a cooperative binding of the proteins could improve the efficiency even more.

Our kinetic framework that embraces the formation of multi-protein complexes and their performance of DNA repair, differs from previous experimentally-based mathematical models that described the kinetic behavior of individual proteins binding to chromatin based on FRAP data [13, 20, 27, 38]. Our model accounts simultaneously for the ki-

netic behavior of seven NER proteins and resolves the seemingly contradicting data of the long-term accumulation and the rapid exchange of the NER factors. Moreover, we obtain quantitative estimates for the kinetic parameters of the individual molecular interactions that explain the experimental data well and fall into biochemical expected ranges. However, some of the parameters can only be understood from a systems perspective. For example, the affinity of XPC and XPA is quite modest. Regarding the isolated parameter, one might expect that a higher affinity of these factors would render a more efficient repair. However, a closer analysis shows that efficiency and specificity of NER are systems properties that are determined by the interplay of the enzymatic action of several repair proteins and chromatin binding.

5.1 Efficiency

According to our model estimations the NER system becomes saturated only at a remarkably high number of DNA lesions. The half-saturation constant obtained from the model is 216,000 lesions per nucleus, which is far from the number of lesions introduced in the local damage experiments of 60 000 damages and far from naturally inflicted UV-damages. It is estimated that sunlight induces up to 30,000 DNA lesions per hour per skin cell. The maximal rate of repair per cell nucleus is estimated to be 6000 lesions per minute in our model, which is consistent with direct measurements of the rate of incision [39, 88]. Previous estimates of ~4 min for incising a single lesion were based on the dissociation rates of individual repair factors from damaged DNA in vivo [35, 66, 90]. However, these dissociation rates reflect the exchange of the protein and do not provide any information if incision has occurred or not and gives consequently no estimate about the time it takes to repair DNA lesions. In our model the repair proteins can bind several times to the same lesion before it is incised, which is consistent with the observed long-term accumulation of repair factors at sites of damage and the rapid exchange of the NER factors. The NER process takes consequently a considerable time to incise lesions, on average ~ 40 min for a single lesion, but with a high variance between the lesions. As our model correctly predicts the NER system is far from saturation and has a large capacity to process lesions in parallel. The average time to incise a single lesion is close to the time for several thousands of them.

5.2 Specificity

Systems analysis reveals that the moderate XPC affinity and the fast reversibility of binding appear advantageous for both specificity and efficiency of NER. Furthermore, the model suggests a mechanism to increase the specificity of the damage recognition process by several orders of magnitude. The specificity of XPC alone ($f_{\min} \sim 10^{-2}$ [30, 32]) would lead to many erroneous incisions. The involvement of multiple factors in damage recognition (XPA and possibly TFIIH) on the one hand, and kinetic proofreading [34] on the other hand greatly increases the specificity of the NER system. The two mechanisms together yield an error fraction in the recognition of lesions as low as $< 10^{-8}$. Kinetic proofreading is naturally realized in our model by the reversible unwinding of DNA. The unwound DNA requires the presence of bound NER factors to prevent re-annealing. When one or several of these factors bind with higher affinity to a lesion than undamaged DNA, the specificity is greatly amplified by this proofreading mechanism, since the stability of the complex is tested several times. This mechanism is particularly effective when the recognition factors cannot easily be saturated with DNA lesions and, indeed, the fit shows rather low initial affinities for XPC and XPA (K_d of $7 - 8 \mu M$, Table 2.4). A too low affinity of XPC, however, would strongly slow down repair. Our theoretical analysis suggests that the observed low XPC affinity guarantees a good trade-off between specificity and efficiency of NER.

In addition, the model indicates that the reversibility of protein binding is beneficial because it prevents the trapping of NER proteins in incomplete and consequently enzymatic inactive repair complexes. The maximal repair rate at an intermediate k_{off} value for XPC reflects this effect that the simplified model for the random binding of two proteins with finite concentrations already unraveled (Figure 3.13 and 4.2B). Moreover, the reversible binding of the proteins involved in NER ensures that they can be used in parallel for other processes, for example TFIIH switches rapidly between NER and transcription, that can still occur on undamaged parts of the DNA [33].

5.3 Validation and limitation of the model

Besides finding a consistent model that describes the available experimental data without making any artificial constraints, we were interested in validating the model with experiments that were not included in the modeling and fitting process. First of all the model correctly predicts the XPG kinetics for an increased amount of damage and proofs thereby that the NER system is indeed not saturated under the given experimental conditions. In-

dependent from that the model predicts the right XPC kinetics in XPA deficient cells. In both cases the amplitude is estimated correctly and gives an evidence that we estimated the kinetic parameters well. The most interesting predication is, however, the degradation of the 6-4PPs. In contrast to other studies that estimated short times of \sim 4 minutes to incise a single lesion, we calculated an average time of \sim 40 minutes for a single lesion. The experiment showed that the time scale of our repair kinetics is approximately correct. However, there are also limitations to our model. Our theoretical analysis of the multi-protein complex assembly suggests that cooperative binding of the individual proteins could improve the efficiency. Since there is no quantitative data on that with the exception of the already included strong cooperative binding of TFIIH after the initial binding of XPC and the binding of ERCC1/XPF that depends on the prior binding of XPA.

5.4 Conclusion: Multi-protein complex assembly on chromatin

This study provides a systems-level view on the assembly and function of a multi-protein machinery acting on chromatin and a connection to the underlying mechanisms of multi-protein complex assembly. Our results for the model of NER show that the repair factors bind rapidly reversible and assemble mainly in a random order into enzymatic active protein complexes that perform the steps of repair. Our theoretical analysis even proofs that a random complex assembly can be fast for a limited number of components. The live-cell imaging data shown in section 2.1 [44] and previous studies [66, 83] contradicts other models based on a irreversible binding of NER factors [63] or NER initiation by proteins other than XPC [40]. In contrast to other models the sequentiality of NER is given by the stepwise enzymatic modifications of the DNA substrate at which the proteins assemble and not by the binding sequence of the repair proteins. Our model includes also strong cooperative protein-protein interactions such as shown between XPA and ERCC1/XPF [83]. Thus, the assembly of NER complexes at chromatin does not only depend on protein-DNA interactions, but also on protein-protein interactions. The theoretical analysis even suggests that a moderate cooperative binding of the proteins leads to a faster assembly time. Importantly, however, cooperative protein-protein interactions generally do not lead to complexes with long-term stability since they are incompatible with the observed rapid exchange of all components. Long-term stability is not required because enzymatically active complexes need only be stable for an interval of time that is sufficient to carry out their catalytic function (e.g. DNA unwinding or dual incision). We even find that reversibility of protein binding is beneficial for NER by ensuring high specificity of lesion

recognition without reducing the efficiency of the repair process. Our analysis demonstrates the enormous potential of kinetic proofreading for damage recognition with the required high specificity.

5.5 Multi-protein complex assembly in other systems

The assembly of multi-protein complexes appears in many different areas of the cell, however, our analysis refers to complexes that perform a catalytic action and usually do not build up a structure with a long-term stability. Nevertheless, there are several processes that employ catalytic active complexes. For example many proteins involved in transcription and DNA replication have enzymatic activity that may affect histones and other proteins determining chromatin accessibility [80]. For the NER model we saw that decoupling of the progressive enzymatic modifications of the chromatin substrate and the binding of the individual proteins, leaves a considerable freedom for the binding mode of the proteins, while the sequence of enzymatic steps defines the underlying structure of the process. We conclude that the formation of chromatin-associated machineries is primarily coordinated in time by progressive enzymatic modifications of the chromatin substrate.

Several live-cell imaging studies together kinetic models have shown that many transcription factors, coactivators and RNA polymerases exchange rapidly in the transcription initiation complex like the components of the NER complex. Previously, polymerases were described as preformed complexes [71], however, more recent live-cell imaging studies imply that the individual components of pol I are assembled at the site of activity and the subunits are rapidly exchanged [20], [27].

The assembly of the transcription initiation complex with exchanging components has been considered inefficient [12], [20, 27]. Regarding our analysis such conclusions may need to be reevaluated from a systems perspective. Optimizing not only the efficiency but also the specificity of the process, leads to different interpretations. Kinetic proofreading based on reversible DNA unwinding enabled and supported by rapidly reversible protein binding, as suggested here for NER, may also ensure specific target site recognition in transcription. We assume that the conflict between specificity and efficiency uncovered here is a general design constraint for chromatin- associated machineries.

Chapter 6

Outlook: Possible continuations

6.1 Improving the quality of the model fit

The considerably large confidence intervals and correlations between the parameters show, that the estimation of the parameters could still be improved. However, a significant improvement could only be achieved with a more extensive and more systematic experimental data set. The size and structure of given experimental data sets for the individual proteins vary a lot. The data sets for RPA and PCNA span several hours and the time intervals between the measurements are small and equally distributed, however, the other proteins are measured for a shorter time in total and with varying intervals between the measurements. For the accuracy of the later parameters, it would also be interesting to measure the decreasing part of the RPA and PCNA curve more extensively.

In addition, to more accurate data sets, further independent measurements are needed to eliminate correlations. The accumulation and FLIP curves are always a sum of the contributions from the different repair intermediates. This leaves a certain freedom to distribute the individual contributions. The measurements for the perturbed NER process allow to constrain the distribution. Thus, more experiments of this type would be very helpful, however, a living cell can obviously not be modified arbitrarily. For sure, some of the wanted manipulations will not be feasible. Another option to gain a better insight in to the contribution of the proteins released from different repair intermediates, would be to perform FLIP experiments at different time points. At the moment, the FLIP experiments were started at the plateau level of the accumulation. The accumulation of the proteins on the different repair intermediates varies in time, hence, the FLIP curve should show different characteristic times reflecting the different weighted contributions.

6.2 Extending the NER model

Our analysis exposed the importance of the recognition process by showing its role for the specificity and efficiency of the whole process. The strong control of the damage recognition factor XPC (cf. Figure 2.12) emphasizes this importance. However, XPC is not the only protein involved in the initiation of the NER process. The UV-DDB complex, has been shown to ubiquitinate XPC resulting in a two-fold increase in XPC affinity for DNA in general [77]. It has been suggested that ubiquitination occurs when XPC is bound to a DNA lesion and is quickly reversed when XPC dissociates [77]. The experiments that were the basis of the kinetic model were obtained in cells that contain low concentrations of functional DDB2 (SV40-immortalized human cells and CHO cells). Therefore, the estimated effective XPC affinity accounts, only in part, for the effect of DDB2. In the complete absence of DDB2 (i.e. an even lower XPC affinity), the model predicts that NER becomes slower (Figure 4.2B), which has also been observed experimentally [55]. It would be of particular interest to study systematically how changes in the concentration of DDB2 affect the kinetics of NER and to understand the recognition process of NER in more detail. Moreover, it would be interesting to include the modification of the chromatin structure in the model. Since NER has to access the DNA directly, it is obvious that the chromatin structure has to be modulated to perform the repair and finally the chromatin must be returned to its original configuration. The exact process of the chromatin modification is yet unknown.

In addition to the UV-DDB complex that participates mainly in early stage of NER, it would be interesting to include later factors such as RFC, XRCC1-Lig3 and DNA polymerase. We expect that the basic conclusions given here on the functionality of the random and readily reversible assembly of NER complexes will not be changed fundamentally by this, however the certain parameter estimated have to be reevaluated.

Including the UV-DDB complex hints to another inaccuracy that could be removed. The UV-DDB complex is essential for the repair of CPDs that are also induced by UV-light and facilitates the repair of the 6-4PPS [77]. In the current model we focused on the repair of the 6-4PPs and neglect the repair of the CPDs, since they are repaired on a much slower time scale. Only a small fraction of CPDs is repaired within the first 8 hours (Figure 2.2).

6.3 Validating the model

Beside the experiments that would improve the quality of the data set, it would be interesting to validate more model predictions. In particular, it would be interesting to measure the kinetics of the repair intermediates. The degradation of the 6-4PP (Figure 2.15) was a the first step in this direction, since it is equivalent to the time development of the sum of the first three repair intermediates. A possible continuation in this direction would be to measure the kinetics of resynthesized DNA, by measuring the incorporation of radioactive nucleotides.

In our theoretical analysis we conclude that the random complex assembly is a general concept of the multi-protein complex assembly on chromatin. To confirm that conclusion it would be very interesting to actually apply this concept on another chromatin associated process like transcription. A concrete model based on our concept could again be experimentally validated and improved by further insights from the experimental side.

Bibliography

- [1] A. Aboussekha, M. Biggerstaff, M. K. Shivji, J. A Vilpo, V. Moncollin, V. N. Podust, M. Proti, U Hübscher, J. M. Egly, and R. D. Wood. Mammalian DNA nucleotide excision repair reconstituted with purified protein components. *Cell*, 80(6):859–868, Mar 1995.
- [2] Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter. *Molecular Biology of the Cell (5th edition)*. Garland Science, 2008.
- [3] S. J. Araújo, E. A. Nigg, and R. D. Wood. Strong functional interactions of TFIIH with XPC and XPG in human DNA nucleotide excision repair, without a preassembled repairosome. *Mol Cell Biol*, 21(7):2281–2291, Apr 2001.
- [4] R. Beukers, A. P. Eker, and P. H. Lohman. 50 years thymine dimer. *DNA Repair (Amst)*, 7(3):530–543, Mar 2008.
- [5] L. J. Blackwell and J. A. Borowiec. Human replication protein A binds single-stranded DNA in two distinct complexes. *Mol Cell Biol*, 14(6):3993–4001, Jun 1994.
- [6] U. Camenisch, R. Dip, S. B. Schumacher, B. Schuler, and H Naegeli. Recognition of helical kinks by xeroderma pigmentosum group A protein triggers DNA excision repair. *Nat Struct Mol Biol*, 13(3):278–284, Mar 2006.
- [7] U. Camenisch, R. Dip, M. Vitanescu, and H. Naegeli. Xeroderma pigmentosum complementation group A protein is driven to nucleotide excision repair sites by the electrostatic potential of distorted DNA. *DNA Repair (Amst)*, 6(12):1819–1828, Dec 2007.
- [8] F. Coin, V. Oksenych, and J. M. Egly. Distinct roles for the XPB/p52 and XPD/p44 subcomplexes of TFIIH in damaged DNA opening during nucleotide excision repair. *Mol Cell*, 26(2):245–256, Apr 2007.

- [9] F. Coin, V. Oksenykh, V. Mocquet, S. Groh, C Blattner, and J. M. Egly. Nucleotide excision repair driven by the dissociation of CAK from TFIIH. *Mol Cell*, 31(1):9–20, Jul 2008.
- [10] F. Coin, L. Proietti De Santis, T. Nardo, O Zlobinskaya, M. Stefanini, and J. M. Egly. p8/TTD-A as a repair-specific TFIIH subunit. *Mol Cell*, 21(2):215–226, Jan 2006.
- [11] D. Coverley, M. K. Kenny, M. Munn, W. D. Rupp, D P Lane, and R. D. Wood. Requirement for the replication protein SSB in human DNA excision repair. *Nature*, 349(6309):538–541, Feb 1991.
- [12] X. Darzacq, Y. Shav-Tal, V. de Turris, Y. Brody, S M. Shenoy, R. D. Phair, and R. H. Singer. In vivo dynamics of RNA polymerase II transcription. *Nat Struct Mol Biol*, 14(9):796–806, Sep 2007.
- [13] X. Darzacq and R. H. Singer. The dynamic range of transcription. *Mol Cell*, 30(5):545–546, Jun 2008.
- [14] W. L. de Laat, E. Appeldoorn, N. G. Jaspers, and J H. Hoeijmakers. DNA structural elements required for ERCC1-XPF endonuclease activity. *J Biol Chem*, 273(14):7835–7842, Apr 1998.
- [15] W. L. de Laat, E. Appeldoorn, K. Sugasawa, E. Weterings, N. G. Jaspers, and J. H. Hoeijmakers. DNA-binding polarity of human replication protein A positions nucleases in nucleotide excision repair. *Genes Dev*, 12(16):2598–2609, Aug 1998.
- [16] W. L. de Laat, N. G. Jaspers, and J. H. Hoeijmakers. Molecular mechanism of nucleotide excision repair. *Genes Dev*, 13(7):768–785, Apr 1999.
- [17] C. Dinant, M. S. Luijsterburg, T. Höfer, G von Bornstaedt, W. Vermeulen, A. B. Houtsmuller, and R. van Driel. Assembly of multiprotein complexes that control genome function. *J Cell Biol*, 185(1):21–26, Apr 2009.
- [18] R. Dip, U. Camenisch, and H. Naegeli. Mechanisms of DNA damage recognition and strand discrimination in human nucleotide excision repair. *DNA Repair (Amst)*, 3(11):1409–1423, Nov 2004.
- [19] R. Drapkin, J. T. Reardon, A. Ansari, J. C. Huang, L Zawel, K. Ahn, A. Sancar, and D. Reinberg. Dual role of TFIIH in DNA excision repair and in transcription by RNA polymerase II. *Nature*, 368(6473):769–772, Apr 1994.

- [20] M. Dundr, U. Hoffmann-Rohrer, Q. Hu, I. Grummt, L I. Rothblum, R. D. Phair, and T. Misteli. A kinetic framework for a mammalian RNA polymerase in vivo. *Science*, 298(5598):1623–1626, Nov 2002.
- [21] J. Essers, A. F. Theil, C. Baldeyron, W. A van Cappellen, A. B. Houtsmailler, R. Kanaar, and W. Vermeulen. Nuclear dynamics of PCNA in DNA replication and repair. *Mol Cell Biol*, 25(21):9350–9359, Nov 2005.
- [22] E. Evans, J. Fellows, A. Coffer, and R. D. Wood. Open complex formation around a lesion during nucleotide excision repair provides a structure for cleavage by human XPG protein. *EMBO J*, 16(3):625–638, Feb 1997.
- [23] E. Evans, J. G. Moggs, J. R. Hwang, J. M. Egly, and R D Wood. Mechanism of open complex and dual incision formation by human nucleotide excision repair factors. *EMBO J*, 16(21):6559–6573, Nov 1997.
- [24] G. Giglia-Mari, F. Coin, J. A. Ranish, D. Hoogstraten, A. Theil, N. Wijgers, N. G. Jaspers, A. Raams, M Argentini, P. J. van der Spek, E. Botta, M. Stefanini, J M Egly, R. Aebersold, J. H. Hoeijmakers, and W. Vermeulen. A new, tenth subunit of TFIIH is responsible for the dna repair syndrome trichothiodystrophy group A. *Nat Genet*, 36(7):714–719, Jul 2004.
- [25] G. Giglia-Mari, C. Miquel, A. F. Theil, P. O. Mari, D Hoogstraten, J. M. Ng, C. Dinnant, J. H. Hoeijmakers, and W. Vermeulen. Dynamic interaction of TTDA with TFIIH is stabilized by nucleotide excision repair in living cells. *PLoS Biol*, 4(6), Jun 2006.
- [26] L. C. Gillet and O. D. Schärer. Molecular mechanisms of mammalian global genome nucleotide excision repair. *Chem Rev*, 106(2):253–276, Feb 2006.
- [27] S. A. Gorski, S. K. Snyder, S. John, I. Grummt, and T Misteli. Modulation of RNA polymerase assembly dynamics in transcriptional regulation. *Mol Cell*, 30(4):486–497, May 2008.
- [28] C. M. Green and G. Almouzni. Local action of the chromatin assembly factor CAF-1 at sites of nucleotide excision repair in vivo. *EMBO J*, 22(19):5163–5174, Oct 2003.
- [29] Z. He, L. A. Henricksen, M. S. Wold, and C. J. Ingles. RPA involvement in the damage-recognition and incision steps of nucleotide excision repair. *Nature*, 374(6522):566–569, Apr 1995.

- [30] T. Hey, G. Lipps, K. Sugasawa, S. Iwai, F. Hanaoka, and G. Krauss. The XPC-HR23B complex displays high affinity and specificity for damaged DNA in a true-equilibrium fluorescence assay. *Biochemistry*, 41(21):6583–6587, May 2002.
- [31] J. H. Hoeijmakers. Genome maintenance mechanisms for preventing cancer. *Nature*, 411(6835):366–374, May 2001.
- [32] D. Hoogstraten, S. Bergink, J. M. Ng, V. H. Verbiest, M S. Luijsterburg, B. Geverts, A. Raams, C. Dinant, J H Hoeijmakers, W. Vermeulen, and A. B. Houtsmuller. Versatile DNA damage detection by the global genome nucleotide excision repair protein XPC. *J Cell Sci*, 121(Pt 17):2850–2859, Sep 2008.
- [33] D. Hoogstraten, A. L. Nigg, H. Heath, L. H. Mullenders, R. van Driel, J. H. Hoeijmakers, W. Vermeulen, and A. B. Houtsmuller. Rapid switching of TFIIH between RNA polymerase I. and II transcription and DNA repair in vivo. *Mol Cell*, 10(5):1163–1174, Nov 2002.
- [34] J. J. Hopfield. Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proc Natl Acad Sci U S A*, 71(10):4135–4139, Oct 1974.
- [35] A. B. Houtsmuller, S. Rademakers, A. L. Nigg, D Hoogstraten, J. H. Hoeijmakers, and W. Vermeulen. Action of DNA repair endonuclease ERCC1/XPF in living cells. *Science*, 284(5416):958–961, May 1999.
- [36] S. Iben, H. Tschochner, M. Bier, D. Hoogstraten, P Hozák, J. M. Egly, and I. Grummt. TFIIH plays an essential role in RNA polymerase I transcription. *Cell*, 109(3):297–306, May 2002.
- [37] C. J. Jones and R. D. Wood. Preferential binding of the xeroderma pigmentosum group A complementing protein to damaged DNA. *Biochemistry*, 32(45):12096–12104, Nov 1993.
- [38] T. S. Karpova, M. J. Kim, C. Spriet, K. Nalley, T J. Stasevich, Z. Kherrouche, L. Heilot, and J. G. McNally. Concurrent fast and slow cycling of a transcriptional activator at an endogenous promoter. *Science*, 319(5862):466–469, Jan 2008.
- [39] W. K. Kaufmann and S. J. Wilson. DNA repair endonuclease activity during synchronous growth of diploid human fibroblasts. *Mutat Res*, 236(1):107–117, Jul 1990.
- [40] K. J. Kesseler, W. K. Kaufmann, J. T. Reardon, T. C Elston, and A. Sancar. A mathematical model for human nucleotide excision repair: damage recognition by

random order assembly and kinetic proofreading. *J Theor Biol*, 249(2):361–375, Nov 2007.

- [41] L. Li, X. Lu, C. A. Peterson, and R. J. Legerski. An interaction between the DNA repair factor XPA and replication protein A appears essential for nucleotide excision repair. *Mol Cell Biol*, 15(10):5396–5402, Oct 1995.
- [42] H. Lodish, C. Berk, A. Kaiser, M. Krieger, M. Scott, A. Bretscher, H. Ploegh, and P. Matsudaira. *Molecular Cell Biology (6th edition)*. Elsevier, 2008.
- [43] M. S. Luijsterburg, J. Goedhart, J. Moser, H. Kool, B. Geverts, A. B. Houtsmuller, L. H. Mullenders, W. Vermeulen, and R. van Driel. Dynamic in vivo interaction of DDB2 E3 ubiquitin ligase with uv-damaged DNA is independent of damage-recognition protein XPC. *J Cell Sci*, 120(Pt 15):2706–2716, Aug 2007.
- [44] M.S. Luijsterburg, G. von Bornstaedt, A.M. Gourdin, A.Z. Politi, M. Moné, D.O. Warmerdam, J. Goedhart, W. Vermeulen, R. van Driel, and T. Höfer. Stochastic and Reversible Assembly of a Multiprotein DNA Repair Complex Ensures Accurate Target Site Recognition and Efficient Repair. *J Cell Biol*, submitted.
- [45] G. Maga and U. Hubscher. Proliferating cell nuclear antigen (PCNA): a dancer with many partners. *J Cell Sci*, 116(Pt 15):3051–3060, Aug 2003.
- [46] O. Maillard, S. Solyom, and H. Naegeli. An aromatic sensor with aversion to damaged strands confers versatility to DNA repair. *PLoS Biol*, 5(4), Apr 2007.
- [47] T. Matsunaga, C. H. Park, T. Bessho, D. Mu, and A. Sancar. Replication protein A confers structure-specific endonuclease activities to the XPF-ERCC1 and XPG subunits of human DNA repair excision nuclease. *J Biol Chem*, 271(19):11047–11050, May 1996.
- [48] I. Mellon, G. Spivak, and P. C. Hanawalt. Selective removal of transcription-blocking dna damage from the transcribed strand of the mammalian dhfr gene. *Cell*, 51(2):241–249, Oct 1987.
- [49] J. H. Min and N. P. Pavletich. Recognition of DNA damage by the Rad4 nucleotide excision repair protein. *Nature*, 449(7162):570–575, Oct 2007.
- [50] T. Misteli. Beyond the sequence: cellular organization of genome function. *Cell*, 128(4):787–800, Feb 2007.

- [51] V. Mocquet, J. P. Lainé, T. Riedl, Z. Yajin, M Y. Lee, and J. M. Egly. Sequential recruitment of the repair factors during NER: the role of XPG in initiating the resynthesis step. *EMBO J*, 27(1):155–167, Jan 2008.
- [52] M. J. Moné, T. Bernas, C. Dinant, F. A. Goedvree, E M. Manders, M. Volker, A. B. Houtsmuller, J. H. Hoeijmakers, W Vermeulen, and R. van Driel. In vivo dynamics of chromatin-associated complex formation in mammalian nucleotide excision repair. *Proc Natl Acad Sci U S A*, 101(45):15933–15937, Nov 2004.
- [53] M. J. Moné, M. Volker, O. Nikaido, L. H. Mullenders, A. A. van Zeeland, P. J. Verschure, E. M. Manders, and R van Driel. Local UV-induced DNA damage in cell nuclei results in local transcription inhibition. *EMBO Rep*, 2(11):1013–1017, Nov 2001.
- [54] J. Moser, H. Kool, I. Giakzidis, K. Caldecott, L H. Mullenders, and M. I. Foustneri. Sealing of chromosomal DNA nicks during nucleotide excision repair requires XRCC1 and DNA ligase III alpha in a cell-cycle-specific manner. *Mol Cell*, 27(2):311–323, Jul 2007.
- [55] J. Moser, M. Volker, H. Kool, S. Alekseev, H Vrieling, A. Yasui, A. A. van Zeeland, and L. H. Mullenders. The UV-damaged DNA binding protein mediates efficient targeting of the nucleotide excision repair complex to UV-induced photo lesions. *DNA Repair (Amst)*, 4(5):571–582, May 2005.
- [56] D. Mu, D. S. Hsu, and A. Sancar. Reaction mechanism of human DNA repair excision nuclease. *J Biol Chem*, 271(14):8285–8294, Apr 1996.
- [57] D. Mu, M. Wakasugi, D. S. Hsu, and A. Sancar. Characterization of reaction intermediates of human excision repair nuclease. *J Biol Chem*, 272(46):28971–28979, Nov 1997.
- [58] L. H. Mullenders, A. A. van Zeeland, and A. T. Natarajan. The localization of ultraviolet-induced excision repair in the nucleus and the distribution of repair events in higher order chromatin loops in mammalian cells. *J Cell Sci Suppl*, 6:243–262, 1987.
- [59] A. O'Donovan, A. A. Davies, J. G. Moggs, S. C. West, and R D. Wood. XPG endonuclease makes the 3' incision in human DNA nucleotide excision repair. *Nature*, 371(6496):432–435, Sep 1994.

- [60] C. J. Park and B. S. Choi. The protein shuffle. sequential interactions among components of the human nucleotide excision repair pathway. *FEBS J*, 273(8):1600–1608, Apr 2006.
- [61] D. Perdiz, P. Grof, M. Mezzina, O. Nikaido, E Moustacchi, and E. Sage. Distribution and repair of bipyrimidine photoproducts in solar UV-irradiated mammalian cells. Possible role of Dewar photoproducts in solar mutagenesis. *J Biol Chem*, 275(35):26732–26742, Sep 2000.
- [62] V. N. Podust, N. Tiwari, S. Stephan, and E. Fanning. Replication factor C disengages from proliferating cell nuclear antigen (PCNA) upon sliding clamp formation, and PCNA itself tethers DNA polymerase delta to DNA. *J Biol Chem*, 273(48):31992–31999, Nov 1998.
- [63] A. Polit, M. J. Moné, A. B. Houtsmuller, D Hoogstraten, W. Vermeulen, R. Heinrich, and R. van Driel. Mathematical modeling of nucleotide excision repair reveals efficiency of sequential assembly strategies. *Mol Cell*, 19(5):679–690, Sep 2005.
- [64] S. E. Polo, D. Roche, and G. Almouzni. New histone incorporation marks sites of UV repair in human cells. *Cell*, 127(3):481–493, Nov 2006.
- [65] William H. Press. *Numerical recipes : the art of scientific computing*. Cambridge University Press, 3 edition, September 2007.
- [66] S. Rademakers, M. Volker, D. Hoogstraten, A. L. Nigg, M. J. Moné, A. A. Van Zee land, J. H. Hoeijmakers, A B. Houtsmuller, and W. Vermeulen. Xeroderma pigmentosum group A. protein loads as a separate factor onto DNA lesions. *Mol Cell Biol*, 23(16):5755–5767, Aug 2003.
- [67] T. Riedl, F. Hanaoka, and J. M. Egly. The comings and goings of nucleotide excision repair factors on damaged DNA. *EMBO J*, 22(19):5293–5303, Oct 2003.
- [68] P. Robins, C. J. Jones, M. Biggerstaff, T. Lindahl, and R D. Wood. Complementation of DNA repair in xeroderma pigmentosum group A cell extracts by a protein with affinity for damaged DNA. *EMBO J*, 10(12):3913–3921, Dec 1991.
- [69] R. Roy, J. P. Adamczewski, T. Seroz, W. Vermeulen, J P. Tassan, L. Schaeffer, E. A. Nigg, J. H. Hoeijmakers, and J M. Egly. The MO15 cell cycle kinase is associated with the TFIIH transcription-DNA repair factor. *Cell*, 79(6):1093–1101, Dec 1994.
- [70] L. Schaeffer, R. Roy, S. Humbert, V. Moncollin, W Vermeulen, J. H. Hoeijmakers, P. Chambon, and J. M. Egly. DNA repair helicase: a component of BTF2 (TFIIH) basic transcription factor. *Science*, 260(5104):58–63, Apr 1993.

- [71] P. Seither, S. Iben, and I. Grummt. Mammalian RNA polymerase I exists as a holoenzyme with associated basal transcription factors. *J Mol Biol*, 275(1):43–53, Jan 1998.
- [72] K. K. Shivji, M. K. Kenny, and R. D. Wood. Proliferating cell nuclear antigen is required for DNA excision repair. *Cell*, 69(2):367–374, Apr 1992.
- [73] M. K. Shivji, V. N. Podust, U. Hübscher, and R. D. Wood. Nucleotide excision repair DNA synthesis by DNA polymerase epsilon in the presence of PCNA, RFC, and RPA. *Biochemistry*, 34(15):5011–5017, Apr 1995.
- [74] A. M. Sijbers, W. L. de Laat, R. R. Ariza, M. Biggerstaff, Y. F. Wei, J. G. Moggs, K. C. Carter, B. K. Shell, E. Evans, M. C. de Jong, S. Rademakers, J. de Rooij, N. G. Jaspers, J. H. Hoeijmakers, and R. D. Wood. Xeroderma pigmentosum group F caused by a defect in a structure-specific DNA repair endonuclease. *Cell*, 86(5):811–822, Sep 1996.
- [75] K. Sugasawa and F. Hanaoka. Sensing of DNA damage by XPC/Rad4: one protein for many lesions. *Nat Struct Mol Biol*, 14(10):887–888, Oct 2007.
- [76] K. Sugasawa, J. M. Ng, C. Masutani, S. Iwai, P. J. van der Spek, A. P. Eker, F. Hanaoka, D. Bootsma, and J. H. Hoeijmakers. Xeroderma pigmentosum group C protein complex is the initiator of global genome nucleotide excision repair. *Mol Cell*, 2(2):223–232, Aug 1998.
- [77] K. Sugasawa, Y. Okuda, M. Saijo, R. Nishi, N. Matsuda, G. Chu, T. Mori, S. Iwai, K. Tanaka, K. Tanaka, and F. Hanaoka. UV-induced ubiquitylation of XPC protein mediated by UV-DDB-ubiquitin ligase complex. *Cell*, 121(3):387–400, May 2005.
- [78] P. Sung, V. Bailly, C. Weber, L. H. Thompson, L. Prakash, and S. Prakash. Human xeroderma pigmentosum group D gene encodes a DNA helicase. *Nature*, 365(6449):852–855, Oct 1993.
- [79] A. Tapias, J. Auriol, D. Forget, J. H. Enzlin, O. D. Schärer, F. Coin, B. Coulombe, and J. M. Egly. Ordered conformational changes in damaged DNA induced by nucleotide excision repair factors. *J Biol Chem*, 279(18):19074–19083, Apr 2004.
- [80] H. van Attikum and S. M. Gasser. Crosstalk between histone modifications during the DNA damage response. *Trends Cell Biol*, 19(5):207–217, May 2009.
- [81] A. van Hoffen, J. Venema, R. Meschini, A. A. van Zeeland, and L. H. Mullenders. Transcription-coupled repair removes both cyclobutane pyrimidine dimers and 6-4

- photoproducts with equal efficiency and in a sequential way from transcribed DNA in xeroderma pigmentosum group C fibroblasts. *EMBO J*, 14(2):360–367, Jan 1995.
- [82] G. Villani and N. Tanguy Le Gac. Interactions of DNA helicases with damaged DNA: possible biological consequences. *J Biol Chem*, 275(43):33185–33188, Oct 2000.
- [83] M. Volker, M. J. Moné, P. Karmakar, A. van Hoffen, W. Schul, W. Vermeulen, J. H. Hoeijmakers, R. van Driel, A. A. van Zeeland, and L. H. Mullenders. Sequential assembly of the nucleotide excision repair factors in vivo. *Mol Cell*, 8(1):213–224, Jul 2001.
- [84] M. Wakasugi, J. T. Reardon, and A. Sancar. The non-catalytic function of XPG protein during dual incision in human nucleotide excision repair. *J Biol Chem*, 272(25):16030–16034, Jun 1997.
- [85] M. Wakasugi and A. Sancar. Order of assembly of human DNA repair excision nuclease. *J Biol Chem*, 274(26):18759–18768, Jun 1999.
- [86] C. A. Weber, E. P. Salazar, S. A. Stewart, and L. H. Thompson. ERCC2: cDNA cloning and molecular characterization of a human nucleotide excision repair gene with high homology to yeast RAD3. *EMBO J*, 9(5):1437–1447, May 1990.
- [87] G. S. Winkler, K. Sugasawa, A. P. Eker, W. L. de Laat, and J H. Hoeijmakers. Novel functional interactions between nucleotide excision DNA repair proteins influencing the enzymatic activities of TFIIH, XPG, and ERCC1-XPF. *Biochemistry*, 40(1):160–165, Jan 2001.
- [88] N. Ye, M. S. Bianchi, N. O. Bianchi, and G. P. Holmquist. Adaptive enhancement and kinetics of nucleotide excision repair in humans. *Mutat Res*, 435(1):43–61, Sep 1999.
- [89] M. Yokoi, C. Masutani, T. Maekawa, K. Sugasawa, Y Ohkuma, and F. Hanaoka. The xeroderma pigmentosum group C protein complex XPC-HR23B plays an important role in the recruitment of transcription factor IIH to damaged DNA. *J Biol Chem*, 275(13):9870–9875, Mar 2000.
- [90] A. Zotter, M. S. Luijsterburg, D. O. Warmerdam, S Ibrahim, A. Nigg, W. A. van Cappellen, J. H. Hoeijmakers, R van Driel, W. Vermeulen, and A. B. Houtsmuller. Recruitment of the nucleotide excision repair endonuclease XPG to sites of UV-induced DNA damage depends on functional TFIIH. *Mol Cell Biol*, 26(23):8868–8879, Dec 2006.

Appendix A

Appendix

A.1 Standard deviation σ for a sequential assembly

The standard deviation can be calculated analytically with Equation (3.11). After some algebra we find the following expression:

$$\sigma_{\text{seq}} = \sqrt{\frac{U}{k^2} + \frac{2V}{k\alpha} + \frac{W}{\alpha^2}}, \quad (\text{A.1})$$

with

$$\begin{aligned} U &= \sum_{i=1}^{N-1} \left[\binom{i+2}{3} + (N-i) i^2 \left(\frac{l}{k}\right)^{i-1} \right] + \sum_{i=N-1}^{2N-2} \binom{2N+1-i}{3} \left(\frac{l}{k}\right)^i \\ &= \frac{\left(\frac{l}{k}\right)^{2+2N} + N - 4 \left(\frac{l}{k}\right)^{1+N} \left(\left(\frac{l}{k}-1\right)N - 1 \right) - \left(\frac{l}{k}\right)(4 + \frac{l}{k} + \frac{l}{k}N)}{\left(-1 + \frac{l}{k}\right)^4}, \end{aligned} \quad (\text{A.2})$$

$$\begin{aligned} V &= \sum_{i=1}^N \binom{i+1}{2} \left(\frac{l}{k}\right)^i + \sum_{i=N+1}^{2N-1} \binom{2n+1-i}{2} \left(\frac{l}{k}\right)^i \\ &= \frac{\left(\frac{l}{k}\right) \left[-1 + \left(\frac{l}{k}\right)^N \left(1 + \left(\frac{l}{k}\right) \left(-1 + \left(\frac{l}{k}\right)^N - 2N \right) + 2N \right] }{\left(\left(\frac{l}{k}\right) - 1\right)^3}, \end{aligned} \quad (\text{A.3})$$

and

$$W = \left(\sum_{i=0}^N \left(\frac{l}{k} \right)^i \right)^2 = \left(\frac{\left(\frac{l}{k} \right)^{N+1} - 1}{\left(\frac{l}{k} \right) - 1} \right)^2. \quad (\text{A.4})$$

Since this expression is very complicated we compute some selected cases. For an irreversible binding of the proteins ($l = 0$), we find

$$\tau_{\text{seq}}^{\text{Irrev}} = \frac{1}{\alpha} + \frac{N}{k} \quad \sigma_{\text{seq}}^{\text{Irrev}} = \sqrt{\frac{1}{\alpha^2} + \frac{N}{k^2}}. \quad (\text{A.5})$$

with the limit

$$\lim_{\alpha \rightarrow \infty} \frac{\sigma_{\text{seq}}^{\text{Irrev}}}{\tau_{\text{seq}}^{\text{Irrev}}} = \frac{\sqrt{\frac{1}{\alpha^2} + \frac{N}{k^2}}}{\frac{1}{\alpha} + \frac{N}{k}} = \frac{1}{\sqrt{N}} \quad (\text{A.6})$$

for a fast enzymatic rate α .

For a balanced reversible binding of the components ($l=k$) we find

$$\sigma_{\text{seq}}^{\text{Bal}} = \sqrt{\frac{\frac{1}{6}N(1+N)(1+N+N^2)}{k^2} + \frac{\frac{1}{3}N(1+N)(1+2N)}{k\alpha} + \frac{(1+N)^2}{\alpha^2}} \quad (\text{A.7})$$

with the limit

$$\lim_{N \rightarrow \infty} \frac{\sigma_{\text{seq}}^{\text{Bal}}}{\tau_{\text{seq}}^{\text{Bal}}} = \sqrt{\frac{2}{3}} \approx 0.82. \quad (\text{A.8})$$

A.2 Reaction matrices for next-neighbor cooperativity

The time development of all possible assembly states y_i can be described by

$$\dot{\vec{y}} = M_n \vec{y}, \quad (\text{A.9})$$

if the system is linear. The dimension of \vec{y} is determined by the number of possible assembly states including the substrate with no proteins bound. The explicit form of M_n with $n \in \{1, \dots, 8\}$ is given in the following.

For one binding partners:

$$M_1 = \begin{pmatrix} -k & l \\ k & -\alpha - l \end{pmatrix}$$

For two binding partners:

$$M_2 = \begin{pmatrix} -2k & l & 0 \\ 2k & -k - l & 2l' \\ 0 & k & -\alpha - 2l' \end{pmatrix}$$

For three binding partners:

$$M_3 = \begin{pmatrix} -3k & l & 0 & 0 \\ 3k & -2k - l & 2l' & 0 \\ 0 & 2k & -k - 2l' & 3l'' \\ 0 & 0 & k & -\alpha - 3l'' \end{pmatrix}$$

For four binding partners:

$$M_4 = \begin{pmatrix} -4k & l & 0 & 0 & 0 & 0 \\ 4k & -3k - l & 2l & 2l' & 0 & 0 \\ 0 & k & -2k - 2l & 0 & l'' & 0 \\ 0 & 2k & 0 & -2k - 2l' & 2l' & 0 \\ 0 & 0 & 2k & 2k & -k - 2l' - l'' & 4l'' \\ 0 & 0 & 0 & 0 & k & -\alpha - 4l'' \end{pmatrix}$$

For five binding partners:

$$M_5 = \begin{pmatrix} -5k & l & 0 & 0 & 0 & 0 & 0 & 0 \\ 5k & -4k-l & 2l & 0 & 2l' & 0 & 0 & 0 \\ 0 & 2k & -3k-2l & l'' & 0 & 2l' & 0 & 0 \\ 0 & 0 & k & -2k-2l'-l'' & 2k & 0 & 2l' & 0 \\ 0 & 2k & 0 & 2l' & -3k-2l' & l & 0 & 0 \\ 0 & 0 & 2k & 0 & k & -2k-l-2l' & 2l'' & 0 \\ 0 & 0 & 0 & 2k & 0 & 2k & -k-2l'-2l'' & 5l'' \\ 0 & 0 & 0 & 0 & 0 & 0 & k & -\alpha-5l'' \end{pmatrix}$$

For six binding partners:

$$M_6 = \left(\begin{array}{ccccccccc} -6k & l & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 6k & -5k-l & 2l' & 2l & 2l & 0 & 0 & 0 & 0 \\ 0 & 2k & -4k-2l' & 0 & 0 & 2l' & l & 0 & 0 \\ 0 & 2k & 0 & -4k-2l & 0 & l'' & l' & 0 & 3l \\ 0 & k & 0 & 0 & -4k-2l & 0 & 0 & 0 & 0 \\ 0 & 0 & 2k & k & 0 & -3k-2l'-l'' & 0 & 0 & 0 \\ 0 & 0 & 2k & 2k & 4k & 0 & -3k-l-2l' & 2l'' & 0 \\ 0 & 0 & 2k & 0 & 0 & 2k & k & -2k-2l'-2l'' & 0 \\ 0 & 0 & 0 & 0 & 0 & k & 0 & -2k-l-2l'-l'' & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2l'' \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -2k-4l' & 3k \\ 0 & 0 & 0 & 0 & 0 & 0 & 2k & -k-2l'-3l'' & 0 \\ 0 & 0 & 0 & k & 0 & 0 & l'' & 0 & 6l'' \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -3k-3l & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & k & 0 & -\alpha - 6l'' \end{array} \right)$$

For seven binding partners:

0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
l''	$2l'$	0	0	0	0	0	0	0	0
l	0	0	0	0	0	0	0	0	0
l'	0	l'	$2l$	0	0	0	0	0	0
l'	0	l'	0	0	0	0	0	0	0
...	0	l	l'	0	0	0	0	0	0
0	0	l'	0	0	0	0	0	0	0
0	0	0	l''	0	0	0	0	0	0
l''	$2l'$	l''	$2l'$	0	0	0	0	0	0
0	0	0	0	$2l''$	$2l'$	$2l'$	0	0	0
$-3k - l - 2l' - 2l''$	0	0	0	$2l''$	$2l'$	$2l'$	0	0	0
0	$-3k - l - 4l'$	0	0	0	l''	$2l''$	0	0	0
0	0	$-3k - 4l' - l''$	0	$2l''$	0	$2l'$	$4l'$	0	0
0	0	0	$3k - 2l - 2l' - l''$	0	$2l''$	0	$2l''$	0	0
k	0	k	0	$-2k - 2l' - 4l''$	0	0	0	$2l'$	0
k	k	0	$2k$	0	$-2k - l - 2l' - 3l''$	0	0	$2l''$	0
k	$2k$	k	0	0	$-2k - 4l' - 2l''$	0	$2l''$	0	0
0	0	k	k	0	0	0	$-2k - 4l' - 2l''$	l''	0
0	0	0	0	$2k$	$2k$	$2k$	$-k - 2l' - 5l''$	$2k$	$-\alpha - 8l''$
0	0	0	0	0	0	0	0	k	

A.3 Explicit calculation of the control coefficients

Control coefficient for sequential assembly

$$\begin{aligned}
C_k^{\text{seq}} &= \frac{1/k}{\tau} \frac{d\tau}{d(1/k)} & (\text{A.10}) \\
&= \frac{1/k}{\tau} \frac{d}{d(1/k)} \left[\frac{1}{\alpha} \sum_{i=0}^N \left(\frac{l}{k}\right)^i + \sum_{i=0}^{N-1} (N-i) \frac{l^i}{k^{i+1}} \right] \\
&= \frac{1}{\tau} \left[\frac{1}{\alpha} \sum_{i=0}^N i \left(\frac{l}{k}\right)^i + \sum_{i=0}^{N-1} (N-i)(i+1) \frac{l^i}{k^{i+1}} \right]
\end{aligned}$$

$$\implies C_k^{\text{seq}} > 0 \quad (\text{A.11})$$

A slower on-time ($1/k$) leads to a slower reaction.

$$\begin{aligned}
C_l^{\text{seq}} &= \frac{1/l}{\tau} \frac{d\tau}{d(1/l)} & (\text{A.12}) \\
&= \frac{1/l}{\tau} \frac{d}{d(1/l)} \left[\frac{1}{\alpha} \sum_{i=0}^N \left(\frac{l}{k}\right)^i + \sum_{i=0}^{N-1} (N-i) \frac{l^i}{k^{i+1}} \right] \\
&= -\frac{1}{\tau} \left[\frac{1}{\alpha} \sum_{i=0}^N i \left(\frac{l}{k}\right)^i + \sum_{i=0}^{N-1} i(N-i) \frac{l^i}{k^{i+1}} \right] \\
&= -\frac{\frac{1}{\alpha} \sum_{i=0}^N i \left(\frac{l}{k}\right)^i + \sum_{i=0}^{N-1} i(N-i) \frac{l^i}{k^{i+1}}}{\frac{1}{\alpha} \sum_{i=0}^N \left(\frac{l}{k}\right)^i + \sum_{i=0}^{N-1} (N-i) \frac{l^i}{k^{i+1}}}
\end{aligned}$$

$$\implies C_l^{\text{seq}} < 0 \quad (\text{A.13})$$

A longer dwell-time ($1/l$) leads to a faster reaction.

$$\begin{aligned}
C_\alpha^{\text{seq}} &= \frac{1/\alpha}{\tau} \frac{d\tau}{d(1/\alpha)} & (\text{A.14}) \\
&= \frac{1/\alpha}{\tau} \frac{d}{d(1/l)} \left[\frac{1}{\alpha} \sum_{i=0}^N \left(\frac{l}{k}\right)^i + \sum_{i=0}^{N-1} (N-i) \frac{l^i}{k^{i+1}} \right] \\
&= \frac{1}{\tau} \frac{1}{\alpha} \sum_{i=0}^N \left(\frac{l}{k}\right)^i
\end{aligned}$$

Control coefficient for random assembly

$$\begin{aligned}
C_k^{\text{rand}} &= \frac{1/k}{\tau} \frac{d\tau}{d(1/k)} & (\text{A.15}) \\
&= \frac{1/k}{\tau} \frac{d}{d(1/k)} \left[\frac{1}{\alpha} \sum_{i=0}^N \binom{N}{i} \left(\frac{l}{k}\right)^i + \frac{1}{k} \sum_{j=0}^{N-1} \left(\sum_{i=1}^{N-j} \frac{(j-1+i)!}{(i-1)!} \frac{(N-j-i)!}{(N+1-i)!} \right) \left(\frac{l}{k}\right)^j \right] \\
&= \frac{1}{\tau} \left[\frac{1}{\alpha} \sum_{i=0}^N i \binom{N}{i} \left(\frac{l}{k}\right)^i + \sum_{j=0}^{N-1} \left(\sum_{i=1}^{N-j} \frac{(j-1+i)!}{(i-1)!} \frac{(N-j-i)!}{(N+1-i)!} \right) (j+1) \frac{l^j}{k^{j+1}} \right] \\
&= \frac{1}{\tau} \left[\frac{1}{\alpha} \sum_{i=0}^N i \binom{N}{i} \left(\frac{l}{k}\right)^i + \sum_{j=0}^{N-1} \sum_{i=1}^{N-j} \frac{j+1}{i+j} \binom{N}{i-1} \frac{l^j}{k^{j+1}} \right]
\end{aligned}$$

$$\implies C_k^{\text{rand}} > 0 \quad (\text{A.16})$$

$$\begin{aligned}
C_l^{\text{rand}} &= \frac{1/l}{\tau} \frac{d\tau}{d(1/l)} & (\text{A.17}) \\
&= \frac{1/l}{\tau} \frac{d}{d(1/l)} \left[\frac{1}{\alpha} \sum_{i=0}^N \binom{N}{i} \left(\frac{l}{k}\right)^i + \frac{1}{k} \sum_{j=0}^{N-1} \left(\sum_{i=1}^{N-j} \frac{(j-1+i)!}{(i-1)!} \frac{(N-j-i)!}{(N+1-i)!} \right) \left(\frac{l}{k}\right)^j \right] \\
&= -\frac{1}{\tau} \left[\frac{1}{\alpha} \sum_{i=0}^N i \binom{N}{i} \left(\frac{l}{k}\right)^i + \sum_{j=0}^{N-1} \left(\sum_{i=1}^{N-j} \frac{(j-1+i)!}{(i-1)!} \frac{(N-j-i)!}{(N+1-i)!} \right) j \frac{l^j}{k^{j+1}} \right] \\
&= -\frac{1}{\tau} \left[\frac{1}{\alpha} \sum_{i=0}^N i \binom{N}{i} \left(\frac{l}{k}\right)^i + \sum_{j=0}^{N-1} \sum_{i=1}^{N-j} \frac{j}{i+j} \binom{N}{i-1} \frac{l^j}{k^{j+1}} \right]
\end{aligned}$$

$$\implies C_l^{\text{rand}} < 0 \quad (\text{A.18})$$

$$\begin{aligned}
C_\alpha^{\text{rand}} &= \frac{1/\alpha}{\tau} \frac{d\tau}{d(1/\alpha)} & (\text{A.19}) \\
&= \frac{1/\alpha}{\tau} \frac{d}{d(1/\alpha)} \left[\frac{1}{\alpha} \sum_{i=0}^N \binom{N}{i} \left(\frac{l}{k}\right)^i + \frac{1}{k} \sum_{j=0}^{N-1} \left(\sum_{i=1}^{N-j} \frac{(j-1+i)!}{(i-1)!} \frac{(N-j-i)!}{(N+1-i)!} \right) \left(\frac{l}{k}\right)^j \right] \\
&= \frac{1}{\tau} \frac{1}{\alpha} \sum_{i=0}^N \binom{N}{i} \left(\frac{l}{k}\right)^i
\end{aligned}$$

Control coefficient for cooperative random assembly

$$\begin{aligned}
C_k^{\text{coop}} &= \frac{1/k}{\tau} \frac{d\tau}{d(1/k)} & (\text{A.20}) \\
&= \frac{1/k}{\tau} \frac{d}{d(1/k)} \left[\frac{1}{\alpha} \sum_{i=0}^N \frac{1}{i!} \left(\frac{l}{k}\right)^i + \frac{1}{k} \sum_{j=1}^N \frac{1}{j!} \sum_{i=0}^{j-1} (j-1-i)! \left(\frac{l}{k}\right)^i \right] \\
&= \frac{1}{\tau} \left[\frac{1}{\alpha} \sum_{i=0}^N \frac{i}{i!} \left(\frac{l}{k}\right)^i + \frac{1}{k} \sum_{j=1}^N \frac{1}{j!} \sum_{i=0}^{j-1} (i+1)(j-1-i)! \left(\frac{l}{k}\right)^i \right] \\
\implies C_k &> 0
\end{aligned}$$

$$\begin{aligned}
C_l^{\text{coop}} &= \frac{1/l}{\tau} \frac{d\tau}{d(1/l)} & (\text{A.21}) \\
&= \frac{1/l}{\tau} \frac{d}{d(1/k)} \left[\frac{1}{\alpha} \sum_{i=0}^N \frac{1}{i!} \left(\frac{l}{k}\right)^i + \frac{1}{k} \sum_{j=1}^N \frac{1}{j!} \sum_{i=0}^{j-1} (j-1-i)! \left(\frac{l}{k}\right)^i \right] \\
&= -\frac{1}{\tau} \left[\frac{1}{\alpha} \sum_{i=0}^N \frac{i}{i!} \left(\frac{l}{k}\right)^i + \frac{1}{k} \sum_{j=1}^N \frac{1}{j!} \sum_{i=0}^{j-1} i(j-1-i)! \left(\frac{l}{k}\right)^i \right] \\
\implies C_l &< 0
\end{aligned}$$

$$\begin{aligned}
C_\alpha^{\text{coop}} &= \frac{1/\alpha}{\tau} \frac{d\tau}{d(1/\alpha)} & (\text{A.22}) \\
&= \frac{1}{\tau} \frac{1}{\alpha} \sum_{i=0}^N \frac{1}{i!} \left(\frac{l}{k}\right)^i
\end{aligned}$$