TECHNISCHE
UNIVERSITÄT
DARMSTADT

Physik

A theoretical investigation of protein recruitment during the DNA damage response and of the dynamics of DNA replication

Vom Fachbereich Physik
der Technischen Universität Darmstadt

zur Erlangung des Grades
eines Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigte Dissertation von
M.Sc. Daniel Löb
aus Erbach

Darmstadt 2013
D17

Referent: Prof. Dr. Barbara Drossel
Korreferent: Prof. PhD. Marco Durante

Tag der Einreichung: 02.07.2013
Tag der Prüfung: 22.07.2013

Abstract (de)

Diese Arbeit befasst sich mit drei theoretischen Untersuchungen zu den zellbiologischen Themen DNA-Doppelstrangbruchreparatur und DNA-Replikation. In der ersten Untersuchung geht es um die Rekrutierung von Proteinen der Schadensantwort bei DNA-Doppelstrangbrüchen zu Schadensstellen. Danach werden die notwendigen Bedingungen für das Auftreten von mehreren stationären Zuständen und Oszillationen in generischen Netzwerken der Proteinkomplexbildung identifiziert. Schließlich werden die der Organisation der DNA-Replikation auf Genomebene zugrunde liegenden Mechanismen analysiert.

Aus experimentellen Untersuchungen ist bekannt, dass sich das Rekrutierungsverhalten des an der Doppelstrangbruchschadensantwort beteiligten pfadunabhängigen Proteins NBS1 jenseits einer gewissen Schadensdichte qualitativ ändert, von schadensdichteabhänig zu schadensdichteunabhängig. Ein minimales Computermodell der Rekrutierung von NBS1 (als Teil des MRN-Komplexes) und mehrerer mit ihm interagierender Proteine wird dargelegt und mit experimentellen Daten verglichen. Durch das Modell zeigt sich, dass die Veränderung in der Rekrutierungsdynamik als eine Folge der sich verschiebenden Wichtigkeit zweier unterschiedlicher MRN-Bindungsinteraktionen interpretiert werden kann. Bei niedrigen Schadensdichten dominiert die Bindung im weitläufigeren Schadensumfeld, während bei höheren Schadensdichten die Bindung direkt an die Doppelstrangenden wichtiger wird.

Es folgt eine Untersuchung generischer Netzwerke der Proteinrekrutierung bzw. Proteinkomplexbildung mit dem Ziel, die Voraussetzungen für komplexe dynamische Phänomene wie Multistabilität und Oszillationen in diesen Netzwerken zu finden. Dabei zeigt sich, dass auf Assoziations- und Dissoziationsreaktionen beschränkte Netzwerke, in denen die Gesamtkonzentrationen der kleinsten, “elementaren” Proteine erhalten sind, mindestens vier dieser elementaren Proteine benötigen um Multistabilität oder Oszillationen zu zeigen. Zudem wird mathematisch bewiesen, dass mehrere stationäre Zustände für Netzwerke mit nur drei elementaren Proteinen unmöglich sind.

DNA-Replikation in Säugetieren und dem Menschen unterscheidet sich qualitativ von der als wohlverstanden geltenden Replikation in einfacheren Eukaryoten wie etwa Hefe. Auf der Skala von Chromosomen und Chromosomsegmenten zeigt der Replikationsprozess einen geordneten Ablauf, während sich die mikroskopische Dynamik stochastisch verhält. Es wird ein stochastisches Computermodell präsentiert, welches die minimalen Bestandteile enthält, die notwendig sind um diese dynamischen Eigenschaften zu reproduzieren. Als Bestandteile wurden identifiziert: eine schnell diffundierende und begrenzt vorhandene Reaktionskomponente, induziertes Auslösen von Replikationsstartpunkten in Abhängigkeit des Abstandes zu nahe gelegenen Replikationsgabeln, eine konstante Geschwindigkeit der Replikationsgabel, die während der frühen S-Phase reduziert wird, und der Beginn des Replikationsprozesses an Startpunkten im Euchromatin.

Die Ergebnisse des Modells sind konsistent, sowohl mit experimentellen Daten, als auch mit der Literatur, so dass es sich bei dem hier präsentierten Replikationsmodell um eines der am umfassendsten mit experimentellen Daten verglichenen Modelle der DNA-Replikation handelt, die derzeit existieren. Die Kombination der Modellergebnisse mit einer durch ein Monte-Carlo-Modell erzeugten dreidimensionalen DNA-Konformation zeigt, dass die Packung des Chromatins die Hauptursache für die bei der Mikroskopie von DNA-Replikation in Säugetierzellen beobachtete Fokusanordnung ist.

Die in dieser Arbeit präsentierten Untersuchungen kombinieren Methoden der Physik und der angewandten Mathematik mit Problemen aus dem Gebiet der Zellbiologie. Dieser inhärent interdisziplinäre Charakter macht die Arbeit für eine Leserschaft von sowohl Physikern als auch Biologen interessant.

Abstract (en)

In this thesis, three interrelated theoretical investigations on the cell-biological topics of DNA double strand break response and DNA replication are presented. The first investigation is concerned with the recruitment of DNA double strand break response proteins to DNA damage sites. In the second, necessary conditions for the appearance of multiple steady states and oscillations in generic protein complex assembly networks are identified. Lastly in the third investigation, the mechanisms underlying the genome-scale organization of DNA replication are analyzed.

It is known from experiment that the recruitment of the pathway-independent double strand break response protein NBS1 qualitatively changes its dynamics beyond a certain damage density, from damage density dependent to damage density independent. A minimal computer model of the recruitment of NBS1 (contained in the MRN complex) and several interacting proteins is developed and compared to experimental data. It becomes evident from the model that the change in dynamics can be interpreted a consequence of the shifting importance of two different MRN binding interactions. At low damage densities, binding in the wider damage site vicinity dominates, while at higher damage densities, binding directly to the damaged double strand ends becomes more important.

Next, generic protein recruitment/protein complex assembly networks are investigated to find the prerequisites of complex dynamical effects such as multistability and oscillation. It is shown that if the networks are limited to association and dissociation reactions and if the protein numbers are conserved for the indivisible "elementary" proteins participating, then at least four such elementary protein species must be present for multistability or oscillations to appear. A rigorous mathematical proof is given that networks with only three elementary species cannot have multiple steady states.

DNA replication in mammals and humans is qualitatively different from the well-understood replication process in simpler eukaryotes such as yeast. Reliable patterns exist in the organization of replication on the scale of chromosomes and chromosome segments, while the microscopic dynamics are known to be stochastic. A stochastic computer model is presented that incorporates the minimal set of model ingredients necessary to reproduce these dynamical properties. The ingredients are a fast-diffusing limiting factor, induced firing of origins depending on proximity to replication forks, a constant replication fork speed that is reduced during early S-phase and the initiation of replication in euchromatic DNA.

Results are consistent with experimental data and the literature, making the model presented here one of the best-benchmarked replication models in existence. A combination of model results with a three-dimensional DNA conformation obtained from a Monte Carlo model shows that chromatin packing is a main cause of the microscopy patterns observed during mammalian DNA replication.

The theoretical investigations presented in this thesis combine methods of physics and applied mathematics with problems from the field of cell-biology. Thus, due to this inherently interdisciplinary character, this thesis is of interest to a readership of both, physicists and biologists.

In memory of my grandparents
Brunhilde Kunkelmann
Johann Kunkelmann

Contents

1 Introduction

A system of interacting parts is called a *complex system* if its behavior is more than the sum of the behaviors of its individual parts [1]. Even though complex systems theory is a relatively young field of science, research into complex phenomena has a long tradition in physics[1]. Particularly the theoretical investigation of phase transitions and critical phenomena in the middle of the 20th century has had significant conceptual and methodological impact.

However, today the realm of complex systems theory extends far beyond the boundaries of any traditional field, making it a truly interdisciplinary science. Systems under investigation range from the microscopic world of cell-biology to the planet-scale domain of social networks and economics. Typically, the interacting parts and possible interactions in the larger complex systems are so numerous that they cannot be easily considered in their entirety. Theoretical investigations of such systems must thus aim to identify and analyze the essential components and interactions, simplifying as far as possible, but always preserving the functional underpinnings of the complex phenomena. This is especially important in the biological systems connected to the phenomenon of life, whose functional parts show an interconnectedness of extreme intricacy.

The field of cell-biology has seen rapid advances in the recent years, boosted by new experimental techniques such as protein labeling with green fluorescent protein (GFP). GFP and other fluorescent proteins give an unprecedented view of protein movement in live cells, making it possible to study previously experimentally inaccessible cellular processes in detail. However, while this new experimental quality allows the direct observation of intracellular kinetics, the important dynamical connections underlying the cellular functions are not trivially evident from the data.

Only through theoretical analysis is it possible to utilize the full information content of the data by identifying and characterizing the important actors and interactions. A constructive theoretical effort will not limit itself to the reproduction of individual data sets, but aims to connect separate experimental perspectives upon a cell-biological phenomenon into a coherent picture, i.e., a consistent model that gives new insight and from which testable predictions can be formulated. Furthermore, once the mathematical representation of a biophysical phenomenon has been found, analyzing the general properties of this representation often leads to new insights and helps to identify targets for additional experimental investigation.

The motivation of this dissertation is therefore to contribute to the on-going effort to uncover the inner workings of cell-biological systems. This is done by numerical modeling, supplemented by some analytical work, of the following patterns of cellular function in eukaryotes: The recruitment of proteins to damage sites in response to DNA double strand breaks, gen-

[1] Usage of the term “complex system”, however, started only in the late 20th century.

eral protein complex assembly dynamics, and the organization an temporal program of DNA replication. In the case of the DNA damage response and DNA replication investigations, the emphasis is firmly put on the particulars of these processes in mammalian cells.

DNA double strand breaks (DSBs), like other kinds of DNA damage, are caused by various types of chemical agents as well as radiation, and can be considered a "fact of life" for all cells. Eukaryotic cells, over the course of their evolution, have adapted to this constant threat by developing several mitigation strategies, ranging from damage repair to cellular suicide. A key process in eukaryotic damage response is the recruitment of response proteins that constitute a network of chemical interactions.

The first topic of this dissertation is an investigation of the dynamical properties of such a network, specifically the interaction network of DSB response protein recruitment. Experiments with heavy ion irradiation of human cells have shown that there is a qualitative change in the recruitment dynamics of the damage response protein NBS1 at high DSB densities[2]. While the recruitment speed increases with damage density in the lower range of DSB density, it stops to do so beyond a certain threshold. A minimal mass action computer model of the DSB response protein network surrounding NBS1 is presented, and results from that model show how the qualitative change in recruitment follows from the DSB density dependence of the different NBS1 binding interactions.

In the next part of this dissertation, the dynamics of protein complex formation is investigated. Protein complexes are groups of proteins that are chemically bound together and constitute an important element of cellular function. For instance, some protein complexes directly process the DNA sequence, such as the DNA replication fork and the various DNA single strand and double strand break repair machineries. Although it is well-known that completely assembled protein complexes participate in chemical processes which are promotive of complex dynamical phenomena, such as phosphorylation cascades (signaling) and protein/DNA modifications, it was so far not clear to what extent the protein accumulation dynamics itself could give rise to complex behavior. In order to answer this question, generic networks of accumulation and dissociation interactions between conserved protein populations are investigated. Using several approaches based on mass-action kinetics, the minimal prerequisites for multistability (and even oscillations) in an accumulation network are identified.

The final theme of this dissertation is the organization of DNA replication in human cells. A key difference between DNA replication in humans and DNA replication in unicellular eukaryotes such as yeast, where it is well understood, is that there are no pre-determined replication initialization sites in human DNA. Existing attempts at modeling DNA replication in humans have shown that the underlying processes are not based on trivially stochastic events, but instead must show some level of organization while remaining fundamentally stochastic. In these previous models several different processes were implemented, with a varying degree of success, and various constraints on the dynamics of replication in humans have been identified through experimental investigations.

In this dissertation, a minimal set of model ingredients is identified and the resulting stochastic one-dimensional model of DNA replication is presented. It is shown to be consistent with various experimental data sets by direct collaborators and from the literature. In addition, an

[2] As typical for heavy ion irradiation, where large numbers of DSBs are created.

attempt is made to relate the observed regularities of the model results to the three-dimensional patterns observed in microscopy by combining the one-dimensional model with a simulation of three-dimensional DNA conformation.

This dissertation is structured as follows: In chapter 2, an introduction to the biological systems and processes that are of relevance to the analysis presented here (such as DNA, proteins, the DSB response and DNA replication) is given. Chapter 3, presents the model for DSB response protein recruitment and contains a detailed discussion of the dynamics of the participating proteins. Next, in chapter 4, generic protein networks consisting only of accumulation and dissociation reactions are characterized and the prerequisites of multistability and oscillations in such networks are identified.

Then, the stochastic computer model of DNA replication is presented and benchmarked on experimental data in chapter 5. An additional Monte Carlo polymer model is used to generate a three-dimensional DNA conformation for comparison of the replication model with microscopy images. While the discussions of the results are contained within each thematically self-contained chapter, chapter 6 will give a summarizing conclusion and outlook.

Since a wide range of mathematical and computational approaches are involved in the investigation of cell-biological systems presented here, they are not all described together. Instead, a short overview of the basics of mass-action kinetics is given at the end of chapter 2 and all more model-specific methods, especially for the replication model, are described in the chapter where they are used.

Parts of the text of this dissertation have been or will be reworked into publications in peer-review journals. This is explicitly noted for each chapter where it is the case. Experimental data for parameters and model benchmarking was contributed by members of the lab of Prof. Cardoso at TU Darmstadt and the GSI biophysics group of Prof. Durante. Contributors are noted separately for each data set in place.

2 Cells, DNA and Proteins

Cells constitute the smallest unit of life as it is known today. The different types of cells (*cell species*) that are present in the biosphere of planet Earth number in the tens of millions [2]. Nonetheless, certain traits are shared by all of them: Cells consume free energy to maintain their existence and function, away from *thermodynamic equilibrium*. Every cell must have the ability to procreate by generating copies of itself. To this end, it must be capable of passing on information (its own "blueprint") hereditarily to its progeny. The information storage system of all known cells is a linear polymer called desoxyribonucleic acid (**DNA**), which can be copied through a process called *DNA replication*.

Most functional operations that a cellular system performs are catalyzed by polymers called *proteins*, which consist of *amino acid* chains. Information is copied from the long-term storage DNA to an intermediate information storage consisting of ribonucleic acid (**RNA**) by the process of *transcription*. RNA, likely an evolutionary predecessor of DNA [3], is chemically very similar to the latter[1]. Proteins are then created from the RNA in a process called *translation*. Both transcription and translation are conserved over all types of cells [2].

Lastly, all cells are bounded by a *plasma membrane*, which serves as a containment barrier to the cellular components and a transport mechanism for chemical agents. This membrane consists of molecules that are in part hydrophilic and in part hydrophobic, which causes them to spontaneously arrange in a closed double-layer. Special transport molecules are embedded into the membrane, so as to enable the transport of nutrients and waste products.

All cells share these traits, but many other features are only part of certain subgroups of cells. The most important division in this regard is that between *prokaryotes* (such as bacteria) and *eukaryotes* (such as yeasts, plant cells and animal cells). Eukaryotes differ from prokaryotes in multiple ways: They have a cell nucleus that contains the overwhelming majority[2] of their DNA, separated from the rest of the cell by its own membrane. Eukaryotes have a *cytoskeleton* (structural support system of protein filaments) and large numbers of internal structures called *organelles*, which fulfill various functions, ranging from energy transformation (*mitochondria*) and waste treatment (*lysosome*) to photosynthesis (*chloroplasts*). Only eukaryotes have the ability to form multi-cellular organisms, although not all eukaryotes are part of a multi-cellular organism. Typically, eukaryotic cells have a volume that is three orders of magnitude larger than that of prokaryotes and more DNA by approximately the same factor [2]. The layout of an example eukaryotic cell (an animal cell) is shown in figure 2.1.

Among the eukaryotes, several organisms hold the status of *model organisms*, meaning that cells of these organisms are studied in great detail by many research groups. Examples of widely used model organisms are *Saccharomyces cerevisiae* (budding yeast, single-cell

[1] RNA has one more hydroxyl group in the sugar backbone and its set of bases includes uracil instead of thymine.
[2] Excluding the DNA contained in some organelles, such as mitochondria and chloroplasts.

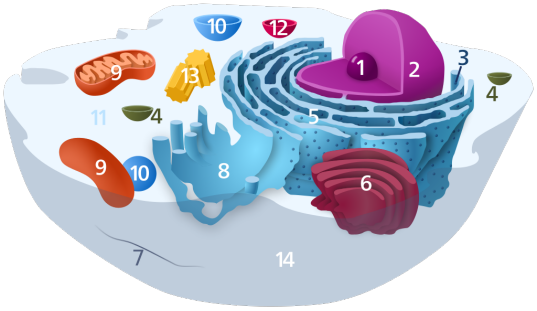

Figure 2.1: **Layout of an animal cell with organelles, as an example for a eukaryote.** The labeled organelles are (1): Nucleolus, (2): Nucleus, (3):Ribosome, (4): Vesicle, (5): Rough endoplasmic reticulum, (6): Golgi apparatus, (7): Cytoskeleton, (8): Smooth endoplasmic reticulum, (9): Mitochondrion, (10): Vacuole, (12) Lysosome, (13): Centriole. Cytosol, the liquid that fills the cell interior is labeled (11) and (14) is the cell membrane. Image source: Wikimedia Commons; License: CC0 1.0.

eukaryote), *Arabidopsis thaliana* (a plant), *Caenorhabditis elegans* (a worm, invertebrate), *Drosophilia melanogaster* (a fly, vertebrate), *Mus musculus* (a mouse, mammal) and the human. Since the model systems are all connected through their evolutionary history, insights gained from the investigation of one system often also apply to others. Some of the most basic (i.e., older) cell functions, such as transcription, function in the same manner in all cells, whereas many specific (i.e., younger) functions, only are similar in very closely related organisms. Investigations of a particular model systems thus not only increase the understanding regarding that system, but often also leads to more general insight on cell-biological processes, sometimes even with implications for medical therapy.

In the remainder of this chapter, those aspects of cellular function that are pertinent to the investigations presented in this dissertation, such as DNA structure, protein dynamics, DNA repair and DNA replication, are covered in more detail. At the end of the chapter, some relevant biochemistry-related experimental and theoretical methods are also discussed.

2.1 DNA

The basic building block of the DNA polymer is the *nucleotide*. Each nucleotide consists of a sugar molecule (desoxyribose) with one of four possible bases attached to it. Two of the four bases, *adenine* (**A**) and *guanine* (**G**) are of the larger *purine* type, while the other two, *cytosine* (**C**) and *thymine* (**T**) are *pyrimidines*. On one side of the sugar molecule (called the *5′ end*), a phosphate group is attached, while on the other side (called the *3′ end*), a hydroxyl binding site for a phosphate group exists. The sugars of multiple nucleotides are linked via this covalent binding "slot", thus forming a *DNA strand*. In its stable form, DNA appears not as a single strand, but as two intertwined strands. The two strands are connected at the bases via hydrogen bonds, with each A on one strand paired to a T on the other and each G paired

to a C. Two connected opposing bases are called a *base pair* (**bp**). It follows that the two strands exactly complement each other in their bases. The two *sugar backbones* (the chain of connected sugars) also run in reversed directions, meaning that the 5′ end of one strand always points in the same direction as the 3′ end of the other. A DNA double strand is wound around its axis, forming a *double helix* [4], at about one rotation every ten base pairs, which allows for more tightly bound bases [2]. Figure 2.2 illustrates some of the structure of the DNA.

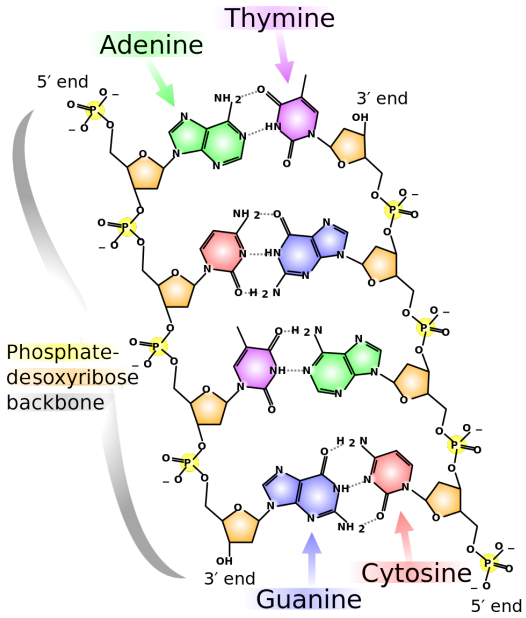

Figure 2.2: **Chemical structure of a DNA double strand segment.** Bases are connected to the sugar backbones left and right. The hydrogen bonds between the bases of different strands are dashed lines. Image source: Madeleine Prince Ball via Wikimedia Commons; License: CC0 1.0.

The DNA of eukaryotes is separated into multiple linear segments, the *chromosomes*. *Somatic cells* (i.e, non-germ cells) in humans contain two copies of their chromosomes, with the exception of the sex chromosomes X, one copy of which is replaced by a Y chromosome in cells of males. In sum, the human *genome* consists of 46 chromosomes with a total of 6.4 billion base pairs (**Gbp**).

Inside the nucleus, the DNA double strand is wrapped around protein complexes[3] called *nucleosomes* [5]. Each nucleosome consists of eight *histone* proteins and there is one nucleosome every 200 base pairs, with a short stretch ($\approx$ 50 bp) of connecting DNA between two adjacent nucleosomes. The conglomerate of DNA, nucleosomes and other chromosomal proteins is called *chromatin*. Nucleosomes with DNA wrapped around them are packed together in a cylindrical structure with a 30 nm diameter, forming the so-called *30 nm fiber* [2].

The 30 nm fiber is further organized into looped domains, which are attached to a scaffolding structure that consists of non-histone proteins [2, 6, 7]. The size of these looped domains

[3] A *complex* in this is a group of proteins that are bound together, see below.

is in the 50 kilobasepair (**Kbp**) to 200 Kbp range [2, 8] and they are fundamental units of chromatin organization [7].

There exists another organizational structure at a chromosomal level (regardless of packing, see below) which confines the chromosomes to separate volumes in the nucleus (*chromosomal territories*), thus keeping them from entangling with each other. While the details of this confinement remain unclear, various studies have shown that DNA looping on a multi-megabasepair scale is a likely mechanism, consistent with experimental data [9, 10].

During transcription, RNA copies of several kilobasepair long segments of DNA are produced by RNA polymerase enzymes. Some of the resulting RNA pieces then function as enzymes, while others (*mRNA*) are processed further (cutting out and splicing together of segments) and are then used as a template for the assembly of proteins in the ribosomes. A segment of the DNA whose pattern is expressed into a protein or a piece of enzymatic RNA, is called a *gene* [2, 11].

When proteins bind together, they are called a protein *complex*. Protein complexes fulfill many functions in the cell, and many proteins can participate in multiple complexes that have different functions [2].

Chemical modifications of the chromatin influence its function and form. These modifications, which are often maintained for prolonged times even after their initial cause has disappeared, include phosphorylation, methylation, and acetylation, among others. If a modification is resilient enough to be passed along to descendant (somatic) cells, it is called an *epigenetic* modification [12]. There are also many (mostly non-local) modifications to the chromatin that are inherently transient and are thus not epigenetic modifications. It has to be noted that a conflicting definition of the word epigenetic is also in widespread use, according to which all chromatin modifications that affect gene expression are considered epigenetic, regardless of their permanence[4] [13]. In recent years it has been shown that epigenetic modifications can also be propagated transgenerationally [14, 15].

One important chromatin property that is caused by epigenetic modifications is the packing state of the chromatin (affecting both DNA and the adjacent proteins). Several different packing states exist: Chromatin in the unpacked state is called *euchromatin*, whereas the various packed states are subsumed under the term *heterochromatin*. Transcription is only active in parts of the genome which are in the euchromatic state and genes become silenced when euchromatin is condensed into heterochromatin [2]. The packing of chromatin is conserved through DNA replication [16].

2.2 Replication

The reproductive process of eukaryotic cells is organized in a temporal program that is called the *cell-cycle*. In order for a cell to be able to procreate by dividing in two, it must duplicate itself completely within one iteration of the cell cycle. In this duplication, the different phases of the cycle play different roles. The entire DNA of the cell is copied via DNA replication

[4] Whenever the word epigenetic is used here, it is used as defined by the first definition, unless explicitly stated otherwise.

during the *S phase*, which takes about 10-12 hours in a mammalian cell. S phase is flanked by phases G1 and G2, which constitute additional time for the cell to grow by duplicating its organelles and producing proteins. The actual division of the cell takes place in M phase, when the chromosome pairs and the cytoplasm are separated in two. Because both DNA replication and cell division must be completed once they have been started, special precautions exist to ensure the viability of completion before their initiation. At the *cell cycle checkpoints*, progression of the cell cycle can be stalled, if conditions for the initiation of these critical processes are unfavorable. For a human cell, the entire cell cycle takes about 24 hours [2].

DNA replication starts with the assembly of pre-replicative complexes on the DNA, which determine the *origins* of replication, that is, the positions where replication is initiated [17]. Once initiation takes place (an origin "*fires*"), the double strand of the DNA is opened up locally and two sets of replication machinery[5] are assembled at the double strand to single strands boundaries. These machineries, called *replication forks*, then move along the original DNA strand and copy the DNA, until they annihilate with a fork coming from the opposite direction or reach a chromosome end or other barrier. A replication fork uses each of the strands of the original DNA as a template, upon which its *DNA polymerase* synthesizes a second strand (see figure 2.3). Thus, both of the resulting two double strands contain one strand of the original DNA and one newly synthesized strand [2]. A region of the DNA that is duplicated by a pair of replication forks from the same origin is termed a *replicon* [18–20]. In human cells, the number of simultaneously active replicons is 6000-7000 [21].

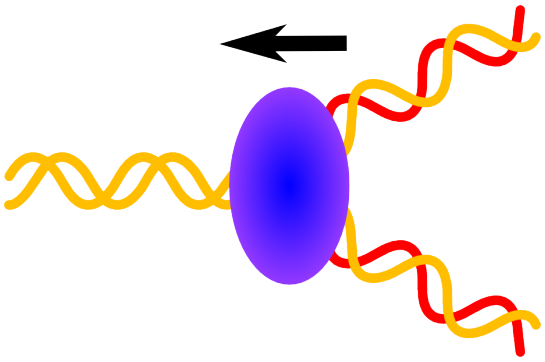

Figure 2.3: **Schematic depiction of replication fork progression.** The fork (blue) moves along the original double strand (yellow) in the direction of the arrow. Both strands of the original DNA are incorporated in one of the resulting two double strands, each of which thus contains one old (yellow) and one new (red) strand.

Even though some features, such as the overall origin initiation timing [22], appear to be very similar for all eukaryotes, differences exists between replication in metazoans and other eukaryotes: The positions and timing of origins in metazoan DNA is influenced by such chromatin features as GC content [23, 24], the presence of specific G-rich sequences [25], histone modifications [26, 27], and "transcription potential" [20, 23, 28–32]. There are, however, no

[5] "Replication machinery" is a commonly used term for the protein complex that performs the actual DNA replication.

specific DNA sequences that pre-determine the origin positions in metazoans [33–35]. This is in contrast to the situation for the model eukaryote S. cerevisiae, where the origin positions and relative firing efficiencies are mapped [36–38]. It has to be noted that the timing of individual origin firing in eukaryotes in general is known not to follow a sequential program, indicating a certain stochasticity of the underlying processes [39–42].

The chromatin-scale timing of replication, on the other hand, is largely conserved for each cell line. Some segments of the genome are reliably early-replicating (*R-bands*), some are reliably late-replicating (*G-bands*), while others are intermediates [41,43]. Euchromatic regions are, on average, replicated early due to their open structure, while heterochromatic regions are replicated late [44, 45]. Replication in metazoan cells also shows changes in its spatial pattern throughout S-phase. In early S-phase (about the first 3 hours), replication foci are small and distributed through the entire nucleus. In middle S phase, replication foci start to group together and align on the nuclear and nucleolar envelopes. In late S-phase, only very few replication foci remain, but these are extremely large. Images of these patterns can be found in the literature [43,46] or below in section 5.11.

Molecular analyses have shown that the positions of active origins, inter-origin distances and the rates of replication fork movement can be very heterogeneous even within individual cells [47–49]. This indicates that the replication program in metazoans has a certain plasticity, thus providing the necessary flexibility to ensure complete genome duplication in the face of developmental and environmental changes [50]. For example, the duration of S-phase can vary between cells of the same organism in dependence on developmental state and influenced by metabolic stress or DNA damage [50–52]. Evidence suggests that the flexibility of replication includes allowing some residual replication to occur in G2 phase [53].

2.3 DNA Damage and Damage Response

In its natural environment, DNA is exposed to a multitude of stresses. These include chemical agents that are always present in the cell but also external influences, such as ultraviolet (*UV*) radiation, human-made chemical compounds and ionizing radiation. Because these stresses cause damage to the DNA, they threaten the viability of cells and even entire organisms[6]. Since this threat has always been present throughout evolution, cells, especially eukaryotic cells, have mechanisms that provide them with a certain resilience to DNA damage [54].

One such mechanism is cell cycle arrest at the G1 cell cycle checkpoint, already mentioned in section 2.2, which allows the cell to avoid initiating DNA replication (and thus avoid exacerbating the situation) if there is too much DNA damage. Another option is apoptosis, a programmed cell death that has many functions, one of which is to act as a mechanism to weed out damaged, but possibly viable, and thus potentially cancer-precursory cells [2,54].

However, there are also a multitude of response pathways that actively counteract the DNA damage by correcting it or by at least restoring the DNA structure. These responses are labeled *DNA repair*, and are grouped according to the properties of the damage that provokes them. If there is damage to the bases of the DNA or to its backbone on one strand, then single strand

[6] Somatic cells with multiple mutations can become cancerous.

break (**SSB**) repair pathways are invoked. These mechanisms remove a base or a part of a DNA strand (*excision*) in order to replace it with a valid segment, usually using the other strand as a template. If the nature of the damage is such that the double strand is deformed ("bulky" damage), then nucleotide excision repair (**NER**) is used [54]. If there is no double strand deformation, the primarily used process is base excision repair (**BER**). BER has two pathways, a short patch pathway in which one nucleotide is replaced, and a long patch pathway, in which 2-13 nucleotides are replaced [55]. Other pathways exist for special cases of damage, such as mismatch repair (**MMR**).

For damage that ruptures the double strand (so called double strand breaks, **DSB**s), special repair pathways exist. *Homologous recombination* uses an existing template of the damaged region to close both strand lesions through DNA synthesis. Even though, in principle, any homologous chromosome could be used in metazoan cells [54], most cells only employ this pathway after DNA replication, when a sister chromatid is available or the chromosome has split into two daughter chromosomes [2].

Non-homologous end-joining (**NHEJ**), on the other hand, rejoins DSBs without need for an intact copy of the DNA sequence. Since NHEJ consists primarily of end processing and ligation, DNA segments can be lost in the process, meaning that the DNA sequence is changed. NHEJ is the prevalent repair mechanism in human cells [56] and has several alternative pathways that maintain its function even for knockout or knockdown of key NHEJ proteins [57]. Finally, there are additional pathways that use homologies between strand ends to fuse them together, such as single-strand annealing [54] and microhomology-mediated end joining [58].

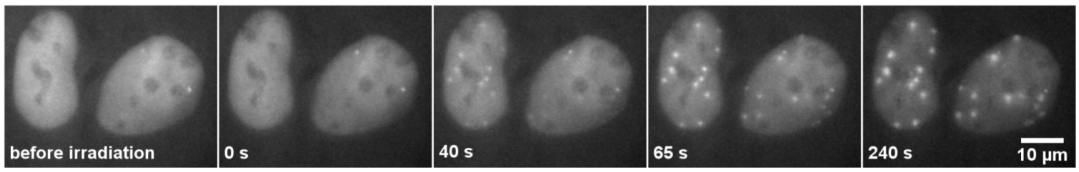

Figure 2.4: **Example image series of live cell protein recruitment fluorescence microscopy.** U2OS cells expressing GFP-tagged NBS1 (a double strand break response protein, see section 3) after DNA damage induced by heavy ions. Before irradiation, the fluorescent proteins are distributed throughout the nucleus, making it visible under the microscope. After irradiation, the response proteins accumulate at damage sites, forming bright foci. Image courtesy of Frank Tobias and Burkhard Jakob.

All DNA damage response pathways share that they are mediated through proteins. This includes cell cycle arrest and apoptosis, which are initiated via recognition (i.e., binding) of the damage by early damage response proteins and subsequent signaling cascades, consisting of modification (such as phosphorylation) of proteins and chromatin. The cell fate decision making process functions through feedback loops in these protein signaling networks, which enable the cell to show switching behavior [54].

A shared trait of the repair pathways is that they require the assembly of proteins into repair "machinery" at the damage site. Over time, free diffusing instances of the required repair proteins are *recruited* to the DNA damage, i.e., they bind to the damage site or to the partially assembled machinery. Figure 2.4 shows microscopy images of this recruitment process. The

binding of proteins at the damage site is a dynamical process, proteins dissociate and are recruited continuously. Correct function of protein complexes such as the repair machinery is ensured by the binding kinetics of the involved proteins: An incorrectly assembled complex (or partial complex) is more likely to dissociate than a correctly assembled one [59] (*kinetic proofreading*).

2.4 Biological Diagnostics

Fluorescent protein labeling is an essential diagnostic tool of cell-biology. Unlike traditional non-fluorescent dyes that require the fixation (i.e., killing) of cells for optical microscopy, fluorescent markers allow the visualization of molecule concentrations (for instance, of proteins) in both fixed cell and live cell microscopy. Typical methods of fluorescent microscopy in fixed cells are staining with directly target-binding dyes, such as the DNA marker *DAPI*, and the use of dyes attached to (sometimes indirectly) target-binding antibodies [2].

The large-scale use of fluorescence microscopy of living cells has been made possible by the isolation of green fluorescent protein (**GFP**) from *Aequorea* jellyfish [60, 61] and the subsequent sequencing and cloning of its gene [62]. GFP, the first of the various fluorescent proteins in use now, is a protein sequence that performs all necessary modifications to attain fluorescence capability itself [63,64]. This significant property makes it possible that GFP and other, similar fluorescent markers can function wherever their gene is expressed, independent of the host DNA or host organism.

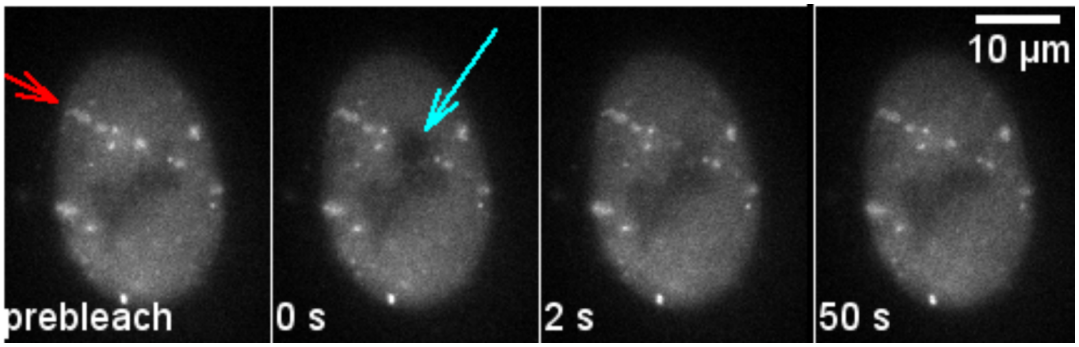

Figure 2.5: **Microscopy images of a typical FRAP measurement.** Before bleaching, the fluorescently tagged damage response protein NBS1 has been recruited to DNA damage sites which are situated along the track of a heavy ion (red arrow). At 0 s, a laser pulse photobleaches the NBS1-GFP proteins in a small region around a damage focus. In the following minutes, the fluorescence signal in the bleached region recovers, as bleached proteins dissociate and fluorescent proteins that have diffused into the bleached region are recruited. Image courtesy of Frank Tobias and Burkhard Jakob.

One important use of fluorescent proteins is the tagging and observation of proteins in living cells. If the DNA of a gene is modified in such a way that the gene will be expressed with a

fluorescent protein attached, and if it can be shown that the function of the resulting protein is retained in spite of the modification, then the dynamics of said protein can be observed through microscopy in vivo. An alternate method is to introduce a plasmid that expresses the GFP-tagged protein, thus flooding the cell with the modified version of the protein, in addition to the unmodified proteins produced by the cell.

Fluorescent recovery after photobleaching (**FRAP**) is a useful method for the quantification of protein reaction and diffusion kinetics. In a typical DNA damage focus FRAP experiment, a clearly visible focus of fluorescent proteins is selected and then photobleached with a laser. The intensity of the photobleached region is recorded and shows a recovering fluorescence signal that stabilizes slightly below the pre-bleach value. The reason for this signal recovery is that, over time, the photobleached proteins in the focus are replaced by non-photobleached proteins via the association and dissociation that continuously happens in the focus (see figure 2.5).

Time scale and exact shape of the recovery curve depend on the binding and dissociation rates in the focus, as well as the diffusion coefficient of the molecule in question. There are limit cases in which one process dominates the curve shape, simplifying the extraction of parameters from FRAP data[7]. In the general case, the diffusion and binding constants, due to their positive correlation [65], can often not be calculated directly from the FRAP data, making additional measurements necessary. However, it is usually possible to directly extract reliable values for the dissociation rate constant. It has to be noted that secondary transient binding of the free protein to partners in the nucleoplasma can complicate the parameter calculation [66].

2.5 Protein Dynamics

Mass action kinetics is a standard approach for the mathematical modeling of chemical reactions that is also applicable in the case of protein interactions in the cell nucleus. If a chemical reaction is elementary, that is, if it consists only of a single reaction step, then the probability of two individual reactant molecules A and B undergoing that chemical reaction together in a defined volume during some time interval has a certain value $\tilde{k}_{A+B}$. If more than one instance of each molecule is present, the total rate of the reaction can be determined by multiplying this value with the population sizes $\tilde{A}$ and $\tilde{B}$ of the molecules: $R_{total} = \tilde{k}_{A+B} \cdot \tilde{A} \cdot \tilde{B}$. It is customary to rescale $\tilde{k}_{A+B}$ by the size of the volume, thus allowing to define a total rate r_{total} in terms of species concentrations A and B: $r_{\text{total}} = k_{A+B} \cdot A \cdot B$. The rescaled constant k_{A+B} is called the reaction *rate constant*. For reactions with more than two reactants the other concentrations are multiplied likewise, as is the case with reactants of which multiple instances participate in a reaction (thus increasing the exponent of the concentration by one).

Since the total reaction rate is the rate at which reactant concentrations are reduced and reaction product concentrations are increased, the dynamics of the participating reactant concentrations can be described by a set of ordinary differential equations $\frac{\mathrm{d}\vec{c}}{\mathrm{d}t} = F(\vec{c})$ in which the concentrations (elements of concentration vector $\vec{c}$) and the rate constant appear. As an exam-

[7] The most prominent example of this is the case of dissociation-limited FRAP dynamics, where the signal recovery follows an exponential curve.

ple, for the reaction of A and B to form a complex AB, written $A + B \rightarrow AB$, these equations are:

$$\frac{\mathrm{d}A}{\mathrm{d}t} = -k_{A+B}A \cdot B, \tag{2.1}$$

$$\frac{\mathrm{d}B}{\mathrm{d}t} = -k_{A+B}A \cdot B, \tag{2.2}$$

$$\frac{\mathrm{d}AB}{\mathrm{d}t} = k_{A+B}A \cdot B. \tag{2.3}$$

$$\tag{2.4}$$

Additional reactions produce additional terms in the differential equations. If a reaction system also contains the inverse of a reaction, in this case $AB \rightarrow A + B$, then it is called *reversible*. The sets of species that participate in or result from a chemical reaction are called *complexes*. To avoid confusion due to the overlap in nomenclature with protein complexes, reaction complexes are only referred to as complexes in chapter 4.

The reversible reaction $A + B \rightleftharpoons AB$ is one of the most simple of the many different reaction types that are possible in mass action kinetics. In the context of cellular protein dynamics, such reactions are part of the binding and dissociation processes of protein complexes. Other very important protein reactions are enzymatic reactions, which are composed of many elementary substeps, and auto-modification reactions, whose mass action representation contains higher order concentration terms. An important condition for the validity of mass action kinetics in protein interaction systems is that the diffusion time scale for all reactants should be faster than the reaction kinetics time scale. The behavior of protein interactions that are diffusion-limited does not conform to mass action kinetics.

In the remainder of this section, dynamical phenomena of interest that can appear in cell-biological protein interaction systems will be discussed, in order to lay the foundations for the more in-depth discussion in chapter 4. It is apparent from reaction system (2.1-2.3) that redundancies exist in the formulation of the mass action differential equations. In the calculation of section 4.3, chemical reaction network theory (**CRNT**), a method that exploits these redundancies to simplify the analysis of reaction dynamics, will be used.

In a cell-biological protein reaction system, an additional restriction applies concerning the protein concentrations: If the protein interaction dynamics is faster than the time scale of gene expression (which is on the order of hours [67]), then the total amount of each protein in the system can be considered constant. This is the case for both, the generic systems under consideration in chapter 4, and the double strand break response dynamics of chapter 3. As a consequence of this restriction, the solution to the differential equations governing such a protein interaction network can never diverge, since all concentrations must be positive and can at most reach their respective maximum value.

Chemical mass action systems are dissipative, meaning that the phase space volume filled by the entirety of the system's trajectories, starting from every possible set of concentrations, shrinks as time advances. In the long run, all trajectories will remain confined to a number of closed subsets of the phase space. These subsets are called *attractors* [68]. Examples of attractors are fixed points, at which all concentrations remain constant, or limit cycles, where

concentrations oscillate indefinitely[8]. Fixed points, for which $\frac{\mathrm{d}\vec{c}}{\mathrm{d}t} = 0$, are said to be stable, if they attract trajectories from some open set in phase space surrounding them, and unstable if they repel trajectories. Stable fixed points can be identified with steady states of a system and systems with more than one steady state are called *multistable*. Unstable fixed points of chemical systems, although they appear for the mass action ODEs, are not realized in nature due to the stochasticity of the underlying processes.

Multistability allows cell-biological systems to implement conditional switching, wherein changes in the concentrations of signal proteins or other chemicals cause the transition from one persistent steady state to another. This functional principle is at the heart of decision making processes in cells, including differentiation of stem cells [69–71], and the function of specialized cells in a multicellular organism [72, 73]. Other examples from the more immediate context of this dissertation include the decision making on cell cycle progression [74, 75] and the selection of DNA damage response pathways [54, 76]. Stable limit cycles are also found in many cell-biological systems. Prominent examples include the oscillation of glycolysis [77, 78] and circadian rhythms [79].

It is important to note that feedback loops are a prerequisite for multistability and oscillations to appear in cellular systems [80–83]. Positive feedback can provide the nonlinear component that is needed to allow a chemical reaction system to have multiple steady states [80, 84], while time-delayed negative feedback does so for oscillations [81]. Chapter 4 will elaborate more on the role of feedback loops.

In line with the importance of complex phenomena for biological function, many mechanisms exist in the chemical interactions of cell-biology that promote the occurrence of such phenomena. Modifications of proteins, such as phosphorylation, allow protein networks to show efficient signal propagation and bistability [85], with the additional benefit that the capacity of some proteins to auto-phosphorylate directly enables multistability [86]. Epigenetic modifications of the DNA are used to regulate the expression of genes, thus enabling fine-grained control over the structure and dynamics of gene networks [87].

From the modeling perspective, it is always possible to make a large reaction system behave like any of its smaller subsystems by choosing near-zero rate constants for some of its reactions[9]. Even if the large reaction system has multiple steady states or oscillations, there are always subsystems that converge to a single steady state. This means that for the larger system there must be some sets of rate constants for which its attractive behavior transitions to a single steady state. In nonlinear dynamics theory, parameter sets at which qualitative changes in the dynamics of a system appear are called *bifurcations* [88]. By the above reasoning, it must be possible to find bifurcations in any cell-biological protein interaction system that shows complex dynamics.

The two most important types of bifurcations for the analysis performed here are the saddle-node bifurcation and the Hopf bifurcation. At a saddle-node bifurcation, a stable and an unstable fixed point are created/destroyed. If there is already a stable fixed point in the system, as is often the case in the closed chemical reaction systems analyzed in this thesis, fixed point

[8] Strange attractors can also appear in chemical systems, but are not pertinent to the results presented here.

[9] Trivially, a system becomes one of its subsystems if all rate constants for reactions that do not belong to the subsystem are zero.

creation by a saddle-node bifurcation means that the system goes from one steady state to two steady states (with an unstable fixed point somewhere in between). The saddle-node bifurcation has been shown to occur in many cell-biological switching systems, especially in development [71, 89].

At a Hopf bifurcation, a stable fixed point (spiral) transitions to an unstable fixed point and a stable limit cycle, allowing persistent concentration oscillation in the system. Hopf bifurcations also appear often in cell-biological systems. They for instance play an important role in the question of robustness in the yeast glycolysis oscillation system [78].

3 Modeling the Kinetics of Protein Recruitment After Double Strand Breaks

In this chapter, a protein recruitment model of the early damage response after DNA double strand damage is presented. A part of the text in this chapter was published in a modified form as the theory section of [90]. All experimental data shown here was generated by Frank Tobias and Burkhard Jakob in the group of Gisela Taucher-Scholz at the GSI Helmholtzzentrum für Schwerionenforschung (**GSI**), under the supervision of Marco Durante. Nicor Lengert, whose bachelor thesis on this topic was overseen by the author, contributed to the model design.

When a DNA double strand break occurs, response proteins are rapidly recruited to the damage site. What follows is a cascade of protein modification, protein recruitment and chromatin modification. The proteins that participate can be divided into proteins that are part of a specific response pathway, such as the NHEJ proteins Ku/DNA-Pk, and pathway-independent response proteins.

The early recognizing *MRN* complex is part of this pathway-independent response [91, 92]. It consists of the proteins MRE11, RAD50 (whose zinc hooks help to keep DSB ends close together [93]) and NBS1. MRN interacts with the protein kinase ATM [94–98], whose phosphorylation targets include the histone H2AX in the break vicinity [99–102] and itself (autophosphorylation). The phosphorylated variant of H2AX, γH2AX, in turn interacts with MRN via the protein MDC1 which then recruits more ATM, leading to further signal amplification [98, 103–107]. While the protein interactions and recruitment order were already uncovered in recent years, the goal of the investigation that is presented here is to characterize the dynamic behavior of the pathway-independent DSB response.

Although many different chemical and radiation-based methods exist to generate DNA damage, heavy ions have the advantage that their dose deposition is highly localized. They thus produce dense clusters of DNA damage sites, among which there are many double strand breaks. A good measure for the damage caused by heavy ion irradiation is the linear energy transfer (**LET**), which is defined as the energy loss of a particle of ionizing radiation per distance traveled through matter. LET has the unit keV/µm and typical values for irradiation with heavy ions are in the 100 keV/µm-15000 keV/µm range.

3.1 LET-Dependence of Recruitment Speed

Before introducing the model, in this section a summary of the results of the accompanying experiments by the GSI group is presented. Human osteosarcoma cells of U2OS cell lines that

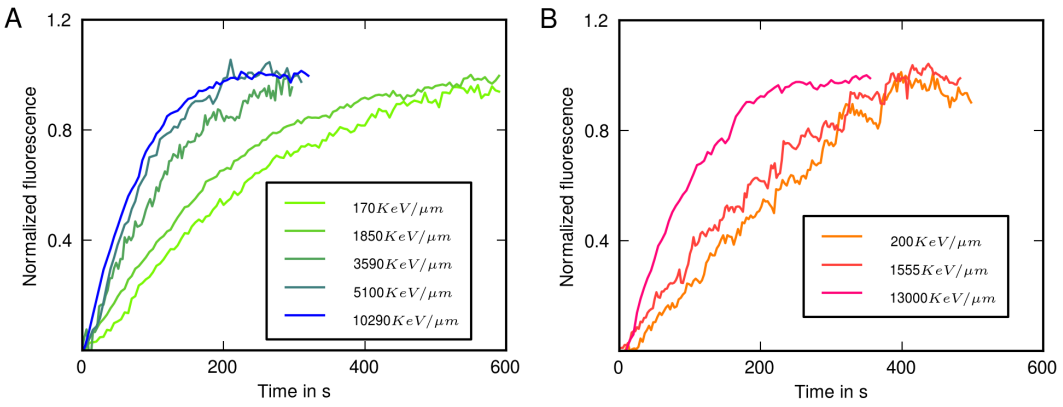

Figure 3.1: **NBS1 recruitment signal after irradiation, in dependence of LET. A** Integrated focus GFP signal after heavy ion irradiation for cells expressing NBS1-GFP, over time, with the end value normalized to one. The recruitment curve of NBS1 becomes steeper with LET, but remains at seemingly constant steepness above 5000 keV/µm. In total, twelve data sets were generated, of which five are shown here. **B** Similar measurements for MDC1. Three data sets were generated using U2OS cells expressing MDC1-GFP. Data courtesy of Frank Tobias and Burkhard Jakob.

express either NBS1-GFP, MDC1-GFP or 53BP1-GFP [108–110] were irradiated with heavy ions at the linear accelerator of GSI and recorded using the beam line microscope of the GSI group [111, 112].

As a result, recruitment curves for NBS1, MDC1 and 53BP1 for different LETs were obtained. From the NBS1 data, a pattern emerged for the LET-dependence of NBS1 recruitment: In the lower LET range (100-2000 keV/µm), the recruitment becomes faster with increasing LET, whereas in the high LET range (above 5000 keV/µm), the recruitment speed remains the same, independent of LET. It is not obvious from the experiment alone what the reason for this qualitative difference is. Figure 3.1 **A** shows a selection of the NBS1 data sets. FRAP results for RAD50 and MRE11 showed that these proteins dissociate on the same time scale as NBS1 (see figure 3.2), indicating that these MRN components dissociate as a complex.

Figure 3.1 **B** shows the recruitment curves for MDC1,which has been shown to be a loading platform for further MRN recruitment in the DSB vicinity, recorded using the MDC1-GFP expressing cell line. As expected from their known interactions, the time scale of recruitment is similar for MDC1 and MRN. An increase in recruitment speed with LET could also be demonstrated in the MDC1 data, although the high-LET behavior could not be observed in detail due to the limited number of data sets. A noticeable difference exists between the MDC1 data and the NBS1 data: While the slope of the NBS1 recruitment curves decreases steadily on the multi-second timescale, the low-LET MDC1 curves have a seemingly constant slope for the first few minutes. This indicates that the connection between the recruitment dynamics of the two proteins is not trivial.

In order to check, whether or not a slope increase with LET is a general feature of the recruitment of DSB damage response proteins, 53BP1 data was also analyzed in a similar

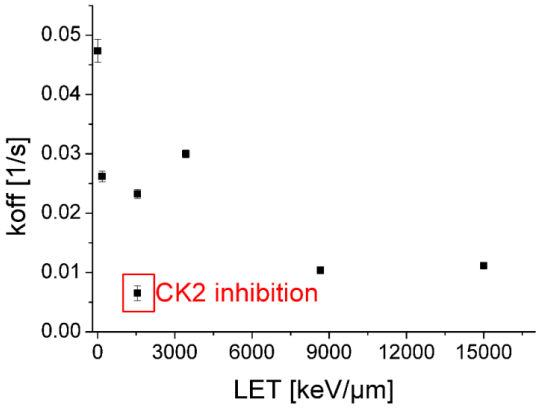

Figure 3.2: **NBS1 dissociation rate constant values over LET.** Values of k_{off}, determined using the FRAP model by Sprague et al [66] decrease as LET increases. In the highlighted experiment, CK2 inhibition was used to impede MRN from binding to MDC1. Notably, the k_{off} value thus achieved is in the same range as the values measured in high LET experiments without inhibition. Figure courtesy of Frank Tobias and Burkhard Jakob.

manner. These measurements showed that 53BP1 is recruited on a much slower time scale and after a slight delay, without any speed-up for higher LET. Data on the recruitment dynamics of ATM at an LET of 14350 keV/µm from a previous study [102] was also available for use in the analysis.

Since a modeling effort of protein recruitment necessarily involves several system parameters, i.e., rate constants in the case of a mass action model, additional FRAP measurements of the proteins NBS1 and MDC1 were made to reduce parametric uncertainty. In these measurements, radiation-induced DSB foci were bleached, so that the resulting recovery curve reflects the binding in the DSB response focus. From the FRAP data, the effective diffusion coefficient D_{eff}, the dissociation rate constant k_{off} and the effective association rate constant k^*_{on} were determined using a radial diffusion-reaction model by Sprague et al [66].

The effective diffusion coefficient can be used to calculate time scales for the movement of the proteins in the nucleus. If the average radius of the nuclei is 9.4 µm, then, under the assumption of approximately cylindrical geometry, the average distance to the nucleus center from each point in the nucleus is 6.3 µm. Using the mean displacement relation for diffusion in three dimensions, $\Delta x^2 = 6D_{\text{eff}}t$, the time scale for diffusive protein exchange in the nucleus, and thus also exchange of proteins in the focus vicinity, can be calculated. It is 26 s (with a measured $D_{\text{eff}} = 0.25\,\mu\text{m}^2/\text{s}$) for NBS1 and 228 s (with a measured $D_{\text{eff}} = 0.029\,\mu\text{m}^2/\text{s}$) for MDC1. The movement of MDC1 in the nucleus is thus significantly slower than that of NBS1. In additional FRAP measurements of MDC1 outside the focus, it was shown that the general mobility of MDC1 is further reduced for high LETs.

The effective association rate constant is the product of the actual association rate constant and the free binding site density. Since these two quantities cannot be separated by use of the data alone, it is not possible to derive an association rate constant for direct use in the model.

It has to be noted that the dissociation rate constants derived from these measurements,

while suitable for modeling use, became smaller with increasing LET (see figure 3.2). To investigate, whether this LET-dependent change depended on a particular binding mode of MRN, additional FRAP experiments with inhibition of casein kinase 2 (CK2) were conducted. CK2 phosphorylates MDC1, an interaction without which NBS1 cannot bind to MDC1. Experiments were made at low (1550 keV/µm) and high (15000 keV/µm) LETs and showed that the NBS1 dissociation constant is the same for both LET values if binding to MDC1 is inhibited. Even though the size of NBS1 foci was reduced significantly, micro-foci, as expected from the work of Bekker-Jensen et al [113], remained. Furthermore, the dissociation constant of the inhibited case had a value very close to that of high-LET measurements in the uninhibited case. A detailed description of the recruitment experiments, FRAP experiments and FRAP data analysis can be found in the PhD thesis of Frank Tobias [114].

In the remainder of this chapter it will be laid out how the qualitative differences in the recruitment behavior of MRN (and MDC1) are a result of a DSB-dependent shift between known recruitment processes. A minimal computer model will be described with which it was possible to reproduce the experimental data and identify the central protein interactions of this early damage response pattern.

3.2 Model

The computer model is based on a minimal subset of the known interactions between damage response proteins and the DNA, an overview of which is given in figure 3.3. From the protein interactions network, a system of ordinary differential equations (**ODE**) is generated that is then solved numerically. The results of such model calculations represent the dynamics of protein concentrations in the fixed volume around damage foci. Because NBS1, RAD50 and MRE11 participate in the DSB response together as MRN complex [95, 106], only the MRN complex was included in the dynamics. Experimental data from section 3.1 showed that NBS1, RAD50 and MRE11 have similar dissociation kinetics, which is essential for the validity of this approach.

The central hypothesis of the recruitment model is that there are two qualitatively distinct processes by which MRN binds in the DSB focus. One of process is the direct binding of MRN to the double strand break ends, thereby forming an "inner focus" whose size is dependent on the number of DSBs. In the other process, MRN binds to MDC1 at γH2AX sites in the wider DSB vicinity, which leads to the formation of an "outer focus". The number of binding sites in this latter interaction, and thus the size of the outer focus, are independent of the number of DSBs. This qualitative difference in the DSB/LET-dependence will later prove crucial for explaining the LET dependence of the recruitment dynamics.

The first reaction in the interaction network of the model is the reversible binding of MRN directly to the double strand break (inner focus binding). The model does not contain the most immediate processing of the various chemically distinct DSB subtypes, but assumes that DSBs are ready for binding with MRN.

MRN bound to the DSB can catalyze the activation of ATM through auto-phosphorylation [100]. In the context of the protein interaction network of the model, this means that free ATM binds to inner focus MRN and subsequently dissociates as activated ATM. Active ATM will

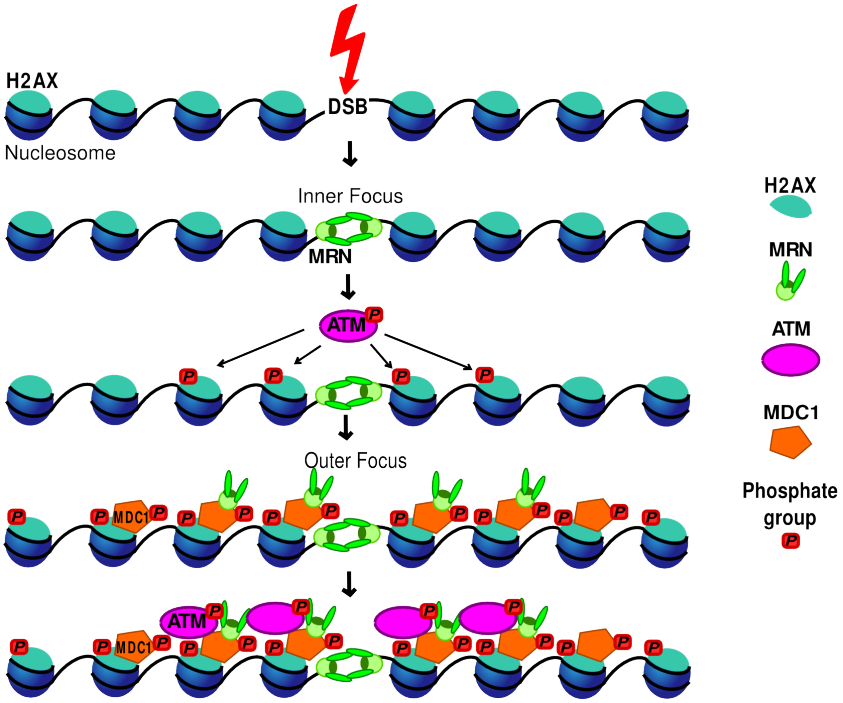

Figure 3.3: **Schematic of interactions in the minimal model.** MRN binds directly to the DSB strand ends. ATM is activated there and subsequently phosphorylates H2AX. MDC1 must be recruited to γH2AX before MRN can bind there. In a final step, ATM also binds to recruited MDC1. For better presentability, only the nucleosomes that contain H2AX are depicted. Graphic courtesy of Nicor Lengert.

then phosphorylate H2AX to γH2AX.

It is known that MDC1 binds directly to γH2AX [104] and that MRN recruitment in the larger vicinity of the DSB is MDC1-dependent [106, 109, 115, 116]. In the model, MRN can thus only be recruited to the DSB surrounding chromatin by binding to already recruited MDC1, which reversibly binds to γH2AX (outer focus binding). It is known that ATM is retained at DSBs through interaction with MDC1 and that phosphorylation of ATM plays an important role in this [100]. It was found that the best model results are obtained when phosphorylated ATM is allowed to bind to recruited MDC1, independently of outer focus MRN binding.

The inclusion of dephosphorylation reactions for γH2AX and active ATM as well as dissociation reactions for ATM at the outer focus did not change the simulation results significantly. Because the main interest of this investigation is to find the essential interactions for a minimal model, these reactions, even though present in nature, were left out for the final calculations.

For some parameters of the model, experimentally obtained values are used as input. The dissociation rates for inner and outer focus MRN as well as MDC1 are extracted from the results of FRAP measurements made by the GSI group. Under the assumption that there is an inner and an outer focus, the dissociation rate of MRN measured through FRAP always reflects a mixture of the inner and outer focus dissociation processes. However, there are limiting

cases, where one of the two binding processes can be characterized without interference from the other.

Since the number of binding sites in the inner focus is proportional to the number of DSBs, for very low LETs it becomes negligible compared to the DSB-independent number of binding sites in the outer focus. An NBS1 dissociation value obtained through X-ray irradiation, which corresponds to an LET of approx 1 keV/µm, was therefore used as the outer focus MRN dissociation rate: $k_{\text{off,o}} = 0.047$ 1/s. Inhibition of kinase CK2 removes the interaction between NBS1 and MDC1 [116]. In the context of the model presented here, this means that binding of MRN at the outer focus is disabled, leaving only inner focus binding. The dissociation rate for NBS1 obtained in FRAP measurements with CK2 inhibition was thus used as the constant for inner focus MRN dissociation: $k_{\text{off,i}} = 0.007$ 1/s.

For the MDC1 dissociation rate, the value $k_{\text{off,MDC1}} = 0.00425$ 1/s that was also determined in FRAP measurements at GSI, was used. It was assumed that the number of DSBs scales linearly with LET [117]. The linear parameter for the number of DSBs was determined to be 28 DSBs at an LET of 170 keV/µm, using the ion fluence ($3 \cdot 10^6$ cm^{-2} in the experiments here), the LET, and the assumption that there are on average 35 DSBs per Gray [118] .

Since H2AX on average appears in every fifth nucleosome, there are approximately 1000 H2AX per Mbp of chromatin. Only H2AX in the focus should be considered in the model, so that the absolute number of H2AX (relative to the number of DSBs) that is returned by the parameter optimization must be in the low thousands. This is the case for all results presented here. The full set of reactions in the computer model is:

$$\text{MRN} + \text{DSB}_s \rightleftharpoons \text{MRN}_i, \tag{3.1}$$
$$\text{ATM} + \text{MRN}_i \rightarrow \text{AMRN}_i, \tag{3.2}$$
$$\text{AMRN}_i \rightarrow \text{ATM}_p + \text{MRN}_i, \tag{3.3}$$
$$\text{H2AX} + \text{ATM}_p \rightarrow \gamma\text{H2AX} + \text{ATM}_p, \tag{3.4}$$
$$\text{MDC1} + \gamma\text{H2AX} \rightleftharpoons \text{M}\gamma\text{H2AX}, \tag{3.5}$$
$$\text{MRN} + \text{M}\gamma\text{H2AX} \rightleftharpoons \text{MM}\gamma\text{H2AX}, \tag{3.6}$$
$$\text{M}\gamma\text{H2AX} + \text{ATM}_p \rightarrow \text{AM}\gamma\text{H2AX}, \tag{3.7}$$
$$\text{MM}\gamma\text{H2AX} + \text{ATM}_p \rightarrow \text{AMM}\gamma\text{H2AX}, \tag{3.8}$$
$$\text{MRN} + \text{AM}\gamma\text{H2AX} \rightleftharpoons \text{AMM}\gamma\text{H2AX}, \tag{3.9}$$

where the compound species names are:

DSB_s: Double strand break ends that are available for direct binding by MRN. Since each DSB consists of two double strand ends, the number of DSB_s is double that of double strand breaks.

MRN_i: MRN bound to a double strand end.

AMRN_i: ATM bound to an MRN_i.

ATM_p: Activated ATM.

MγH2AX: MDC1 bound to γH2AX in the surrounding chromatin.

MMγH2AX: MRN bound to MDC1 in the surrounding chromatin.

AMγH2AX: ATM bound to recruited MDC1.

AMMγH2AX: ATM bound to recruited MDC1 that also binds MRN.

Equation (3.9) is added to keep binding of ATM and MRN fully independent. There are no new reaction parameters introduced with equations (3.8) and (3.9) because they use the same reaction rate parameters as equations (3.7) and (3.6).

All simulations were performed using the netdyn python package for chemical reaction computing that was developed by the author and is available online at `www.danielloeb.eu/netdyn.html`. The package automatically generates ODEs from the chemical reaction network, which are then solved using the Runge-Kutta Cash-Karp method [119].

For the optimization of the model parameters, a series of twelve recruitment data sets for NBS1, three data sets for MDC1 and one ATM recruitment data set was used (see section 3.1). In each optimization step, for each NBS1 and ATM data set a calculation was performed, with all of calculations using identical parameters for reaction rates (shown in table 3.1), total concentrations (shown in table 3.2) and data set scaling (shown in table 2 in appendix 3). The only parameter that changed between simulation runs is the number of DSBs, which was obtained from the LET value of each data set. Least squares between data points and the corresponding function values were summed up over all calculations to serve as the optimization measure.

Reaction	**Rate constant**
$\mathrm{MRN} + \mathrm{DSB}_s \rightarrow \mathrm{MRN}_i$	$1.01244 \cdot 10^{-7}$
$\mathrm{ATM} + \mathrm{MRN}_i \rightarrow \mathrm{AMRN}_i$	$3.63249 \cdot 10^{-6}$
$\mathrm{AMRN}_i \rightarrow \mathrm{ATM}_p + \mathrm{MRN}_i$	0.98329
$\mathrm{H2AX} + \mathrm{ATM}_p \rightarrow \gamma\mathrm{H2AX} + \mathrm{ATM}_p$	$1.81244 \cdot 10^{-4}$
$\mathrm{MDC1} + \gamma\mathrm{H2AX} \rightleftharpoons \mathrm{M}\gamma\mathrm{H2AX}$	$3.59072 \cdot 10^{-8}$
$\mathrm{MRN} + \mathrm{M}\gamma\mathrm{H2AX} \rightarrow \mathrm{MM}\gamma\mathrm{H2AX}$ $\mathrm{MRN} + \mathrm{AM}\gamma\mathrm{H2AX} \rightarrow \mathrm{AMM}\gamma\mathrm{H2AX}$	$6.89539 \cdot 10^{-7}$
$\mathrm{M}\gamma\mathrm{H2AX} + \mathrm{ATM}_p \rightarrow \mathrm{AM}\gamma\mathrm{H2AX}$ $\mathrm{MM}\gamma\mathrm{H2AX} + \mathrm{ATM}_p \rightarrow \mathrm{AMM}\gamma\mathrm{H2AX}$	$3.30756 \cdot 10^{-6}$

Table 3.1: **Optimization results for the rate constants of the DSB response model.** All rate constants for reactions that are not listed here were determined from experiment (see text).

Due to the slow diffusion of MDC1, less agreement between the model and MDC1 recruitment data was expected than between the model and other data sets[1]. For this reason,

[1] A side note on why this is expected: For a slowly diffusing protein, the dynamics will at some point become diffusion-limited, with areas where the free population of that protein is locally reduced or depleted. Since the entire free population of a protein is treated as a single variable in the mass action kinetics equations, the results of these equations cannot reproduce such transport effects.

Reactant	Concentration
ATM_0	253260
$MDC1_0$	177163
MRN_0	133479
$H2AX_0$	3494.97

Table 3.2: **Optimization results for the conserved total concentration values of the participating reactants.**

the MDC1 data sets were only used to determine the time at which MDC1 recruitment is saturated. A term was added to the optimization measure that, for the measurement LETs, punished saturation at a later time than in the data. All optimizations were done using the Nelder-Mead downhill simplex algorithm [120] provided by the python scipy package.

3.3 Results

Figure 3.4 shows a comparison of the model and recruitment data for three representative NBS1 data sets, as well as the ATM data set. For low LETs, NBS1 recruitment saturates at later times than for high LETs, which is in agreement with the experimental data. Another notable feature of the experiments is that beyond an LET of 5000 keV/µm, the shape of the recruitment signal ceases to change with further LET increase. This effect is also observed in the simulation results and can be explained by the properties of the two distinct MRN binding interactions.

In the model, binding at the inner focus contributes significantly to the MRN concentration at high LETs, while it is almost negligible at very low LETs. This is a consequence of the direct proportionality between the number of inner MRN binding sites and LET. Taken alone, the shape of the inner focus MRN recruitment curve is independent of LET (with the exception of a minor overshoot due to ATM binding around 200 s). This is not surprising, since the total binding site number for substrate binding can be eliminated from the corresponding differential equations through rescaling[2].

The number of binding sites in the outer focus is always the same, independently of LET. There are small differences in the shape of the outer focus recruitment curve that are caused by the dependence of outer focus binding site availability on inner focus activity (i.e., H2AX has to be phosphorylated by ATM in order to bind MDC1 and MRN). However, these differences are marginal compared to the changes in curve shape with LET of the entire focus recruitment.

The inner focus dynamics, whose contribution increases with LET, is faster than the outer focus dynamics. Consequently, the faster saturation for higher LETs reflects a shift in the shape of the recruitment curve from that of the outer focus to that of the inner focus. Once the

[2]This is an important point that has to be emphasized: The number of inner focus binding sites for MRN is LET-dependent. But since it can be eliminated from the differential equations through variable rescaling, the resulting normalized curves all have the same shape, which is independent of LET.

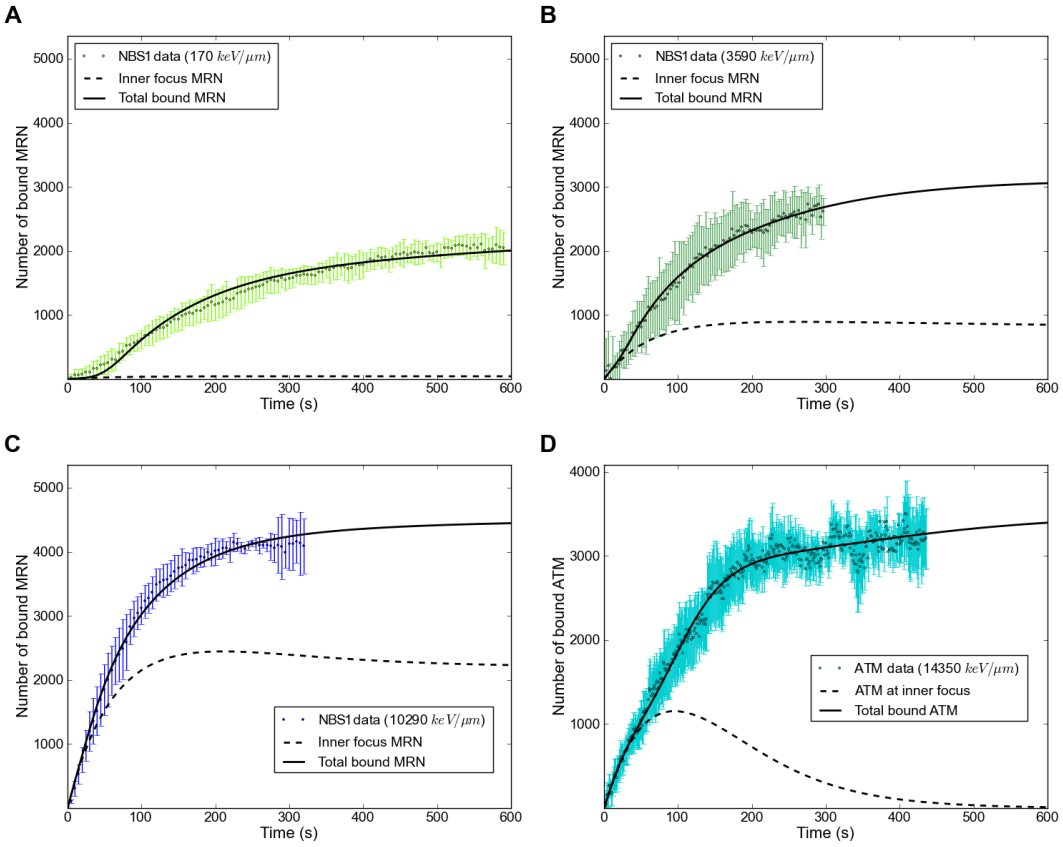

Figure 3.4: **Comparison of NBS1 and ATM recruitment data with model results.** (**A-C**) NBS1 data and NBS1 signal calculated from the recruitment model for LETs of 170 keV/µm, 3590 keV/µm and 10290 keV/µm. Dashed lines indicate the NBS1 signal contribution of MRN recruited to the inner focus (MRN_i), whereas solid lines indicate total recruited NBS1 signal. (**D**) ATM recruitment data and model for an LET of 14750 keV/µm. Dashed line indicates ATM bound at the inner focus ($AMRN_i$), solid line indicates total recruited ATM. The concentration of H2AX in the focus, which limits binding sites for MRN and ATM in the outer focus, has a value of 3364 relative to the scales used here. Additional figures for all recruitment data sets can be found in appendix 3.

inner focus recruitment is the dominating contributor, the curve shape will be very similar to that of inner focus recruitment. As a consequence, the shape of the curve will cease to change with LET. Agreement between the simulations and the experimental data confirms that this is the mechanism of the qualitative change in the LET-dependence of NBS1 recruitment.

ATM recruitment for high LET, as shown in figure 3.4 **D**, is reproduced by the model. The number of bound ATM is divided into those bound at the inner focus and those at the outer focus MDC1 sites. The latter species steadily increases, until all MDC1 sites are occupied, whereas the inner focus bound ATM only exists transiently, as long as there are still free non-activated ATM available for binding. This becomes relevant for very high LETs, where all the effectively available ATM is activated on the recruitment timescale, i.e., within 10 minutes.

The disappearance of the inner focus binding ATM after all ATM has been activated in high LET simulations has an effect on the MRN recruitment, causing a slight decrease in inner focus bound MRN between 200 s and 500 s (see figure 3.4 **C**). The reason for this is that in the model, binding of ATM to the inner focus impedes the dissociation of MRN there, thus shifting the bound concentration towards which the curve converges to a higher value. When the bound ATM at the inner focus decreases as the last remaining non-activated ATM becomes activated, the curve converges towards a reduced concentration. For lower LETs (figure 3.4 **A** and **B**), a state in which all ATM is activated is never reached during the simulation, so that this effect does not appear.

Figure 3.5 **A** shows the activation of ATM in the model for the LET corresponding to the recruitment curve of figure 3.4 **D** and for a representative low LET value. The high-LET curve saturates as all the available ATM are activated, which is reflected in the decreasing number of ATM bound at the inner focus. This effect also leads to a slight premature slope reduction of the ATM recruitment model curve (figure 3.4 **D**) around the time 300 s, where the steady increase in recruitment in the outer focus is temporarily neutralized by the decrease in the inner focus. The curve then continues on to saturate at time 600 s, when almost all binding sites for ATM in the outer focus are occupied.

A comparison of the experimental MDC1 recruitment data and MDC1 recruitment in the simulation shows that only general agreement could be achieved between the model and experiment, due to the relatively slow diffusion of MDC1. Figure 3.5 **B** shows a comparison between a low-LET MDC1 data set and the corresponding model result. The experimental curve has a quasi-constant slope that remains below the simulation result between 100 s and 300 s. This is consistent with a situation in which the MDC1 concentration is locally decreased and the diffusive influx of MDC1 becomes rate-limiting for the recruitment reaction. To test this hypothesis, the model was modified so that the total amount of available MDC1 increases with $(4Dt)^{1/2}$, where $D = D_{\text{eff,MDC1}} = 0.029\,\mu\text{m}^2/\text{s}$ obtained in FRAP measurements. This corresponds to the scaling behavior of diffusion in a cylindrical system. The result of this modified model, shown as the dashed curve in figure 3.5 **B**, agrees better with the MDC1 recruitment data at low LET.

As was noted in the model description, the dissociation of ATM from the outer focus binding sites was not included in the model due to it having no influence on the end results. It has to be clarified that this does not mean that the model requires ATM to bind persistently at the outer focus. If the dissociation of outer focus ATM is included and its parameter is artificially

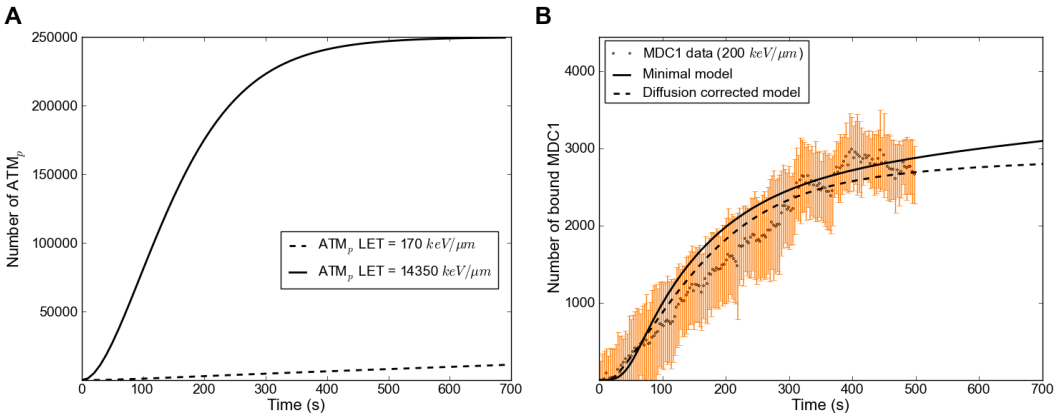

Figure 3.5: **Active ATM in the model and comparison of MDC1 in model and experiment.** (**A**) Activation of ATM in the model for an LET of 170 keV/µm and of 14750 keV/µm. The high LET curve goes into saturation as all of the available ATM is activated. It has to be noted that the absolute maximum value for ATM is a relative value that represents the effective concentration of ATM (due to its fast diffusion throughout the nucleus). (**B**) MDC1 data set for an LET of 200 keV/µm and the corresponding simulation results (solid curve). In this particular calculation, the steady state concentrations for MDC1 are not reached in the first 700 s. The fit at low LET can be considerably improved by taking into account the slow diffusion of MDC1. When the amount of available MDC1 in the simulation is made to increase with $(4Dt)^{1/2}$, as would be the case for diffusion-limited influx in an approximately cylindrical geometry, the dashed curve is obtained.

pushed away from zero (via a penalty function), then there is no change in the quality of the fit. In that case, the large number of available ATM ensures that most binding sites remain occupied with ATM, even in spite of high dissociation turnover. The author considers such transient binding a realistic scenario for outer focus ATM. Nonetheless, since the dissociation parameter increases the parameter space without adding information to the dynamics, it is left out of the minimal model.

Inspired by the numerous ATM interactions reported in the literature, it was investigated how the inclusion of such interactions affects model quality. However, neither auto-phosphorylation of free ATM (see equation (3.10)), as suggested by Mouri et al [86], nor participation of outer focus bound ATM in the phosphorylation of further H2AX (equations (3.11-3.12)) as described by Lou et al [121] improved result quality.

$$\mathrm{ATM}_p + \mathrm{ATM} \rightarrow 2\mathrm{ATM}_p, \tag{3.10}$$

$$\mathrm{H2AX} + \mathrm{AM}\gamma\mathrm{H2AX} \rightarrow \gamma\mathrm{H2AX} + \mathrm{AM}\gamma\mathrm{H2AX}, \tag{3.11}$$

$$\mathrm{H2AX} + \mathrm{AMM}\gamma\mathrm{H2AX} \rightarrow \gamma\mathrm{H2AX} + \mathrm{AMM}\gamma\mathrm{H2AX}. \tag{3.12}$$

Therefore, these two ATM interactions were left out of the minimal model. It has to be noted that this does not mean that auto-phosphorylation as such does not play a role in the model. Indeed, the activation of ATM by MRN directly at the DSB functions through MRN facilitating the auto-phosphorylation of ATM [100]. Furthermore, ATM has multiple auto-phosphorylation sites [122] and fulfills many functions during damage response, so that auto-phosphorylation of free ATM at sites that are not relevant to MRN recruitment likely occurs.

3.4 Discussion

For the recruitment model, the chemical reaction network of the damage response proteins was translated into a set of ODEs, which has been the standard approach for damage response protein dynamics modeling in recent years [86,123,124]. Minimality of the model was ensured by including only those protein interactions that are absolutely necessary to reproduce the data and by further reducing the free model parameters, using FRAP results to pinpoint the dissociation rate constants of both MRN binding modes and MDC1.

Because of its proportionality to the number of double strand breaks, binding of MRN at the inner focus is negligible for the lowest LET data sets investigated with the model, while it becomes the dominating binding type for the highest LETs. The LET-dependence of MRN/NBS1 recruitment seen in the experimental data can thus be explained as a shift from the dynamical behavior of outer focus recruitment at low LET to the behavior of inner focus recruitment at high LET. Both the increase of NBS1 recruitment speed in the low LET range and the independence of LET in the high LET range are therefore reproduced by the simple model presented here.

In addition to the direct comparison of model MRN recruitment with experimental NBS1 recruitment data, there are several consistency considerations indicating that this is indeed the mechanism of LET-dependence in the MRN dynamics. Figure 3.4 **D** shows that the model reproduces the particular shape of the ATM recruitment curve for high LET accurately. Notably,

both the interaction of ATM with the inner focus MRN and the binding of ATM to the outer focus are necessary to achieve this degree of model-experiment agreement.

Fluorescence microscopy imaging showed that when the interaction between MDC1 and MRN is inhibited, NBS1 foci are reduced but do not disappear, indicating that two distinct types of MRN binding are indeed active. Furthermore, the FRAP measurements of the NBS1 dissociation constants showed a relatively large value for low LET which then tended towards the much lower value obtained at CK2 inhibition as LET increases (see figure 3.2). This is consistent with the model result that the inhibited process (i.e. the MRN-MDC1 interaction) becomes less important with increasing LET.

As noted in the results section, the experimental recruitment data of MDC1 goes into saturation faster for high LET measurements than predicted by the model (see figure 9 in appendix 3). A possible explanation is that the model does not contain the experimentally observed nucleus-wide activation of H2AX [125] and subsequent binding of MDC1 for high LETs. A largely reduced free MDC1 population due to MDC1 binding at γH2AX in the entire nucleus could cause such premature saturation. FRAP measurements that showed a reduced mobility of non-focus MDC1 in high LET experiments further indicate this. A preliminary modification of the model to include nucleus-wide interactions showed improved agreement with high-LET MDC1 recruitment data.

In the comparison of ATM activation for different LETs (figure 3.4 **A**), it becomes apparent that only a small fraction of ATM is activated in the first minutes of low-LET irradiation, but all ATM is activated for high LETs. Consequently, outer focus binding sites become available slower in the low-LET case. However, the difference between the outer foci for the lowest and highest LET that were investigated is only a delay in the tens of seconds range. This is not surprising, as the lowest LETs used here (170 keV/µm) corresponds to 28 DSBs, meaning that even the lowest LETs under consideration produce what would be considered a large number of DSBs in a natural environment. Results recently reported by Hable et al [126] on slower MDC1 recruitment for proton irradiation (LET=2.6 keV/µm) support this hypothesis.

It has to be noted that there are several proteins, such as the NHEJ proteins DNA-PKcs and KU70/80, for which involvement in the early damage response has been shown, but which were not included in the model. All investigations here were restricted to the pathway-independent damage response and thus the MRN-ATM-H2AX-MDC1 subunit of the response protein network. It was implicitly assumed that pathway-dependent proteins do not affect the interactions of the model significantly. Since redundancies come into play when the pathway-dependent protein response is considered (for instance between ATM and DNA-PKcs [127]), selective protein knockdown and knockout experiments could help to extend the model in this regard. If the model is to be developed further in this direction, existing results of pathway-dependent modeling [128, 129] should also be taken into account.

4 Multistability and Oscillation in Protein Complex Formation Networks

As was already noted in chapters 1 and 2, protein complex formation plays an important role in many cell-biological processes. The investigation presented in this chapter identifies the minimal prerequisites for multistability in the protein interaction networks that lead to protein complex formation. Because complex assembly itself is the process of interest here, the analysis is restricted to networks that consist exclusively of association and dissociation reactions. After a short description of the type of generic network under consideration, a simple example of a system with multiple steady states will be constructed. Next follows a discussion of the mathematical proof that networks with only three elementary species cannot have multiple steady states. Finally, an example of a system that shows oscillations will be shown and further system properties will be discussed. A paper manuscript based on the text of this chapter will be written and submitted to a peer review journal shortly.

Unfortunately, there is some overlap in nomenclature between the three different scientific disciplines that intersect in this analysis. In the previous chapters, the noun "complex" was used to describe multiple proteins that are chemically bound together. However, "complex" also appears in the context of chemical reaction network theory, where it is used to describe multisets of chemical species. As an example, for a reaction $A + B \rightarrow AB$, the two multisets $A + B$ and AB are the *complexes* of that reaction. For the rest of this chapter, multiple proteins bound together will always be referred to as "protein complex", whereas a multiset of chemical species will be called either "reaction complex" or "complex". The terms "complex system" and "complex dynamics" continue to be used in the sense of "exhibiting dynamical complexity".

4.1 Protein Aggregation Model

It is necessary to first define clearly, what constitutes a generic protein complex formation model in the context of this analysis. In this chapter, all protein interaction systems are regarded as chemical reaction networks, consisting of reactants, reactions and reaction complexes. In order to keep the structure of the systems as simple as possible, the following restrictions are imposed on them:

(i) Protein species denoted by a single letter, such as A and B, are considered "elementary" proteins. These proteins are indivisible reaction participants.

(ii) Protein complexes are reactant species that consist of multiple elementary proteins. Every elementary protein can appear at most once in any protein complex. Protein

complexes consisting of the same combination of elementary species are considered as indistinguishable, independently of their order of assembly. The name of any protein complex, such as AB, is assembled from the names of the elementary proteins it consists of.

(iii) The system is closed and the total amount of each elementary protein is conserved.

(iv) The only chemical reactions that are allowed in the network are the assembly of two reactants[1] to a protein complex and the dissociation of a protein complex into two reactants:

$$A + B \rightleftharpoons AB.. \tag{4.1}$$

Either of these two reactions may also appear without its reversed counterpart.

It follows that the number of possible protein species, i.e., possible combinations, increases with the number of elementary protein species. According to these definitions and restrictions, an aggregation system with three elementary species A, B, and C can have the compound protein species AB, AC, BC and ABC. Its complete set of possible reactions is

$$A + B \rightleftharpoons AB, \tag{4.2}$$
$$A + C \rightleftharpoons AC, \tag{4.3}$$
$$B + C \rightleftharpoons BC, \tag{4.4}$$
$$ABC \rightleftharpoons \begin{cases} AB + C, \\ AC + B, \\ BC + A. \end{cases} \tag{4.5}$$

While it is self-evident that network (4.1) with only two elementary species A and B will always settle into a single, stable steady state, this is not obvious for networks such as (4.2-4.5), with three elementary species and more.

4.2 Minimal Multistable Network

Thomas's conjecture [130] states that a positive feedback loop is a necessary (but not sufficient) condition for multiple fixed points in a dynamical system. Christophe Soulé [84] proved that the conjecture holds for differential mappings within an open finite dimensional real vector space, thus validating it for systems of chemical reactions with non-zero concentrations. A positive feedback loop in this context means that for the system there must exist some set of parameters, at which one of the systems species concentrations, in a value range above some threshold, causes an accelerating increase in itself. More strictly, it must be possible to find a concentration vector x for which a closed circuit of entries $J_{i,j}(x), J_{j,k}(x)...J_{m,i}(x)$ of the system

[1] Whenever the word reactant is used in this chapter, it means any possible reaction participant, independent of whether it is an elementary protein or a protein complex.

Jacobian matrix $J(x)$ exists that is positive, i.e., for which the product of all entries' signs is positive.

Soulés proof holds only on open sets, because on closed sets (i.e., if concentrations of zero are allowed) zeros can appear on the set boundaries. This limitation is of no consequence at this point, since the aim of this analysis is to find a minimal network that fulfills the feedback loop requirement of Thomas's conjecture, and which conforms to the restrictions of the previous section. For the sake of clarity, it is useful to first show the minimal bistable reaction system and then discuss the role of each of its parts and how they relate to the Thomas/Soulé conditions.

The minimal bistable protein aggregation system consists of the following reactions:

$$AB \rightarrow A + B, \tag{4.6}$$
$$A + BC \rightarrow ABC \rightarrow AB + C, \tag{4.7}$$
$$B + C \rightarrow BC, \tag{4.8}$$
$$A + C \rightleftharpoons AC, \tag{4.9}$$
$$B + D \rightleftharpoons BD. \tag{4.10}$$

It has nine protein species (A, B, C, D, AB, AC, BC, BC, ABC) and four elementary protein species. Figure 4.1 shows a bifurcation diagram of the AB concentration in this system for variation of the rate constant for reaction $ABC \rightarrow AB + C$. The corresponding rate constant values can be found in table 4.1

$AB \rightarrow A + B$	0.21071	$A + C \rightarrow AC$	50.9297
$A + BC \rightarrow ABC$	53.2370	$AC \rightarrow A + C$	0.57276
$ABC \rightarrow AB + C$	Parameter	$B + D \rightarrow BD$	1.00966
$B + C \rightarrow BC$	1.51774	$BD \rightarrow B + D$	1.7182818

Table 4.1: Rate constant values for all reactions, as used to generate figure 4.1.

If a system such as (4.6-4.10) can have two fixed points α and β (shown in figure 4.1) for the same set of rate constants and total elementary protein concentrations, this has certain implications for the protein flux[2] through its feedback loop. At one fixed point, there must be a sustained flux through the feedback loop that is different from the flux at the other fixed point.

If, for example, the reaction $AB \rightarrow A + B$ is a part of the feedback loop, a sustained flux means that this reaction is continuously taking place. But for this reaction to be sustained, there must be reactions that again generate AB from the A and B that the reaction produces, or else the reaction will stop due to depletion of AB.

Trivially, merely adding the reversed reaction $A + B \rightarrow AB$ would not produce a feedback

[2]Flux here means the protein concentration per time that is consumed/produced by a reaction.

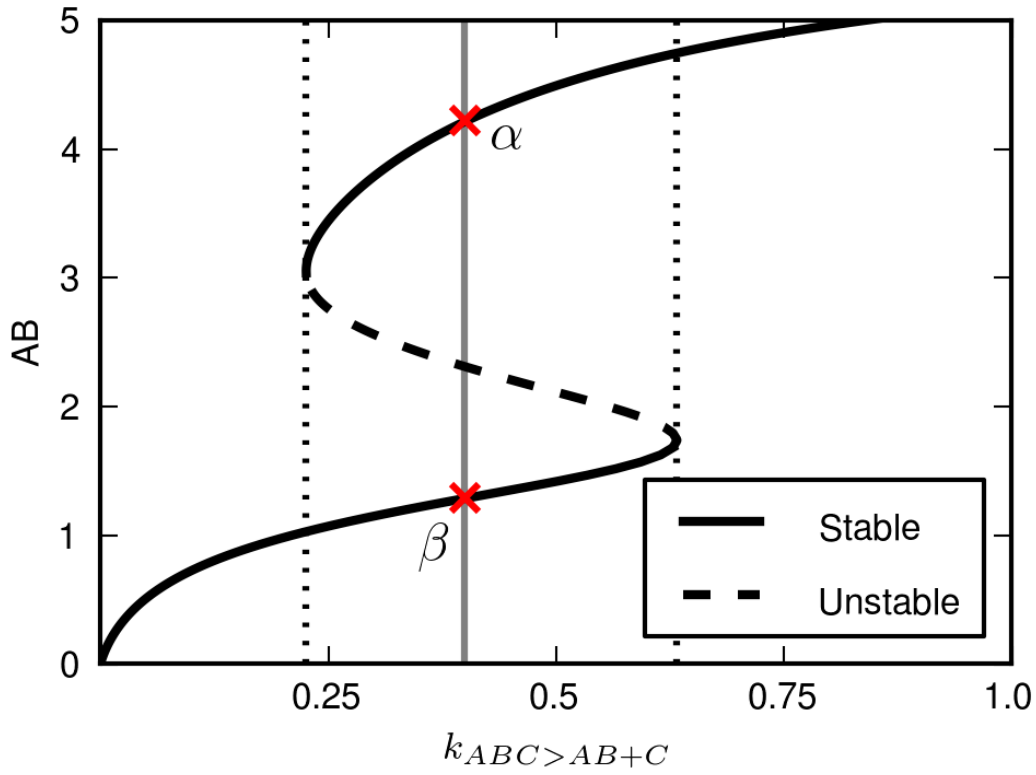

Figure 4.1: **Bifurcation diagram of multistable system with four elementary proteins.** The reaction rate constant for the reaction $ABC \rightarrow AB + C$ in system (4.6 4.10) was varied from 0 to 1 using the bifurcation analysis tool auto [131]. Protein species AB is used to show the zone of bistability between rate constant values of 0.23 and 0.63. The rate constants for all other reactions are shown in table 4.1. Table 4.2 shows the concentrations of all protein species for the fixed points α and β at a $ABC \rightarrow AB + C$ rate constant value of 0.4.

Reactant	Value at α	Value at β
A	$2.28687 \cdot 10^{-2}$	1.45693
B	0.72849	4.96146
C	0.80335	$3.60407 \cdot 10^{-2}$
AB	4.21547	1.28802
AC	1.63359	4.66906
BC	0.72958	$3.49904 \cdot 10^{-3}$
ABC	2.22057	$6.78486 \cdot 10^{-1}$
D	1.51531	$5.52683 \cdot 10^{-1}$
BD	0.64864	1.61127

Table 4.2: **Reactant species concentrations for fixed points α and β in figure 4.1.**

loop[3]. Thus, if the reaction $AB \rightarrow A + B$ is part of a feedback loop, there must be other reactions that regenerate AB from A and B on an indirect path. The most simple reaction system that does so is:

$$AB \rightarrow A + B, \tag{4.11}$$

$$A + BC \rightarrow ABC \rightarrow C + AB, \tag{4.12}$$

$$B + C \rightarrow BC. \tag{4.13}$$

The Jacobian of this system has a positive circuit, and no reaction can be removed without eventually causing depletion of a protein species.

Interestingly, even though a positive circuit exists, this system cannot support a functioning feedback loop yet. The reason again is the closedness of the system, more specifically, the conservation of elementary reaction species. Supposing that a feedback loop were to increase the concentration of A, it follows from mass action kinetics that $A + BC \rightarrow ABC$ will result in an increase of ABC, which in turn leads to an increase in AB, closing the flux loop of A via $AB \rightarrow A + B$. This, however, is impossible, because it is a simultaneous increase of all species that contain the elementary species A, and thus in conflict with the requirement that the total amount of A be conserved. This same argument can be repeated for all species involved in the system and will always result in a violation of one of the systems conservation laws.

The solution to this problem is to introduce "buffer" species whose concentration can be reduced to accommodate growth in the concentrations of other species. In the case of system (4.11-4.13), one such species could be AC, coupled to A and C via the reversible reaction:

$$A + C \rightleftharpoons AC. \tag{4.14}$$

Since these are the only reactions in which AC participates, its concentration is coupled to those of A and C. Through this interaction, changes in the concentration of one of the latter

[3] While adding this reaction to the system in principle does not have any adverse effect regarding the involvement of $AB \rightarrow A + B$ in a feedback loop, it cannot provide the reverse concentration flux needed. If these two reactions balance each other out, they cannot at the same time increase each other's rates.

two species can be used to "adjust" the concentration of the other. If, for instance, C were to be decreased, the concentration of AC would adjust through the reactions such that the relation between the three species again becomes:

$$k_{A+C\to AC}\, A \cdot C = k_{AC\to A+C}\, AC. \tag{4.15}$$

In the process, the concentration of A and A-based species in the loop will increase as the concentration of AC decreases. This process becomes particularly efficient if $k_{A+C\to AC} \gg k_{AC\to A+C}$, that is, if much A is stored in AC.

Thus, the introduction of a buffer species allows the concentration adjustments needed in loop activity, but only for either one of two elementary species. However, as noted before, three elementary species participate in the circular reaction path, all of which need the flexibility to adjust their species concentrations. Since with the inclusion of AC all possible species in a system of three elementary species are now in use, it follows that a system with a working feedback loop is impossible via this approach if only three elementary species are available.

Adding a fourth elementary species to the system increases the number of available reactant species from 7 to 15. It is thus possible to introduce the buffer species BD and the reaction:

$$B + D \rightleftharpoons BD. \tag{4.16}$$

$$\tag{4.17}$$

It is necessary to include the elementary species D itself to make these reactions possible. With this, conservation of all three elementary protein species that participate in the feedback loop structure can be upheld and the fully bistable reaction system (4.6-4.10) is obtained. It has to be noted that the feedback loop itself is only a necessary but not sufficient condition for multistability.

Making all of the participating reactions reversible will not qualitatively change the dynamics of the system, as long as the introduced reactions are sufficiently weak.

It is useful to consider the biological meaning of the bistable system presented here. Table 4.2 shows the differences between concentrations at a pair of steady states in the bistable parameter range. The concentrations of proteins AB and AC are especially interesting, because their fixed point concentrations "change roles" between the two fixed points. This could be interpreted as a decision making process in the sense that, depending on some condition[4], either the protein complex AB is assembled or the protein complex AC is assembled.

The concentration differences between fixed points for each of these two protein species is only a factor of four, but this could be increased by an adjusted choice of rate constants.

So far, it was shown by example that an association/dissociation system with four elementary species can have multiple steady states. In the construction of the system, the fourth species had to be introduced in order to give the other species the necessary flexibility to participate in the feedback loop. However, this demonstration that at least four elementary species are necessary to construct a multistable system based on the reaction structure (4.11-4.13) does not constitute a rigorous proof that multistability is impossible in a system with only three elementary species. Such a proof will be given in the next section.

[4] "Some condition" in terms of the dynamical system presented here means either a change in species concentrations or, alternatively, a change in one of the rate constants.

4.3 No Multistability in ABC Systems

The calculations presented in this section are in part based on the master thesis of Christopher Priester [132], which was overseen by the author in the course of his thesis work. Parts of the formula typesetting were adapted from that thesis.

In the previous section, it was shown that no positive feedback loop (as required for multistability by Thomas [130] and Soulé [84]) could be constructed in a network with only three elementary protein species. While this constructive approach in principle shows that multistability in a network with three elementary species is impossible, this section provides a mathematically rigorous proof. The proof will state that multiple steady states, regardless of their stability, are impossible in the most generic such reaction network for three elementary species. The proof is phrased in the context chemical reaction network theory [133–140] (**CRNT**), an extensive mathematical framework for the analysis of chemical reaction networks, making use of the *deficiency one algorithm* [135] of CRNT. The full network for three species, as already shown in equations (4.2-4.5) is:

$$A + B \rightleftharpoons AB, \tag{4.18}$$

$$A + C \rightleftharpoons AC, \tag{4.19}$$

$$B + C \rightleftharpoons BC, \tag{4.20}$$

$$ABC \rightleftharpoons \begin{cases} AB + C, \\ AC + B, \\ BC + A. \end{cases} \tag{4.21}$$

CRNT analysis uses the structural properties of chemical reaction networks to derive qualitative information about their dynamics. To this end, the network (4.18-4.21) is described in terms of the set of species $\mathscr{S}$ (here: $\{A, B, AB, C, AC, BC, ABC\}$), the set of *complexes*[5] $\mathscr{C}$ (here: $\{A + B, AB, A + C, AC, B + C, BC, AB + C, BC + A, AC + B, ABC\}$) and the set $\mathscr{R}$ of all reactions in the system (here: equations (4.18-4.21) themselves). Complexes $y \in \mathscr{C}$ always have a corresponding basis vector ω_y in "*complex space*" $\mathbb{R}^{\mathscr{C}}$, and a vector in "*species space*" $y \in \mathbb{R}^{\mathscr{S}}$, with y_i being the stochiometric coefficient of species i in complex y.

From the notation of system (4.18-4.21), it becomes apparent that the reaction network is sub-structured into groups of complexes that are connected by reactions. These groups are called *linkage classes* and are the equivalent of connected components in graph theory. If there is a directed connection between every two complexes in a linkage class, it is called a *strong linkage class*; if no complex in a strong linkage class reacts to a complex outside of it, it is called a *terminal strong linkage class*.

If x is a vector in species concentration space, then the dynamics of the chemical reaction system is given by the differential equation $\dot{x} = F(x)$. The system is said to be in a *steady state*, if for its concentration vector x, the relation $\dot{x} = 0$ holds (which means that the system will remain in that state indefinitely). Concentrations of steady states must thus be in the kernel of the nonlinear map F. The structure of the chemical reactions allows to decompose F into

[5] Here and in the rest of this proof, the word "complex" is used to describe reaction complexes (as opposed to protein complexes).

three independent maps $F = YA_\kappa\Psi$, where Ψ maps a vector of species concentrations onto the corresponding representation in complex space, A_κ contains the reactions between complexes in complex space, and Y maps the complexes back into species space.

Since for each of the maps the null-vector is contained in the kernel, the kernel of F consists of $\ker(\Psi)$, $\mathrm{Im}(\Psi) \cap \ker(A_\kappa\Psi)$ and $\mathrm{Im}(A_\kappa\Psi) \cap \ker(YA_\kappa\Psi)$. Vectors in $\ker(\Psi)$ can be neglected since these are combinations of species not able to form any complex, i.e., are not part of the network. For vectors in $\mathrm{Im}(\Psi) \cap \ker(A_\kappa\Psi)$, a steady state exists due to the balancing out of production and consumption of complexes. More formally, it has been shown that if a steady state exists in $\ker(A_\kappa\Psi)$, then there can only be one steady state in each stochiometric compatibility class [137].

The flux through a complex is the difference between the amount of the complex generated through its incoming reactions and the amount consumed by its outgoing reactions, per unit time. Elements in $\mathrm{Im}(A_\kappa\Psi) \cap \ker(YA_\kappa\Psi)$ represent steady states with nonzero complex fluxes, but for which the species concentrations, and thus the summed contributions of the complexes to each species, nonetheless balance out. The dimension of $\mathrm{Im}(A_\kappa\Psi) \cap \ker(YA_\kappa\Psi)$, also called the deficiency [133] δ, is important for the existence of multiple steady states:

$$\delta := \dim\left(\mathrm{Im}(A_\kappa\Psi) \cap \ker(YA_\kappa\Psi)\right). \tag{4.22}$$

It was shown in [134] that δ can be calculated using

$$\delta = n - l - s, \tag{4.23}$$

where n is the number of complexes, l is the number of linkage classes in the reaction-network and s is the rank of the stochiometric matrix [133]. Based on the deficiency of the network, it can be decided, which of the theorems and algorithms of **CRNT** can be applied. With 10 complexes, 4 linkage classes and a stochiometric rank of 4, network (4.18-4.21) has a deficiency $\delta = 2$, even though all linkage classes individually have a deficiency of zero.

The deficiency one algorithm will be used to analyze this network, even though it is not a deficiency one network. It is possible to do so since, as will be explained below, there are symmetries in this network that effectively reduce its deficiency to one. A mathematical derivation of the deficiency one algorithm can be found in [134].

The deficiency one algorithm can only be applied to networks matching the following criteria, all three of which are fulfilled by network (4.18-4.21):

(i) A set of rate constants (and conserved quantities) is known, such that the chemical system exhibits a steady-state.

(ii) Each linkage class of the network contains exactly one terminal strong linkage class.

(ii) Removing a reaction (both reactions for a reversible reaction) from a terminal strong linkage class leads to a decomposition of the terminal strong linkage class.

The precise manner in which the production and consumption of each species are balanced at the steady state(s) depends on the reactive flux through all of the reactions. In the context of the deficiency one algorithm, the reactive flux at the steady state is defined as the *confluence vector* $g \in \mathscr{R}^n$, a vector whose components g_y are the net flux for complex y. A confluence vector must have the following properties:

(i) $\sum_{y\in\mathscr{C}} g_y y = 0$

(ii) $\sum_{y\in\mathscr{L}} g_y = 0$ for each linkage class $\mathscr{L}$ of the network

(iii) $\sum_{y\in\mathscr{L}'} g_y > 0$ for each terminal strong linkage class $\mathscr{L}'$ which is not identical to the linkage class containing it.

Property (i) is the steady state condition, (ii) is the equivalent of Kirchhoff's first law and (iii) ensures that no flux can leave a terminal strong linkage class. The confluence vectors of the network span the basis of $\mathrm{Im}\,(A_\kappa\Psi) \cap \ker\,(YA_\kappa\Psi)$. In reversible networks, the negative multiples of these vectors are also valid confluence vectors and have to be included in all further considerations. For network (4.18-4.21), the map Y is:

$$Y = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \tag{4.24}$$

where each complex in the system has a column containing the stochiometric coefficients of its species. Since the deficiency of network (4.18-4.21) is two, a valid basis must be spanned by two linearly independent confluence vectors. A pair of such vectors is:

$$g_1 = (-1, 1, 0, 0, 1, -1, -1, 0, 1, 0) \qquad \text{and} \tag{4.25}$$

$$g_2 = (0, 0, -1, 1, 1, -1, 0, -1, 1, 0)\,, \tag{4.26}$$

which allows for possible confluence vectors of the form:

$$g = (-\alpha, \alpha, -\beta, \beta, \alpha+\beta, -\alpha-\beta, -\alpha, -\beta, \alpha+\beta, 0) \tag{4.27}$$

with $\alpha, \beta \in \mathbb{R}$. It is apparent from the network itself that the system should be symmetrical regarding permutations of the elementary species A, B and C. This symmetry can also be observed in the behavior of the generic confluence vector expression for this network (4.27). The structure of vector (4.27) allows six distinguishable sign configurations for α and β:

$$\begin{aligned} &\{\alpha > 0, \beta > 0\}, && \{\alpha < 0, \beta < 0\}, \\ &\{\alpha > 0, \beta < 0, |\alpha| > |\beta|\}, && \{\alpha < 0, \beta > 0, |\alpha| > |\beta|\}, \\ &\{\alpha > 0, \beta < 0, |\alpha| < |\beta|\}, && \{\alpha < 0, \beta > 0, |\alpha| < |\beta|\}. \end{aligned}$$

However, any one of these sign configurations can be transformed into any other, using only elementary species permutation and vector sign reversal. The most notable difference between the sign configurations concerns the fluxes in the largest linkage class (4.21). For the

configurations listed on the left, influx into ABC passes through one complex, while outflux passes through two. Since the configurations on the right side are sign reversed to those on the left side, this relation is also reversed. With all variations of α and β effectively resulting in the same confluence vector, system (4.18-4.21) can be treated as if it was a deficiency one network.

Using the confluence vectors furthermore allows to find the net flux through each reaction of the network. For a given confluence vector g the flux through a reaction $y \rightleftharpoons y'$ is given by

$$[g, y \rightleftharpoons y', y] := \sum_{\bar{y} \in L(y)} g_{\bar{y}} \quad , \tag{4.28}$$

where L is the linkage class containing the complex y after removing the reaction $y \rightleftharpoons y'$ from the network. Since the confluence vector can be rescaled, only the sign of $[g, y \rightleftharpoons y', y]$ is relevant for the algorithm.

The existence of two steady-states for a network $\{\mathscr{S}, \mathscr{C}, \mathscr{R}\}$ with rate constants $\mathbf{k} \in (\mathbb{R}^+)^{\dim(\mathscr{R})}$ requires the existence of two nonequal, strictly positive concentration vectors $c^{\star\star}, c^{\star} \in (\mathbb{R}^+)^n$ solving the equations

$$\sum_{y \longrightarrow y' \in \mathscr{R}} k_{y \longrightarrow y'} \left(c^{\star}\right)^y (y' - y) = 0, \tag{4.29}$$

$$\sum_{y \longrightarrow y' \in \mathscr{R}} k_{y \longrightarrow y'} \left(c^{\star\star}\right)^y (y' - y) = 0. \tag{4.30}$$

The notation c^y here, corresponds to the mass-action term $\prod_{\forall i} c_i^{y_i}$. Assuming the existence of $c^{\star}$ solving (4.29), expanding $(c^{\star\star})^y$ to $(c^{\star\star})^y \cdot \left(\frac{c^{\star}}{c^{\star}}\right)^y$ and introducing the vector $\mu_s = \ln\left(\frac{c_s^{\star\star}}{c_s^{\star}}\right)$, the equations can be modified to:

$$\sum_{y \longrightarrow y' \in \mathscr{R}} \kappa_{y \longrightarrow y'} (y' - y) = 0 \qquad \text{and} \tag{4.31}$$

$$\sum_{y \longrightarrow y' \in \mathscr{R}} \kappa_{y \longrightarrow y'} e^{y \cdot \mu} (y' - y) = 0, \tag{4.32}$$

where $\kappa_{y \longrightarrow y'} = k_{y \longrightarrow y'} \left(c^{\star}\right)^y$. The newly introduced vector μ has some useful properties. Conservation of the elementary species means that, for the set of all species containing a particular elementary species (as for instance $\{A, AB, AC, ABC\}$ in the case of elementary species A), there must always be entries in c^{**} that are bigger than their counterpart in c^*, as well as entries that are smaller than their counterpart. Otherwise, the total amount of the elementary species would have to increase or decrease, in violation of the conservation restriction. This in turn means that the entries or μ for that same set of species cannot all have the same sign. From the properties of the logarithm follow further restrictions on the signs of the components of μ:

$$\operatorname{sign}\left(c_i^{**} - c_i^*\right) = \operatorname{sign}\left(\ln \frac{c_i^{**}}{c_i^*}\right) = \operatorname{sign}\left(\mu_i\right) \qquad \forall i \in \mathscr{S}. \tag{4.33}$$

This means that the sign of all entries in μ must be the same as the sign of the corresponding entries in the difference between the two steady state vectors. Those steady states are in the

same stochiometric subspace, i.e., it must be possible to go from one to the other by using only the chemical reactions of the system. It follows that the signs of the entries in μ have to be such that a vector with the same combination of signs can be constructed from the reaction vectors $y' - y$ of the system. This property is called the condition of sign compatibility with the stochiometric subspace. Using the maps Y and A_κ introduced above, equations (4.29-4.30) simplify to:

$$YA_\kappa \omega_{\mathscr{C}} = 0 \tag{4.34}$$

$$YA_\kappa \left(\sum_{y \in \mathscr{C}} e^{y \cdot \mu} \omega_y \right) = 0 \tag{4.35}$$

The vectors ω_y form the standard basis of $\mathbb{R}^{\mathscr{C}}$ and $\omega_{\mathscr{C}}$ is defined as $\omega_{\mathscr{C}} := \sum_{y \in \mathscr{C}} \omega_y$. Equations (4.29) and (4.30) are then equivalent to the condition that $\omega_{\mathscr{C}}$ and $\sum_{y \in \mathscr{C}} e^{y \cdot \mu} \omega_y$ should be elements of $\ker(YA_\kappa)$. The assumed existence of one steady-state warrants that equation (4.34) and thereby $\omega_{\mathscr{C}} \in \ker(YA_\kappa)$ holds true.

The kernel $\ker(YA_\kappa)$, as discussed by Feinberg [135] has several important properties:

(i) The dimension $\dim(\ker(YA_\kappa))$ equals $1 + l$, where l is the number of linkage classes.

(ii) A basis of $\ker(YA_\kappa)$ can be constructed by taking $\ker(A_\kappa)$ and the vector $\omega_{\mathscr{C}}$ if the latter is linearly independent of $\ker(YA_\kappa)$.

(iii) The basis $\{b_1, b_2, \ldots, b_l\} \subset (\mathbb{R}^+)^N$ of $\ker(A_\kappa)$ has the property:

$$\operatorname{supp}(b_i) = \Lambda_i \tag{4.36}$$

where $\operatorname{supp}(x) := \{y \in \mathscr{C} \mid x_y \neq 0\}$ and $\{\Lambda_1, \Lambda_2, \ldots, \Lambda_l\}$ is the set containing all terminal strong linkage classes.

Since $\omega_{\mathscr{C}} \in \ker(YA_\kappa)$, the existence of a second steady-state reduces to the requirement of a vector μ solving equation 4.35. Using the basis $\{\omega_{\mathscr{C}}, b_1, b_2, \ldots, b_l\}$ equation (4.35) becomes equivalent to:

$$\sum_{y \in \mathscr{C}} e^{y \cdot \mu} \omega_y = \lambda_0 \cdot \omega_{\mathscr{C}} + \lambda_1 \cdot b_1 + \lambda_2 \cdot b_2 + \cdots + \lambda_l \cdot b_l. \tag{4.37}$$

To calculate the basis $\{\omega_{\mathscr{C}}, b_1, b_2, \ldots, b_l\}$ of $\ker(YA_\kappa)$ it is necessary to explicitly state A_κ for system (4.18-4.21):

$$A_\kappa = \begin{pmatrix} -\kappa_1 & \kappa_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \kappa_1 & -\kappa_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\kappa_3 & \kappa_4 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \kappa_3 & -\kappa_4 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -\kappa_5 & \kappa_6 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \kappa_5 & -\kappa_6 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -\kappa_7 & 0 & 0 & \kappa_8 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\kappa_9 & 0 & \kappa_{10} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\kappa_{11} & \kappa_{12} \\ 0 & 0 & 0 & 0 & 0 & 0 & \kappa_7 & \kappa_9 & \kappa_{11} & -\kappa_8 - \kappa_{10} - \kappa_{12} \end{pmatrix}. \tag{4.38}$$

Using the basis vectors derived from the above expression, equation (4.37) becomes:

$$\begin{pmatrix} e^{\mu_A+\mu_B} \\ e^{\mu_{AB}} \\ e^{\mu_A+\mu_C} \\ e^{\mu_{AC}} \\ e^{\mu_B+\mu_C} \\ e^{\mu_{BC}} \\ e^{\mu_{AB}+\mu_C} \\ e^{\mu_{AC}+\mu_B} \\ e^{\mu_{BC}+\mu_A} \\ e^{\mu_{ABC}} \end{pmatrix} = \lambda_0 \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} + \lambda_1 \begin{pmatrix} 1 \\ \frac{\kappa_1}{\kappa_2} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} 0 \\ 0 \\ 1 \\ \frac{\kappa_3}{\kappa_4} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \lambda_3 \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ \frac{\kappa_5}{\kappa_6} \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \lambda_4 \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ \frac{\kappa_8}{\kappa_7} \\ \frac{\kappa_{10}}{\kappa_9} \\ \frac{\kappa_{12}}{\kappa_{11}} \end{pmatrix}. \tag{4.39}$$

Here, for typesetting reasons, the κ names were replaced using this scheme:

$\kappa_{A+B\to AB}$	κ_1	$\kappa_{B+C\to BC}$	κ_5	$\kappa_{AC+B\to ABC}$	κ_9
$\kappa_{AB\to A+B}$	κ_2	$\kappa_{BC\to B+C}$	κ_6	$\kappa_{ABC\to AC+B}$	κ_{10}
$\kappa_{A+C\to AC}$	κ_3	$\kappa_{AB+C\to ABC}$	κ_7	$\kappa_{BC+A\to ABC}$	κ_{11}
$\kappa_{AC\to A+C}$	κ_4	$\kappa_{ABC\to AB+C}$	κ_8	$\kappa_{ABC\to BC+A}$	κ_{12}

Because each basis vector b has nonzero entries only for a limited number of complexes, equation (4.37) can be decomposed into a set of equations that each involve only one complex and a few of the λ coefficients. Depending on the signs of the coefficients λ, these equations impose conditions upon the relative sizes of the exponent terms $y \cdot \mu$. Two stochiometrically compatible steady states are only possible if a vector μ can be found which solves equation (4.37) for some combination of λ signs and if that vector is sign compatible to the stochiometric subspace of the network.

The inequality systems are constructed as follows: Using the signs of the coefficients λ_i, it is possible to group the strong linkage classes (and thereby the contained complexes) into three different sets:

$$U := \cup\{y \in \Lambda_i \mid \lambda_i > 0\}, \tag{4.40}$$

$$L := \cup\{y \in \Lambda_i \mid \lambda_i < 0\}, \qquad \text{and} \tag{4.41}$$

$$M := \{y \in \mathscr{C} \mid y \text{ is educt of a reaction}, y \notin U, y \notin L\} \tag{4.42}$$

The latter contains all complexes in non-terminal strong linkage classes as well as the set

$$\cup\,\{\Lambda_i \mid \lambda_i = 0\}. \tag{4.43}$$

Using these sets and eq. (4.37), it is possible to derive several requirements on μ. For any complex $y \in M$ the equation

$$e^{y\cdot\mu} = \lambda_0 \tag{4.44}$$

has to hold true. Since the exponential function is injective, it follows that, for a pair of complexes $y, y' \in M$, the relation $y \cdot \mu = y' \cdot \mu$ holds.

Evaluation of eq. (4.37) for complexes $y^U \in U$, $y^M \in M$ and $y^L \in L$ with $y^U \in \Lambda_1$ and $y^L \in \Lambda_2$ leads to:

$$e^{y^U \cdot \mu} = \lambda_0 + \lambda_1 (b_1)_{y^U}\,, \tag{4.45}$$
$$e^{y^M \cdot \mu} = \lambda_0, \tag{4.46}$$
$$e^{y^L \cdot \mu} = \lambda_0 + \lambda_2 (b_2)_{y^L}\,, \tag{4.47}$$

with $(b_1)_{y^U}$ and $(b_2)_{y^L}$ being the components of the basis vectors $b_1, b_2 \in \mathbb{R}^n$ belonging to the respective complex. Since $\lambda_1 > 0 > \lambda_2$ as well as $(b_i)_y \geq 0$ and since the exponential function is strictly increasing it follows that:

$$y^U \cdot \mu > y^M \cdot \mu > y^L \cdot \mu. \tag{4.48}$$

For two complexes y, y' in the same terminal strong linkage class Λ_i and connected by a reaction, eq. (4.37) becomes:

$$e^{y \cdot \mu} = \lambda_0 + \lambda_i (b_i)_y\,, \tag{4.49}$$
$$e^{y' \cdot \mu} = \lambda_0 + \lambda_i (b_i)_{y'}\,. \tag{4.50}$$

If both complexes are in U so $\lambda_i > 0$, this leads to:

$$\operatorname{sign}(y \cdot \mu - y' \cdot \mu) = \operatorname{sign}\left((b_i)_y - (b_i)_{y'}\right). \tag{4.51}$$

In contrast, if both complexes are in L it follows that $\lambda_i < 0$ and :

$$\operatorname{sign}(y \cdot \mu - y' \cdot \mu) = -\operatorname{sign}\left((b_i)_y - (b_i)_{y'}\right) \tag{4.52}$$

Using a confluence vector g and the results from [135], stating that

$$\operatorname{sign}\left((b_i)_y - (b_i)_{y'}\right) = \operatorname{sign}[g, y \leftrightarrow y', y] \quad , \tag{4.53}$$

one of the following relations ensue:

(i) If $y, y' \in U$ holds:

$$y \cdot \mu > y' \cdot \mu \quad \text{if} \quad [g, y \leftrightarrow y', y] > 0, \tag{4.54}$$
$$y \cdot \mu = y' \cdot \mu \quad \text{if} \quad [g, y \leftrightarrow y', y] = 0, \tag{4.55}$$
$$y \cdot \mu < y' \cdot \mu \quad \text{if} \quad [g, y \leftrightarrow y', y]\,. < 0 \tag{4.56}$$

(ii) If $y, y' \in L$ holds:

$$y \cdot \mu < y' \cdot \mu \quad \text{if} \quad [g, y \leftrightarrow y', y] > 0, \tag{4.57}$$
$$y \cdot \mu = y' \cdot \mu \quad \text{if} \quad [g, y \leftrightarrow y', y] = 0, \tag{4.58}$$
$$y \cdot \mu > y' \cdot \mu \quad \text{if} \quad [g, y \leftrightarrow y', y] < 0. \tag{4.59}$$

If there is a choice for g and the sets U, M and L, for which the resulting system of in-/equalities has a solution μ that is sign-compatible with the reactions of the network, then the network has the capacity for multiple steady states.

Applying the deficiency-one algorithm to a network amounts to testing all possible configurations of g and the sets U, M and L. For networks with conservation of concentrations for some species, as is the case with conservation of elementary species here, additional requirements on μ arise. Since the change of all species contributing to a conserved quantity have to sum up to zero, the signs of $\mu_s = \ln\left(\frac{c_s^{\star\star}}{c_s^{\star}}\right)$ for the respective species cannot all be equal.

Another restriction on the signs of the components of μ arises from the way in which the conservation laws are connected among each other through shared species. For instance, the conservation of A requires that the sum of the species concentrations A, AB, AC, ABC remains constant. Likewise, conservation of B demands the same for the species B, AB, BC, ABC. Both AB and ABC are shared between these two conservation laws. If μ_A and μ_{AC} both are positive, it follows that, in sum, the amount of AB and ABC must be lower at c^{**} than at c^{*}. In turn, if both μ_B and μ_{BC} are negative, the total amount of AB and ABC at c^{**} must be higher than at c^{*}. Since these two requirements contradict each other, it follows that, if μ_A and μ_{AC} are both positive, μ_B and μ_{BC} cannot both be negative. Relations of this kind in the system give three conservation exclusion rules:

1st rule: If μ_A and μ_{AC} have the same sign, then μ_B and μ_{BC} cannot both have the opposite sign (and vice versa).

2nd rule: If μ_A and μ_{AB} have the same sign, then μ_C and μ_{BC} cannot both have the opposite sign (and vice versa).

3rd rule: If μ_B and μ_{AB} have the same sign, then μ_C and μ_{AC} cannot both have the opposite sign (and vice versa).

Since all possible U, M, L combinations for the network have to be verified, the number of inequality systems that have to be checked is quite large. For the sake of brevity, these calculations are therefore relegated to appendix 1. They show that there is no valid solution to any of the possible inequality systems for network (4.18-4.21). It thus follows that there cannot be multiple steady states in a system of this type with only three elementary species, regardless of rate constants and initial concentrations.

The steps that were taken in the previous section to ensure that the network fulfills the feedback loop condition are analogous to some elements of the deficiency one algorithm. Having a substructure in the network that allows the circular flux of species means that the rank of the stochiometric matrix for that substructure is less than the number of reactions involved. It follows that the network must have a deficiency of at least one.

The argument for introducing the buffer species in the previous section was that they act as a counterbalance to the concentration differences between two steady states, thus allowing the conservation of elementary species. In this section, this requirement is mirrored by the restrictions on the sign of μ.

4.4 Oscillations

The existence of oscillating chemical systems has long since been proven in theory [141, 142] and experiment [143–145]. In this section, an example of a protein association/dissociation system with four elementary species that shows oscillations is presented. Because the analysis here restricts itself to the systems described in section 4.1, the nonlinear function terms that are typically applied in the modeling of oscillations, such as Hill functions [81–83, 146], cannot be used.

In a biochemical context, oscillating behavior can be obtained by reaction pathways that include a negative feedback loop with a sufficient time delay and some sort of nonlinear component. It has been shown that three different time delay options are available at the differential equation level for biochemical systems [81]: Explicit time delay, intermediate reactions and an additional positive feedback loop for a participating species. Consequently, the construction of an oscillating system while adhering to the restrictions of section 4.1 can be achieved by using one of these three patterns.

Because explicit delay is impossible within the restrictions of the systems analyzed here, only time delay and positive feedback are viable options. Using a positive feedback loop increases the number of reactions by approximately the number of reactions in system (4.6-4.10), so that time delay, realized by intermediate chemical reactions, was chosen instead. The following system was obtained:

$$A + B \rightarrow AB, \tag{4.60}$$
$$AB + C \rightarrow ABC, \tag{4.61}$$
$$ABC + D \rightarrow ABCD, \rightarrow AC + BD, \tag{4.62}$$
$$D + AC \rightarrow ACD \rightarrow A + CD, \tag{4.63}$$
$$CD + B \rightarrow BCD \rightarrow C + BD, \tag{4.64}$$
$$BD \rightarrow B + D. \tag{4.65}$$

It consists of a step-by-step buildup of the protein complex $ABCD$, which then splits into the two smaller protein complexes AC and BD. These continue along independent reaction paths to then dissociate again. To see why these particular reactions are needed for the network to sustain an oscillation, it is helpful to consider how the individual reactions contribute.

As described in section 2.5, chemical mass-action systems are dissipative. A consequence of this is that a disturbance to the species concentrations in a loop-like[6] system of this type typically smears out over time and finally disappears while it moves through the loop structure. If the system was on a fixed point before the disturbance, it will again end up there.

For a sustained oscillation, some element in the system must counteract the dissipation in such a way that the particular "disturbance" of that oscillation can maintain its shape and move through the system indefinitely[7]. To this end, in system (4.60-4.65) the special property

[6] Loop-like in the sense in which the following reactions form a loop: Some protein X reacts with another protein to become Y, which then goes through a reaction to become Z, which then reacts back to X again.

[7] A sustained oscillation in a dissipative system is itself an attractor, meaning that trajectories that start close by will converge to it.

of the $ABCD$ species that it can be separated into two other non-elemental species is used. Downstream from the reaction $ABCD \rightarrow AC + BD$, there are two reaction branches: One involving AC and one starting with BD. If the circular flux of reactants is set up in such a way that one or more of the reactions upstream from $ABCD$ involve reaction products from both downstream branches, it becomes possible to recover the original temporal shape of the oscillation from its smeared out downstream shape.

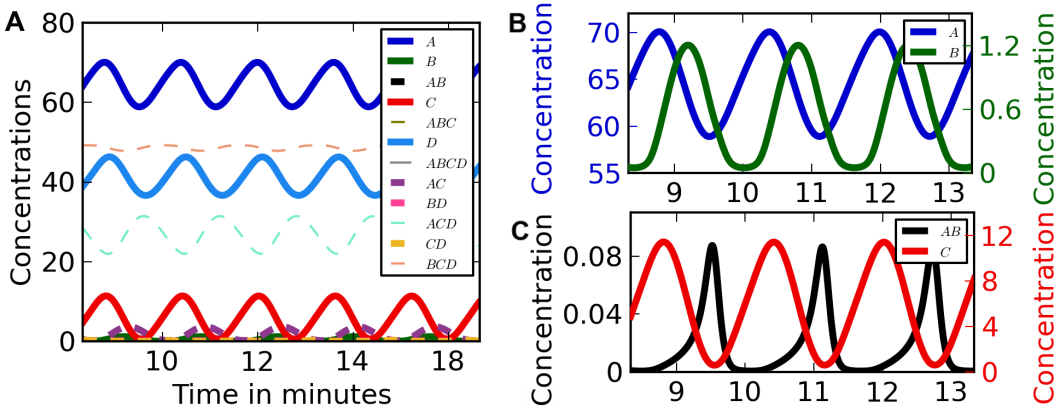

Figure 4.2: **Oscillation in the negative feedback system (4.60-4.65).** All species in the system are plotted in **A**. Subfigures **B** and **C** show the driving species of reactions $A+B \rightarrow AB$ and $AB+C \rightarrow ABC$, respectively, with scaling adjusted for visibility. Vertical lines t_1 and t_2 indicate the times for flux figure 4.3.

The key property of this construct is that the two reaction branches must have different round trip times for their reactants. If this is the case, the upstream reaction involving reactants from both branches can be frustrated by the absence of one reactant, thus provoking a buildup of the other reactant. Once the concentration of the reactant from the other branch rises again, this built up "lump" of reactant travels through the oscillatory loop. The reactant buildup and subsequent release thus boost the oscillation amplitude. The intermediate reactions of equations (4.63) and (4.64) are necessary to make the coordination of round trip times possible.

Figure 4.2 shows the species concentrations during a sustained oscillation in the system, figure 4.3 is a diagram of reactive fluxes in the system at different times in the oscillatory cycle. The oscillation in the reaction branch going out from AC is dampened down so far that its output of reactants A and C fluctuates very little (almost no change in arrow thickness). The BD branch, however, directly relays the oscillation to the concentration of B via $BD \rightarrow B+D$. Thus, the reaction $A + B \rightarrow AB$ functions as the concentration buildup bottleneck. Since A is continuously supplied by $ACD \rightarrow A + CD$, its concentration increases as long as the concentration of B is low. Once the concentration of B rises, $A + B \rightarrow AB$ rapidly increases in strength, dropping off again as B is reactively depleted. The same process also happens at reaction $AB + C \rightarrow ABC$, where C takes the role of A and AB serves as a proxy for the oscillation of B.

A video of the flux dynamics in this system is available at *www.danielloeb.eu/research/ABCD_oscillation_phd.mp4*. Figure 4.4 shows the Hopf bifurcation for this system for the

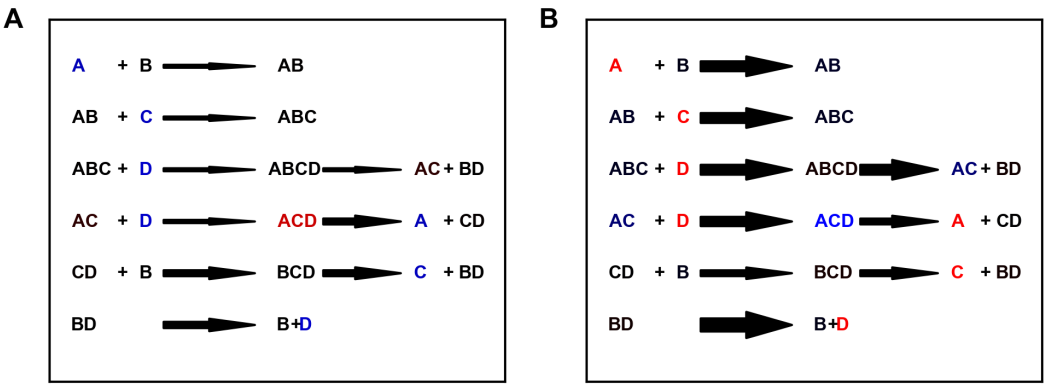

Figure 4.3: **Reactive flux and concentration changes during the oscillation of system (4.60-4.65).** Figure **A** shows flux and concentration changes at the minimum of B and figure **B** shows the same at the maximum of B. The arrow thickness indicates the absolute reaction strength, as given by the mass-action terms $k_{X+Y>XY}\, X \cdot Y$ or $k_{XY>X+Y}\, XY$. The colors of species names represent the reactive change of their concentrations, i.e., the sum of all fluxes involving that species. Red indicates a concentration decline, blue a concentration increase and black indicates an unchanged concentration. Both the main loop upstream from $ABCD$ and the downstream BD branch oscillate, whereas there is very little fluctuation beyond ACD in the downstream AC branch. Concentration colors are scaled so that the maximum/minimum concentration change for each species gives pure blue/red.

rate constant of the reaction $A + B \rightarrow AB$. The requirement that the system must maintain the oscillation "signal" against its dissipative nature is a more precise version of the demand of other authors that the system should be "sufficiently nonlinear to destabilize the steady state" [81].

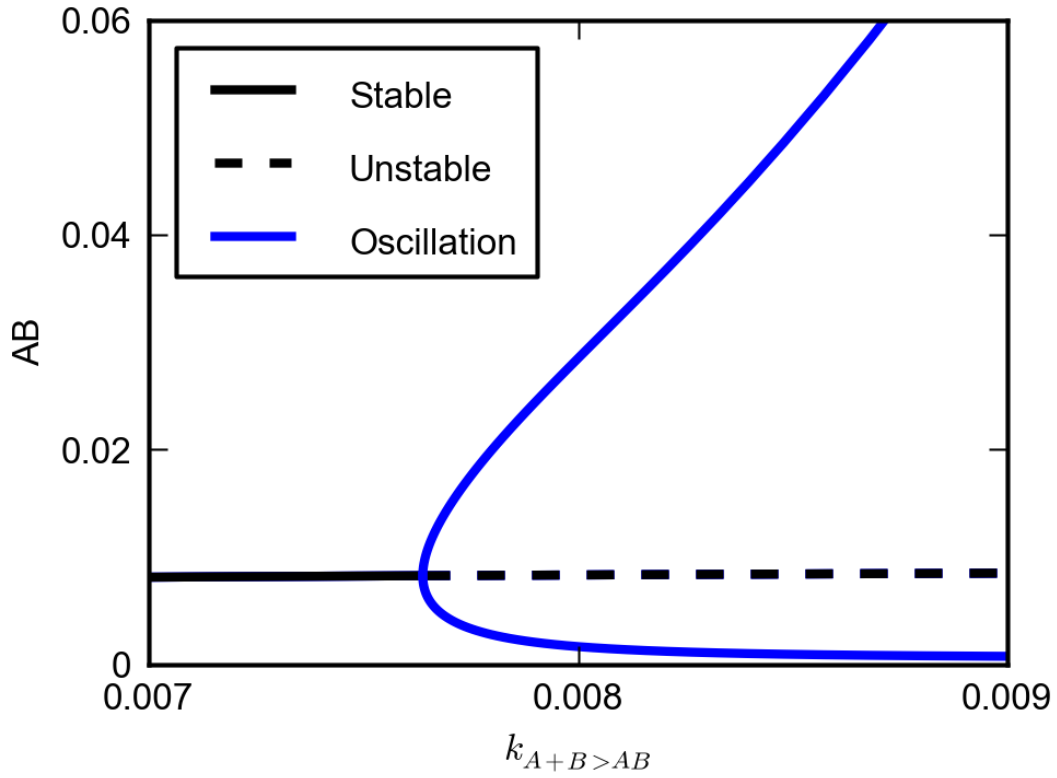

Figure 4.4: **Hopf bifurcation for parameter $k_{A+B>AB}$ in system (4.60-4.65).** At the supercritical Hopf bifurcation ($k_{A+B>AB} = 0.00764$, all other parameters as listed in the supplement), the stable fixed point (solid black curve) becomes unstable (dashed curve), while a stable limit cycle appears. The blue curves denote the maximum and minimum values of AB in the oscillation.

On a side note: It was remarked earlier that Hill and Michaelis-Menten function terms are commonly used components in the analysis of cell-biological dynamics. While the systems investigated here rely on a very restricted subset of nonlinear dynamics, they can reproduce these function terms. The premise of Michaelis-Menten kinetics [147] is that an enzyme E reversibly binds to a substrate S, forming the aggregate protein ES. A catalytic step and subsequent dissociation then convert ES into the original enzyme E and a product P:

$$E + S \rightleftharpoons ES \rightarrow E + P. \tag{4.66}$$

To reproduce the Michaelis-Menten kinetics, the reactants above are identified with generic protein species, as used throughout this section:

Michaelis-Menten	Generic Proteins
E	A
S	B
P	BC

Using these generic proteins, the following system can be constructed, which under certain conditions behaves as the above Michaelis-Menten reaction system:

$$A + B \rightleftharpoons AB, \tag{4.67}$$

$$AB + C \rightarrow ABC, \tag{4.68}$$

$$ABC \rightarrow A + BC. \tag{4.69}$$

If the reaction $AB+C \rightarrow ABC$ is so fast that it is instantaneous compared to the other reactions, the protein species AB exists only transiently. Thus ignoring the very small concentration of AB, the concentration of A can be expressed as $A = A_{sum} - ABC$, with A_{sum} being the total concentration of elementary protein A in the system.

If $A + B \rightarrow AB$ is so fast that ABC reaches its steady state concentration much faster than the other proteins, then, via $\frac{\mathrm{d}ABC}{\mathrm{d}t} = 0$, the following concentration value can be obtained:

$$ABC = \frac{k_{A+B>AB} A_{sum} \cdot B}{k_{ABC>A+BC} + k_{A+B>AB} B} = \frac{A_{sum} \cdot B}{\Theta + B}. \tag{4.70}$$

Here, $k_{A+B>AB}$ and $k_{ABC>A+BC}$ are the rate constants of reactions (4.67) and (4.69) and $\Theta = k_{ABC>A+BC}/k_{A+B>AB}$. Equation (4.70) is identical to the corresponding expression for ES in Michaelis-Menten kinetics. The product BC is then produced according to:

$$\frac{\mathrm{d}BC}{\mathrm{d}t} = k_{ABC>A+BC} \cdot A_{sum} \frac{B}{\Theta + B}. \tag{4.71}$$

Reactions (4.67-4.69) are thus equivalent to a Michaelis-Menten term and could be used as a building block for systems with multiple instances of Michaelis-Menten kinetics. Since the class of systems under consideration here allows only one instance of each elementary protein in a protein complex, a trivial implementation of Hill kinetics in a similar manner is not possible. However, if larger reaction systems with more elementary species are used, a reasonable approximation to the higher order terms can be achieved by using intermediate protein complexes.

In this chapter it was demonstrated that systems with at least four elementary species can have multiple steady states. The result of the deficiency one algorithm in section 4.3 in turn means that systems with three or less elementary species will have at most one steady state for any set of reaction rates and total concentrations.

Using another system as an example, it was shown that systems with four elementary species can have sustained oscillations. It was demonstrated how the Michaelis-Menten function term, which is an important building block of cell-biological oscillation analysis, can be generated in the simple reaction systems under investigation here.

This investigation has shown that even relatively small protein accumulation/dissociation systems can give rise to complex dynamics. Since the dynamical phenomena bistability and oscillation are important functional patterns in cell biology, these results contribute to the understanding of cellular function in general.

5 Stochastic DNA Replication Model

In this chapter, a model of DNA replication is presented that was developed with the aim of identifying the essential functional components of the DNA replication process in human cells. It it will be shown that it was possible to identify a minimal set of model ingredients that allows the dynamics of the model to reproduce a wide range of experimental evidence. A manuscript based on the text of this chapter will be submitted to a peer-review journal. All experimental data presented here (unless noted otherwise) was produced by the group of Cristina Cardoso at TU Darmstadt, with data contributions by Vadim Chagin, Corella Casas-Delucchi and Marius Reinhart.

5.1 Existing Models

Several key properties have already been identified in the literature as being essential to replication modeling. It was noted in section 2.2 that the timing of origin firing is stochastic in eukaryotic cells, which is therefore a standard component of replication models, be it in yeast [36, 38, 148] or metazoans [149–151]. Flow cytometry images show that the overall speed of DNA replication remains constant for most of S-phase (see figure 5.3). This, however, means that the firing of origins cannot be simple random firing throughout S-phase, because random firing would cause the total replication speed to continuously rise, as more forks become active through firing, and then fall off again when the origins are depleted. Therefore, random firing would result in a peak-like time course of the total replication speed, which is in conflict with the flow-cytometry data. Another mechanism must thus exist that imposes the constant replication speed upon the S-phase dynamics. While there are some reports of differences in replication fork speed between the subphases of S-phase [43, 152], these differences are not large enough to produce a constant overall replication speed for completely stochastic origin firing. A limiting factor has been proposed as a firing-restricting mechanism to alleviate this inconsistency [150].

In models that use a constant replication fork speed, it takes very long to replicate the last stretches of DNA, giving S-phase a tail of steadily decreasing replication activity [36]. In order to ensure a timely end to S-phase, some models increase the origin efficiency over time [150, 153]. This adjustment might be unnecessary, however, because some residual replication has been observed experimentally the G2 phase [53].

Jackson and Pombo [154] showed that replicons are clustered even in early S-phase. A possible explanation for this is the coordinated simultaneous firing of groups of origins, which has been implemented in some recent models [150, 151]. An alternate explanation for the clustered appearance of replicons is a stochastic domino-like mechanism, in which the firing of origins is promoted by the proximity of replication forks [155, 156]. Since the latter mech-

anism is simpler than coordinated firing of origin groups, it could be a more elegant answer to the question of clustered origins. At the time of writing, it is a unique feature of the model presented here that it implements this mechanism without resorting to complicated supporting firing patterns [157].

A model comparison study conducted by Shaw et al [150] suggests that there is no strict temporal separation of R- and G-band replication [150]. Furthermore, the recent extensive study of replication in HeLa cells by Guilbaud et al. [155], shows that the temporal transition regions (i.e. regions that are neither clearly early- nor clearly late-replicating) are also covered by origins which are sequentially activated.

In light of the spatio-temporal flexibility of S-phase structure [31,46] and because metazoan origins can be initiated at many non-predetermined sites [33], many models of genome duplication in metazoans are based on intricate stochastic mechanisms. It follows that such models, while necessarily more elaborate than yeast models, easily become overly complicated. Moreover, due to the large size of metazoan genomes and the complexity of their nuclear structure, models of genome duplication in metazoans often rely on aggregate biological data obtained in separate studies and/or for different cell lines [150, 155, 158]. In contrast, the model presented here strives to isolate the most basic ingredients of metazoan DNA replication and obtains all of its parameters from a coordinated study of a single human (HeLa) cell line [21], supported by publicly available banding data [159].

5.2 Euchromatic and Heterochromatic Zones in the Model

An important observation from experimental replication data is that early replication occurs preferentially in euchromatin, while later replication occurs mostly in heterochromatin. For this reason, a replication model must include the patterning of DNA into zones of different chromatin type [160]. In the model presented here, the genome is conceived as a one-dimensional DNA string with a length of about 10^{10} base pairs (bp), which is characteristic of the HeLa genome [21].

Partitioning of the DNA into chromosomes is implemented by dividing it into sections separated by barriers, which cannot be overcome by replication forks and block induced firing events. In contrast, replication forks can move through boundaries between eu- and heterochromatin zones. Therefore, the zones only differ with respect to their accessibility at the beginning of S-phase.

The sizes and positions of eu- and heterochromatin zones were derived from human genome giemsa [161] band data of the USC Genome Browser project [159]. In order to adjust the model to HeLa cells, extra copies of those chromosomes that are contained more than twice in HeLa cells [162] were added. Abnormal chromosomes were replaced by unaltered copies of their ancestral human chromosome. This resulted in about $N_z = 1400$ zones of euchromatin and heterochromatin, respectively, with a total chromatin content of $4.4 \cdot 10^9$ bp for eu- and $6. \cdot 10^9$ bp for heterochromatin (including centromeres).

5.3 Correlated and Limited Firing of Origins

Replication origins are not associated with specific DNA sequences. Therefore, the potential replication origins are distributed randomly on the DNA at the beginning of a simulation. Assuming that potential origins defined by MCM complexes are on average located every 10-20 Kbp [17, 153], the total number of potential origins is about $N_0 = 5 \cdot 10^5$ for the HeLa genome. Due to their random positioning, the distances between neighboring potential origins follow an exponential distribution.

It is generally accepted that the first firing events (i.e. the initiation of replication) occur in the euchromatin zones. As an idealized scenario, the model presented here initiates replication by firing one random origin in every euchromatin zone at time $t = 0$ of a simulation.

All other firing events are "induced" firing events, i.e., they occur only in the proximity of active replication forks. In the past, various models incorporated fork-induced firing of origins either explicitly [157] or implicitly [163].

Induced firing is implemented by introducing a relative probability density $p_f(d)$ that depends on its distance d to the closest replication fork. The probability density used here, for reasons explained below, decreases with distance from an active fork and is zero in extreme fork proximity:

$$p_f(d) = \begin{cases} 0 & d < d_i, \\ \exp\left(-\frac{d^2}{2\sigma^2}\right) & d_i < d < d_{max}, \\ 0 & d_{max} < d, \end{cases} \tag{5.1}$$

as depicted in figure 5.1. This is a Gaussian distribution with a cutoff at large and small distances. The standard deviation was chosen to be $\sigma = 280$ Kbp. A value of a few hundred Kbp is suggested by the observation that the empirically observed distance between fired origins is of the order of hundreds of Kbp. The value chosen here leads to the best agreement between simulations and data. The cutoff at large distances, $d_{max} = 600$ Kbp, was chosen such that $p_f(d_{max}) = 0.1$ is 10 percent of the maximum probability density. Furthermore, a value d_{max} of this order is in good agreement with the maximum distance between fired origins seen in empirical data and reflects the idea that DNA replication is regulated at the level of ~ 1 Mbp chromatin domains [154, 164].

The cutoff d_i at small distances corresponds to about half the length of looped chromatin domains, which represent fundamental units of chromatin organization [7] and can also be considered as units of decondensation for replication at the level of the 30 nm fiber [2, 8]. The size of chromatin loops was shown to correlate with replicon size [6] and with distances between preferentially activated origins [165, 166]. One possible explanation for this correlation is that when an origin is activated inside a chromatin loop, other potential origins in the loop tend to be passively replicated. For simplicity's sake, this was implemented as a distance-based inhibition in the model. Figure 5.1 shows a schematic of the induced firing process in the model.

The flat S-phase profile evident in cytometry histograms made available by the Cardoso group (see figure 5.3) suggests that the rate of DNA synthesis is approximately constant for

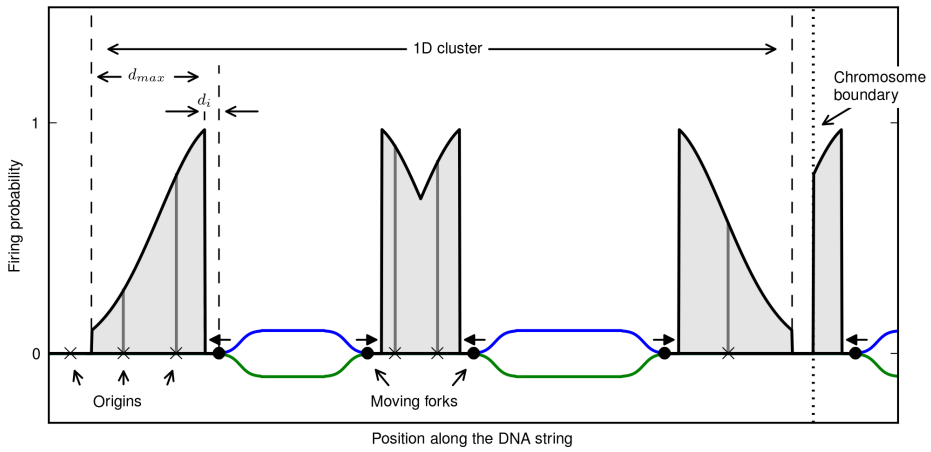

Figure 5.1: **Mechanics of origin firing in the model.** The firing probability of origins that are close to forks follows the probability density defined in equation 5.1, indicated here as shaded areas next to the forks. Firing at positions closer than $d_i = 67$ Kbp to a fork is inhibited and the probability density is cut off at $d_{max} = 600$ Kbp. The relative probabilities of individual origins are indicated by dark grey bars. All four forks to the left of the chromosome boundary belong to a single one-dimensional fork cluster (assuming that neighboring forks are less than 1 Mbp apart). The chromosome boundary near the right edge of the image isolates chromatin belonging to different chromosomes and thus cuts off the induced firing range of the rightmost fork.

most of middle and late S-phase. For this reason, a limit on the total number of replication forks that can be simultaneously active is imposed in the model. This agrees with the suggestion by other researchers [167, 168] that there exist mechanisms that control the progression of S-phase and distribute firing events over a longer period of time. This could be due to a structural program according to which the origins are activated, or due to a component of the replication machinery that is available only in limited numbers. Since it is, at the moment, impossible to discern between these effects, it was decided to limit the rate of DNA synthesis in the model by using a "limiting factor", which is a necessary component that is associated with each replication fork. Such a limiting factor has been used in the past in models of metazoan DNA replication to obtain realistic origin activation profiles and synthesis rates [22, 150]. The assumptions made here are that the limiting factor starts to become available once the cell enters S-phase, and that its number increases during the first hour until it reaches a maximum value L_{max} that is maintained until the end of S-phase. Since the diffusion of small proteins in the cell is very fast [169], it is assumed for the model that the movement of a free limiting factor to an origin, where assembly of the replication machinery could be completed, is instantaneous. As a consequence, the number of replication forks is always identical to the total number of limiting factors in the nucleus.

Experimental data suggest that the total number of replicons is between 6000 and 7000 [21]. A replicon is considered to consist of two forks, meaning that the number of active

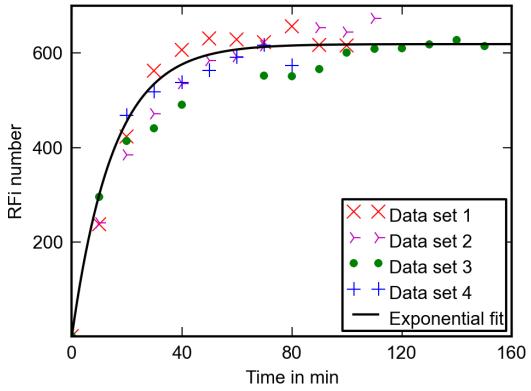

Figure 5.2: **Confocal replication foci (RFi) measurements as an indicator of the initial limiting factor increase.** This RFi data was used in order to obtain a timescale for the initial increase of the limiting factor in the model. The solid line shows equation 5.2 with a time scale of $\tau = 15$ min. Data courtesy of Vadim Chagin.

replication forks is about 12000. Accordingly, in the model, the maximum value of the number of replication forks is set to $L_{max} = 12000$. With this value, the total replication time obtained in the computer simulation agrees with the empirically found replication time.

In order to model the increase of the limiting factor $L(t)$ in the beginning of S-phase, the function

$$L(t) = L_{max}\left(1 - e^{-t/\tau}\right) \tag{5.2}$$

was used, with $\tau = 15min$, as obtained from the increase in replication foci numbers found in measurements (see figure 5.2).

5.4 Reduced Fork Speed During Early S-Phase

When an origin has fired, two replication forks start moving in opposite directions from its position with a fork speed of $v = 28$ bp/s. This is the average fork speed obtained from measurements made in the Cardoso group [21], and it also lies in the range of replication fork speeds given in the literature [2, 152].

A fork moves along the DNA until it collides with another fork that moves in the opposite direction, whereupon both forks annihilate. Consequently, forks do not only appear in pairs but are also removed in pairs, freeing two limiting factors. Forks freely travel from one chromatin type into another, but are stopped at the boundaries between chromosomes, setting a limiting factor free.

In order to make the model complete, it has to be taken into account that the proportion of DNA synthesized in early S-phase is much less than what would be expected from the fact that it lasts for 2.8 hours of S-phase (see figure 5.3). Less than 15% of the DNA is replicated during this time, even though it takes approximately a quarter of S-phase. This can be a consequence

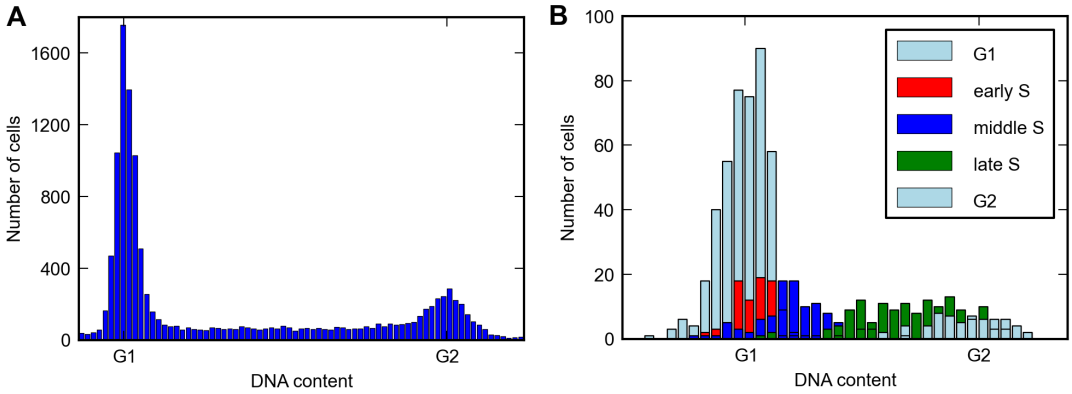

Figure 5.3: **DNA content frequency throughout the cell cycle.** Cells were are binned by DNA content (DAPI signal), with the abscissa showing the DNA content of the bins in arbitrary units. **A** shows the DNA content distribution for 15143 DAPI-stained cells, obtained by flow-cytometry. The distribution remains at an approximately constant value throughout S-phase, i.e., between the G1 and G2 peaks, meaning that the overall speed of replication is constant. Image **B** shows the frequency of specific DNA content intervals in an ensemble of 840 cells, dependent on their cell cycle position. HeLa Kyoto [170] cells expressing fluorescent PCNA were grown, after which they were fixed and stained using DAPI. Through optical inspection of the PCNA signal, the cells were sorted into early, middle and late S-phase, as well as non-S-phase cells. The non-S-phase cells were grouped into G1 and G2 phase cells based on the DAPI signal and one histogram was drawn for each subphase population. It is notable that the number of early S-phase cells drops off steeply at 15 % of the DNA replicated, even though early S-phase lasts for a quarter of S-phase. Data courtesy of Vadim Chagin.

either of the interplay between replication and transcription leading to reduced replication fork speed, or due to transient stalling of replication complexes, which is equivalent to a reduced effective fork speed. The cytometry histograms of figure 5.3 do not show a reduction in early S-phase replication, because the first part of DNA replication is masked by the G1 peak.

To verify that the cause of this effect is equivalent to a reduced effective fork speed, the Cardoso group studied nucleotide incorporation rates for different sub-phases of S-phase, using EdU labeling (see figure 5.4). Nucleotide incorporation was significantly reduced during early S-phase, thus also indicating a reduced effective fork speed. In order to account for this effect in the model, the fork speed was reduced to $v_s = 14\,\text{bp/s}$ (i.e. half its original value) during the first 2.8 hours. In simulations without decreased fork speed during the first 2.8 hours, significantly more heterochromatin is replicated during early S-phase.

Fork speed and maximum value of the limiting factor determine the rate of DNA replication. Thus, except for the start and end of S-phase, the rate of DNA replication is either $12000 \times 14\,\text{bp/s} = 168\,\text{Kbp/s}$ or $12000 \times 28\,\text{bp/s} = 340\,\text{Kbp/s}$. Only 1.4 Gbp are replicated by all forks during the 2.8 hours when the fork speed is reduced. The remaining 9 Gbp are replicated at the full speed, which takes about 7.5 hours, giving a total S-phase duration time of 10.3 hours.

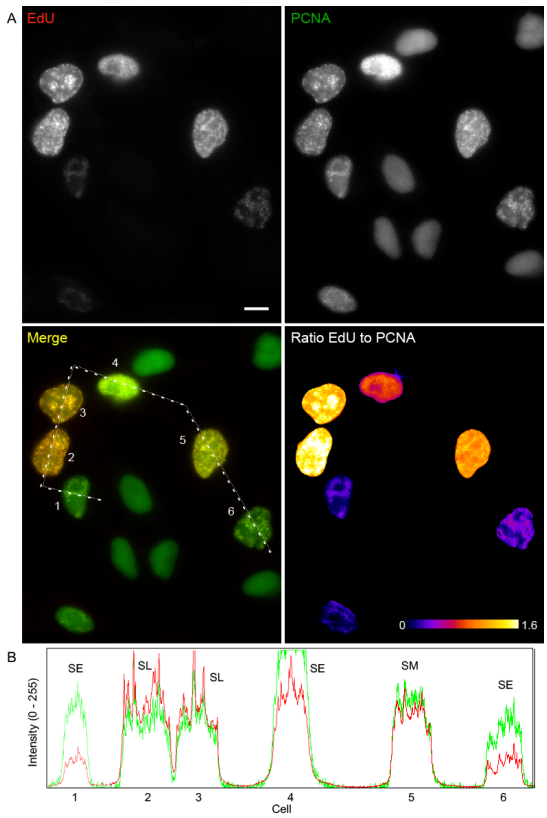

Figure 5.4: **Rate of DNA synthesis at beginning of S-phase is lower than during the rest of S-phase. A** Modified nucleotides (EdU [171]) were introduced into HeLa Kyoto cells expressing mCherry-PCNA for 15 minutes before fixation. The scale bar is 10 µm. **B** Line profiles of the signal intensity of EdU and the PCNA for the six marked cells. The amount of nucleotides incorporated in relation to the recruited PCNA is reduced in the early S-phase cells compared to the middle and late S-phase cells. This indicates that the synthesis rate in early S-phase is lower than during the rest of S-phase. Images and data courtesy of Corella Casas Delucchi.

This is in good agreement with the S-phase duration that was measured for HeLa cells (see figure 5.3).

Together with the constant maximum amount of limiting factor, the above rule for the fork speed leads to a cell cycle profile that is consistent with the experimental data and the constant rate of replication that is observed for most of S-phase [30].

5.5 Simulation Algorithm

The computer simulation of the model was implemented in the C++11 standard of the C++ programming language and can be built and compiled using the GNU toolchain. The package and its installation instructions will be made available online at `www.danielloeb.eu/replication.html` under a free software license (GPLv3) upon publication of the results in a peer-review journal. In addition, a package of data containers for stochastical simulations, developed by the author, was used extensively. It is available under a GPLv3 license at `http://www.danielloeb.eu/simtools.html`.

For the algorithm of the simulation, the replication model is reduced to the task of maintaining multiple sorted lists. The central data structure in the system is the event heap which is a binary heap data structure that at any given time contains all future collision events between the objects that are currently in the system (forks, chromosome barriers, chromatin zone transitions), sorted by time of occurrence. Thus the root element in the heap always holds the next event in the system. In each simulation step, the root element of the heap is removed and time is advanced to its time of occurrence.

If the removal triggers a chromatin zone boundary crossing or a firing event (because a limiting factor has been freed), then the addition and removal of future collision events becomes necessary. In order to keep such operations efficient, ordered lists are maintained for barriers, potential origins, left-going forks and right-going forks. These lists are implemented using a special red-black tree that, in addition to standard red-black tree behavior, allows indexed element access scaling $O(\ln(N))$ with the number of elements N (all nodes keep track of the number of their children).

For instance, if it is determined that an origin has to be fired, a random origin is picked from the available origins and checked if it has been passively replicated by the active forks. If not, its relative firing probability is calculated (a value between 0 and 1) and a random number between 0 and 1 is drawn. Should the random number be lower than the probability, the origin is fired, otherwise the process is repeated. Firing of the origin means that two forks, one in each direction, will be created, which have to be inserted into the fork lists, and for which collision events have to be calculated. Since the positions on the DNA are floating point values in the simulation, this only requires position searches in the list of barriers and in the lists of forks with opposed headings, all of which scale logarithmically, due to the data structures used. Once the earliest collision is found for each fork, they are inserted into the event heap. Because each fork also has a pointer to its future collision, collision events that have become invalid due to new insertions are easily removed from the collision heap.

Various smaller extensions to this algorithm are implemented to cover edge cases (initial origin firing, origins firing with low origin numbers). Whenever a fork is removed from the

system, all origins along its path are removed also. Since the positions of forks are not updated in each time step but only calculated from initial positions and time when needed, output operations that require fork positions are computationally expensive.

5.6 The Four Stages of S-phase Dynamics

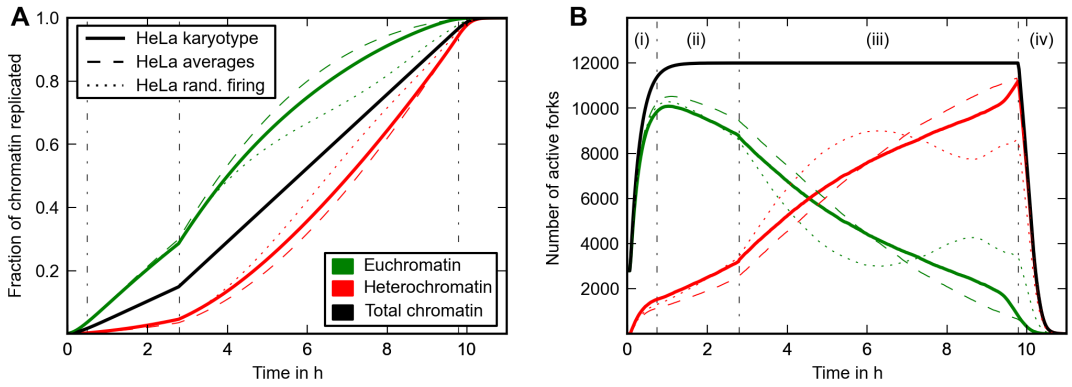

Figure 5.5: **Replication progression for heterochromatin, euchromatin and sum of both.** Solid lines show simulations with one initial firing event in every euchromatin zone and karyotype-based chromatin patterns. The other lines are the result of modified model versions in order to asses the impact of the chromatin pattern and the importance of having one firing event per euchromatin zone (discussed in section 5.7). Dashed lines represent simulations with chromatin zones of identical sizes in each chromatin type, and dotted lines are karyotype-based simulations with initial firing events randomly distributed over euchromatin. **A** shows the fraction of chromatin replicated as function time. Half of euchromatin is replicated during the first 4 hours, whereas most heterochromatin is replicated during the last 3 hours. Figure **B** shows the number of forks in each chromatin type and in total. When the fork speed is increased from 14 bp/s to 28 bp/s at 2.8 hours, fork annihilation and creation occur two times as fast as before.

Figure 5.5 shows the progress of replication in simulations with human-like (HeLa) cell parameters. There are four qualitatively different stages.

(i) **S-phase entry stage** During the first 45 minutes, the number of forks and the replication rate both increase, because the limiting factor increases.

(ii) **Housekeeping replication** For the next 2 hours, during which mostly housekeeping genes are replicated, replication operates at the maximum limiting factor level, but still with the reduced fork speed.

(iii) **Processive S-phase** Because the maximum limiting factor is reached at stage (ii), the number of forks remains constant until the finalizing stage (figure 5.5B). However, the

rate of replication, which depends not only on the number of forks but also on fork speed, jumps by a factor of two when the initially reduced fork speed is set to its regular value of 28 bp/s after 2.8 hours. From then on, the replication rate assumes its maximum value, maintaining it for most of the remainder of S-phase. The transition is visible as a change of slope of the total replication curve in figure 5.5A. Most of the genome is replicated during this stage.

(iv) **Finalizing stage** Shortly before replication is completed, the last remaining potential origins are fired. After that, no origins are left that can fire, but there are still DNA stretches of various lengths that have to be replicated. The number of replication forks decreases steadily during this finalizing stage, but a few of these forks persist for more than an hour.

The average duration of this final stage in the model depends on the number N_0 of potential origins. It is shorter when the number of origins is larger. In simulations with $N_0 = 500000$ origins, it typically lasts between 1 and 2 hours, which is on the same timescale as RFi number decay at the end of S-phase [21] and is consistent with data reported by Widrow et al. [53]. This means that the assumptions used in the model and the number of origins chosen are in agreement with what is found in nature.

5.7 Importance of Euchromatin Zone Sizes and Initial Firing Locations

In order to assess to what extent the simulation results are sensitive to the size distribution of chromatin zones and to the condition of initial firing in every euchromatin zone, two types of control simulations were made. In one version, all euchromatin zones and all heterochromatin zones have the same size (instead of the sizes taken from chromatin data), and in the other, initial firing occurs completely randomly in euchromatin (instead of firing one origin in each euchromatin zone). All other parameters remained unchanged. The results of these simulations are shown as dashed and dotted lines in figures 5.5, 5.7 and 5.6.

With identical zone sizes, less heterochromatin is replicated during the early stages. This is because there are no small euchromatin zones. Due to all euchromatin and heterochromatin zones having the same sizes, the initial distances between groups of forks spawned by initial firing events (i.e. 1D clusters, see below) have a more narrow distribution, increasing the time until they start merging. The total number of origins fired during the entire simulation drops from 49000 to 45000.

The differences between the two types of simulations are not large, since having equal sizes for all zones is not very different from the karyotype-based size distribution. There, about 75% of chromatin zones are between 1 Mbp and 6 Mbp in size. Distributions that are less similar to the karyotype, such as an exponential distribution, lead to larger deviations in replication timing and number of fired origins (see figure 1 in appendix 2). This comparison thus shows that specific chromatin patterns imprint on the replication dynamics but that the temporal structure of S-phase is not very sensitive to the size distribution of eu- and heterochromatin

zones, as long as this distribution has most of its weight between 1 and 6 Mbp.

To assess the importance of having one single initial firing event in each chromatin zone, the author performed simulations where the initial firing occurs at random sites in the euchromatin. The total number of initially firing origins was still identical to the number of euchromatin zones. The results are indicated as dotted lines in figures 5.5, 5.7 and 5.6.

One effect of random initial firing is a less realistic timing of replication of the two chromatin types (see figure 5.5). From hour 7 on, the number of forks in euchromatin actually increases, leading to more euchromatin replicated in late S-phase. The reason for this is that when the initially firing origins are distributed randomly, but their number is identical to that of euchromatic regions, part of the euchromatic regions will not receive any initial firing event. These euchromatic regions will be replicated only after a series of induced firing events has traversed one of the adjacent heterochromatic regions, which may happen late in S-phase.

5.8 Induced Firing and the Inter-Origin Distance Distribution

It is known that the total number N_a of active origins involved in the replication of an entire mammalian genome lies in between 30000 and 50000 [172]. Other authors have shown that there is an excess of available potential replication origins in eukaryotic cells [173, 174]. The most important parameter influencing the number of fired origins in the model presented here is σ, the standard deviation used for the induced firing probability. Because the number of active forks and the rate of replication are both independent of σ, its primary role is to determine the average distances between fired origins. Smaller σ will increase the number of fired origins and larger σ will decrease it. With an inhibition distance of $d_i = 67$ Kbp and a value $\sigma = 280$ Kbp 49000 fired origins are obtained, which is consistent with the experimental average distance between fired origins of 200 Kbp.

Within the parameter range used for all simulations shown here, the total number of available origins does not influence the total number of fired origins. The reason for this is that for the number of available origins to be relevant, there would have to be origin depletion at some point during the simulation. But because the average inter-origin distance of $l/N_0 = 20$ Kbp is significantly smaller than σ, origin depletion is extremely unlikely. As far as the process of induced firing is concerned, a number of $N_0 = 500000$ available origins is indistinguishable from an infinite number of available origins. The only moment where the number of origins plays a role is during the final stage of replication, as mentioned above.

Chromatin size and number of fired origins trivially determine the average distance between origins that have fired (*inter-origin distance*) to be 10^{10} bp/49000 $\approx$ 200 Kbp. Using DNA combing, the Cardoso group obtained data on these distances, seen in figure 5.6, that shows a peak slightly below 200 Kbp and a longer tail that extends to 600 Kbp. These features of the distance distribution of origins are reproduced by the model with the chosen values of σ and d_i. Table 5.1 gives an overview of the complete set of parameters used for the simulations shown here, and shows the rationale for each value.

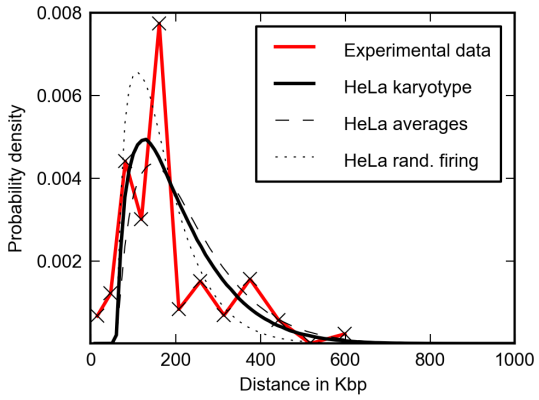

Figure 5.6: **Distribution of distances between adjacent fired origins** DNA combing data on the distances between fired origins for HeLa Kyoto cells indicates a peak close to but below the 200 Kbp mark and a heavy tail extending to the 400 Kbp to 700 Kbp range. The distance distribution obtained from the model calculations does have both of these features. Inhibition of firing for distances below 67 Kbp forces the simulation data to zero in that range. Without inhibition, this part of the distribution would be exponential. The model data is averaged over 100 simulations and there are 50 data points for the experimental data.

5.9 One-Dimensional Replication Fork Clusters

Since all firing events after time $t = 0$ are induced firing events, firing occurs always in the vicinity of active forks, leading to clusters of active forks on the one-dimensional model representation of DNA. The larger such a 1D cluster becomes, the larger is the probability that the next firing event will occur in the proximity of or within this cluster. The size of the cluster increases each time an induced firing event occurs outside the outermost forks of the cluster.

Thus, clustered replication in this model is maintained through individual firing and annihilation events. Other approaches, such as the model by Shaw et al. [150], include an explicit mechanism of clustered origin activation.

Two adjacent forks are considered to belong to the same cluster if their distance is less than 1 Mbp, which is consistent with the distance over which induced firing can occur in the model and the characteristic size of chromatin domains [164]. Clusters can therefore split into two parts that move in opposing directions when large stretches of DNA within them have been replicated. In order to see in more detail how replication activity is distributed over the DNA, the number of clusters and the rate of replication fork annihilation were evaluated as a function of time, see figure 5.7.

For the first three hours of replication, the total number of clusters cannot exceed the number of initially fired origins, i.e. the number of euchromatin zones. The activation of origins follows the limiting factor dynamics during the first hour of S-phase and then stays constant until about the end of the third hour, which is approximately the end of the early S-phase subperiod (see figure 5.3). Figure 5.7 shows significantly less one-dimensional clusters for the random initiation calculations than for those with initiation in every euchromatin zone. Here,

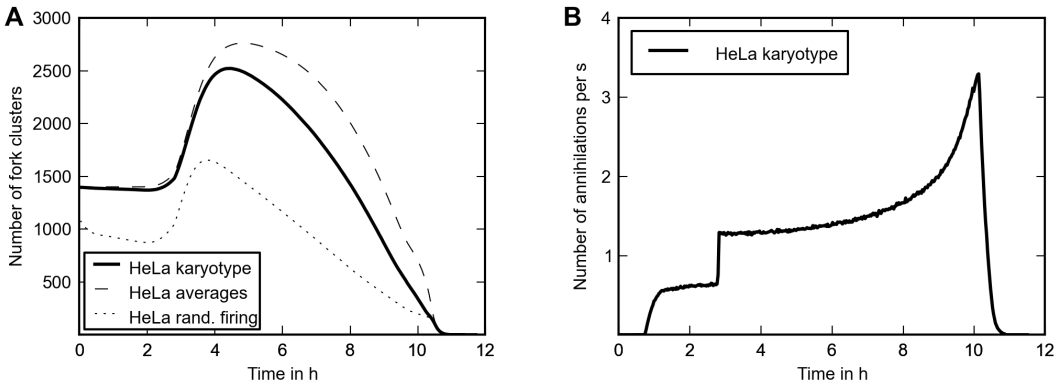

Figure 5.7: **Induced firing and number of annihilation events over time.** Figure **A** shows the number of replication clusters as a function of time, considering all active forks that are less than 1 Mbp apart as belonging to the same cluster. There is very little cluster merging during the initial period of slow dynamics. About 3.5 hours into S-phase, the number of clusters almost doubles, due to clusters splitting into two. Later, the number of clusters declines, as more and more clusters merge. Figure **B** shows the rate of annihilation of forks throughout S-phase. Due to inhibition of firing within distances smaller than d_i of active replication forks, there are no annihilation events during the first $d_i/(2 \cdot v_s) = 40$ minutes. After this period, the rate of annihilations approaches a steady-state value. Due to the doubling in fork speed, the rate of annihilation (and thus also firing) increases by a factor of two at 2.8 hours. As the proportion of replicated DNA becomes large, the clusters become more densely packed with forks. This causes the fork annihilation rate to increase towards the end. Each curve is an average of 100 simulations.

Parameter	**Value**	**Underlying experimental data and consistency arguments**
Genome size	$l \approx 10.4\,\text{Gbp}$	Directly measured in [21]
Number and sizes of eu- and heterochromatin zones	$N_z = 1400$	Giemsa band data from [159]
Number of potential origins	$N_0 = 500000$	Distances between MCM complexes [17, 153], late S-phase duration consistency with [53]
Limiting factor	$L_{max} = 12000$	Double the number of replicons measured in [21], consistency with fork speed and duration of S-phase
Initial limiting factor growth timescale	$\tau = 15\,\text{min}$	Taken from RFi number growth (figure 5.2)
Fork speed	$v = 28\,\text{bp/s}$	Directly measured in [21], consistency with limiting factor and duration of S-phase
Reduced fork speed during first 2.8 hours	$v_s = 14\,\text{bp/s}$	Set so that fraction of the genome replicated during that time stays below 15%, as indicated by the DNA content counts of figure 5.3. This value is also consistent with the nucleotide incorporation measurements (see figure 5.4).
Distance parameter of induced firing	$\sigma = 280\,\text{Kbp}$	Chosen so that model reproduces measured distances between fired origins (figure 5.6)
Distance parameter of firing inhibition	$d_i = 67\,\text{Kbp}$	Chosen so that model reproduces measured distances between fired origins (figure 5.6), value consistent with known size of looped domains

Table 5.1: **All parameters of the model.** For each parameter, the known/measured quantities from which its value is determined are listed. With the exception of σ and d_i, the experimental values for all parameters were inserted into the model a priori.

initiations can happen close together, so that the 1D clusters they spawn are counted only as one cluster. This follows from the distribution of distances between initiation events, which for random initiation positions is an exponential. When the total number of 1D clusters is reduced through the annihilation of 1D clusters that are close together, the fork density increases in the remaining clusters. Accordingly, the total number of activated origins in these simulations increases to 60000.

After the first three hours, clusters start splitting into two, and the number of clusters increases for two hours. The splitting of a cluster occurs when a continuous stretch larger than 1 Mbp within the cluster has been replicated. After hour 6, the merging of clusters becomes

more dominant than the splitting of clusters, and the number of clusters decreases steadily from then on.

The 1D clusters described here cannot be easily identified with replication foci visible microscopically in the experiment. The reason for this is that the 1D clusters are defined along a one-dimensional representation of the DNA, and that including forks with a distance of almost 1 Mbp in the same cluster amounts to using a low resolution. Due to the three-dimensional arrangement of DNA in the nucleus, replication forks that are much farther apart than 1 Mbp may be close to each other in the experiment, and due to the high resolution of the experiment, replication forks that are closer to each other than 1 Mbp can often be separated. Only during late S-phase, where replication forks form few and large foci, can the 1D clusters be related to replication foci (see section 5.11).

The rate of fork annihilation is zero during the first 80 minutes, because there is inhibition of firing close to forks. Only after the minimum distance d_i has been covered by two forks running towards each other, can the first forks annihilate (see figure 5.7 B). After this period, the rate of annihilations approaches a steady-state value, which jumps by a factor of two at 2.8 hours, when fork speed is increased. All these transitions are barely visible in the number of 1D clusters. On the other hand, the increase in the number of clusters at 3.5 hours (see figure 5.7 A) is not visible in the fork annihilation rate (shown in figure 5.7 B). The fork annihilation rate increases first slowly, and then increasingly fast, because the number of replication clusters and, during the last hour, also their size, decreases, leading to a larger fork density within these clusters.

5.10 Comparison to Replication Timing Measurements

In the recent literature [30, 175, 176], experimental data on the replication timing of all regions of specific human chromosomes has been presented. The author performed an analogous evaluation of replication timing for the simulations presented here, shown in Figure 5.8 for chromosome 6, which also includes the microarray data of figure 2 A (and S5 A supplementary data) in [30].

In order to mimic the situation in the experiments, the points in time at which DNA at the experimental sampling positions is replicated are taken from the model. Figure 5.8 shows that the results obtained from the simulations resemble the experimental data, with a correlation coefficient of 0.36 between the two data sets. Both theory and experiment exhibit distinctive peaks due to early replication in the euchromatic zones, including the smallest euchromatin zones. The presence of these peaks in the experiment indicates that there indeed are early firing events in all euchromatic zones, regardless of their size. If initial firing were to happen at random positions, early firing in small euchromatic zones would be rare.

While general agreement between model and experiment is observed, the positions of peaks relative to the chromatin zones are clearly more asymmetric in the experimental averages than in the simulation averages. This can be attributed in part to the statistics of the experiment, since every experimental data point is an average over only four measurements. Based on the known stochasticity of initiation events at the single cell level, individual measurements are expected to show variation in the peak positions. The same effect is visible in the simulation

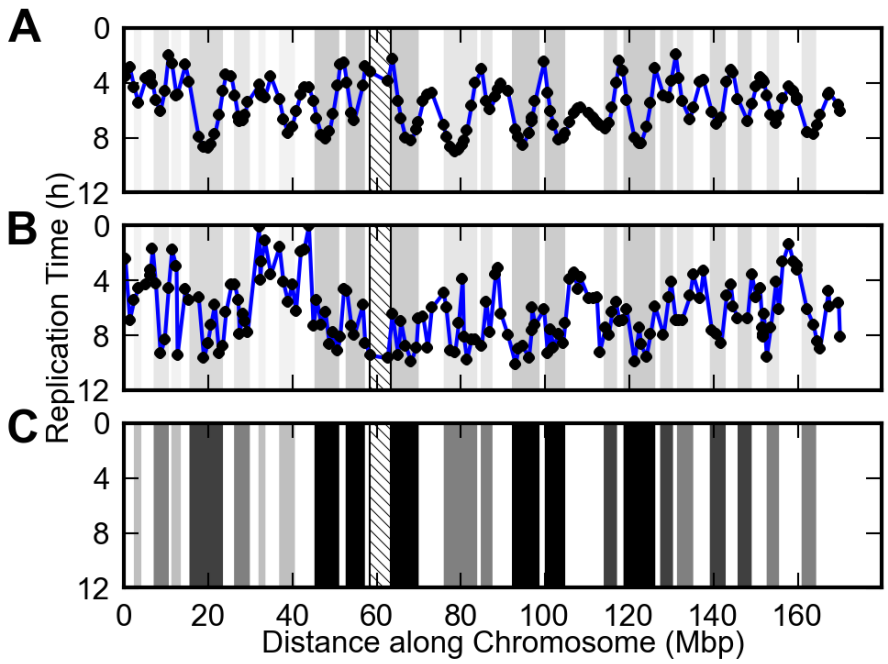

Figure 5.8: **Comparison of the model with replication timing data for chromosome 6 from Woodfine et al..** **A** Replication timing for human chromosome 6 in the model, averaged over 100 simulations. Sampling positions are identical to the positions in the experimental data [30]. For individual calculations, the euchromatic peaks start at time zero, but because of averaging and the sampling interval size, the peaks in the graph are less extreme. The time axis of this diagram is inverted, as is the convention in replication timing diagrams. **B** Microarray replication timing data of human chromosome 6 at 1 Mbp resolution that was published by Woodfine et al. [30]. All data points are averages of four measurements. The Pearson's correlation coefficient between the theoretical and experimental data shown here is 0.36. **C** Giemsa staining pattern [159] for chromosome 6, where white regions are euchromatic and shaded regions heterochromatic. The centromere is indicated as a striped pattern. Analogous figures for other human chromosomes can be found in appendix 2.

results. When a large number of data sets are averaged, the results follow a more regular pattern: Centers of euchromatic regions are on average replicated first and the centers of heterochromatic regions are replicated last.

However, from the simulation results the author found that this effect alone only reduces the correlation to $0.8 - 0.9$. There are two additional effects that lead to a total correlation of about 0.3. On the one hand, the experimental data contains numerous data points with completely different values than their immediate neighbors (for example, the sudden heterochromatic spikes at 65.7 Mbp, 80.4 Mbp and 96.9 Mbp in figure 7B), which could also be alleviated by increased experiment statistics. On the other hand, the experimental data shows that groups of contiguous chromatin zones are collectively replicated earlier or later than others. For instance in figure 7B, in the area between 25 Mbp and 45 Mbp both eu- and heterochromatin are replicated much earlier than in the area between 70 Mbp and 90 Mbp. This effect could be attributed to the influence of the three-dimensional chromatin arrangement, which is not included in the model. Comparative images similar to figure 7 for 23 human chromosomes can be found in appendix 2.

The slope of the replication timing curves is determined by the progression of induced firing and can be estimated by the following considerations. After initial firing, a one-dimensional replication cluster starts expanding in each euchromatin zone. Once the limiting factor has reached its stationary value of $L_{max} = 12000$, the average amount of DNA replicated within each cluster per unit time is given by vL_{max}/N_z, with N_z being the number of euchromatic zones. This means that the two fronts of a cluster each move with a "*wave speed*" of

$$v_w = \frac{v \cdot L_{max}}{2N_z}. \tag{5.3}$$

During the first 2.8 hours with $v = v_s = 14$ bp/s, this has a value of about 60 bp/s, while it becomes 120 bp/s once the fork speed has been reset to $v = 28$ bp/s and then progressively increases as the number of 1D clusters declines. This latter number matches the slopes of replication timing measurements in the literature [30, 175].

5.11 Pseudo-Microscopy Images and Late S-Phase Foci

It is known from fluorescence microscopy [177] that there are distinct patterns in the three-dimensional arrangement and size of replication foci for each of the sub-phases of S-phase. To compare the dynamics of the one-dimensional replication clusters in the model with the experimentally observed characteristics of replication foci, pseudo-microscopy images of the model results were generated. To this purpose, the author created a Monte Carlo simulation based on the random loop model for long polymers by Bohn et al. [178], which has already been successfully used to describe chromatin folding of human DNA [10].

In the random loop model, a polymer (i.e. the DNA) is approximated as a chain of beads with harmonic springs between adjacent beads (*Gaussian chain*). Non-adjacent beads are linked randomly, such that loops are generated at an average incidence of 5 loops per 10 Mbp. Because this random linking generates loops on all size scales (i.e. possibly connecting any

two positions on a chromosome), they serve to restrict chromosomes to the limited volume associated with chromosomal regions.

For the Monte Carlo implementation of the model, the potential for a Gaussian chain with N_{beads} beads is:

$$U_{Gauss} = \sum_{i=1}^{N_{beads}-1} \frac{\kappa_i}{2} ||\mathbf{x_i} - \mathbf{x_{i+1}}||^2, \tag{5.4}$$

with the spring constant κ_i here being $5 \cdot 10^{-7}$ for heterochromatin and $1 \cdot 10^{-8}$ for euchromatin. HeLa karyotype [162] and Giemsa staining [159, 161] data was used to generate the bead chains for the chromatin zones of all chromosomes and to ensure that no connections between beads of different chromosomes are made. Random loop connections within chromosomes give an additional potential term:

$$U_{Loop} = \sum_{\substack{i=1 \\ k_i, j_i \in [0, N_{beads}]}}^{5000} \frac{\kappa_L}{2} ||\mathbf{x_{k_i}} - \mathbf{x_{j_i}}||^2, \tag{5.5}$$

where the total number of 5000 connections is based on a comparison of random loop model results with experimental genomic distance data by Mateos-Langerak et al. [10]. For the average loop size, a value of 2 Mbp was chosen, which is towards the low end of their loop size estimate. The spring constant for these links is $\kappa_L = 5 \cdot 10^{-7}$.

Since no cellular scaffolding or membrane interactions are included in this model, it is necessary to implement a small repulsive force between chromosomes to ensure that each chromosome has its own nuclear territory. For this a reversed gravity approach was used:

$$U_{Rep} = \sum_{\substack{\forall m \neq n \\ m,n \in \text{chromosomes}}} \kappa_R \frac{W_m \cdot W_n}{\sqrt{||\mathbf{x_m} - \mathbf{x_n}||^2}}. \tag{5.6}$$

Here, vectors $\mathbf{x_m}$ and $\mathbf{x_n}$ are the center positions of chromosomes m and n, W_m and W_n are the chromosome weights (i.e. number of beads). In all simulations presented here, $\kappa_R = 2 \cdot 10^{-4}$ was used, which means that the per-bead contribution of the repulsive potential is significantly smaller than the contribution of the bead connection potential. For the total potential, the three terms are added together:

$$U = U_{Gauss} + U_{Loop} + U_{Rep}. \tag{5.7}$$

Movement of beads is restricted to an oblate ellipsoid with a vertical semi-axis of 3.5 µm and two horizontal semi-axes of 7.5 µm and 5 µm. The standard Metropolis algorithm was used to let the beads relax into equilibrium with a temperature reservoir at 290 K. Replication fork positions from the 1D replication model are then mapped onto the chromatin, thus generating a coordinate in three dimensions for each fork.

When the rationale for origin firing inhibition was laid out in section 5.3, the argument was based on looped domains on a $d_i = 67$ Kbp (equals half a domain) scale. Since the inter-bead distance used in the random loop model simulations is 100 Kbp, these domains are not

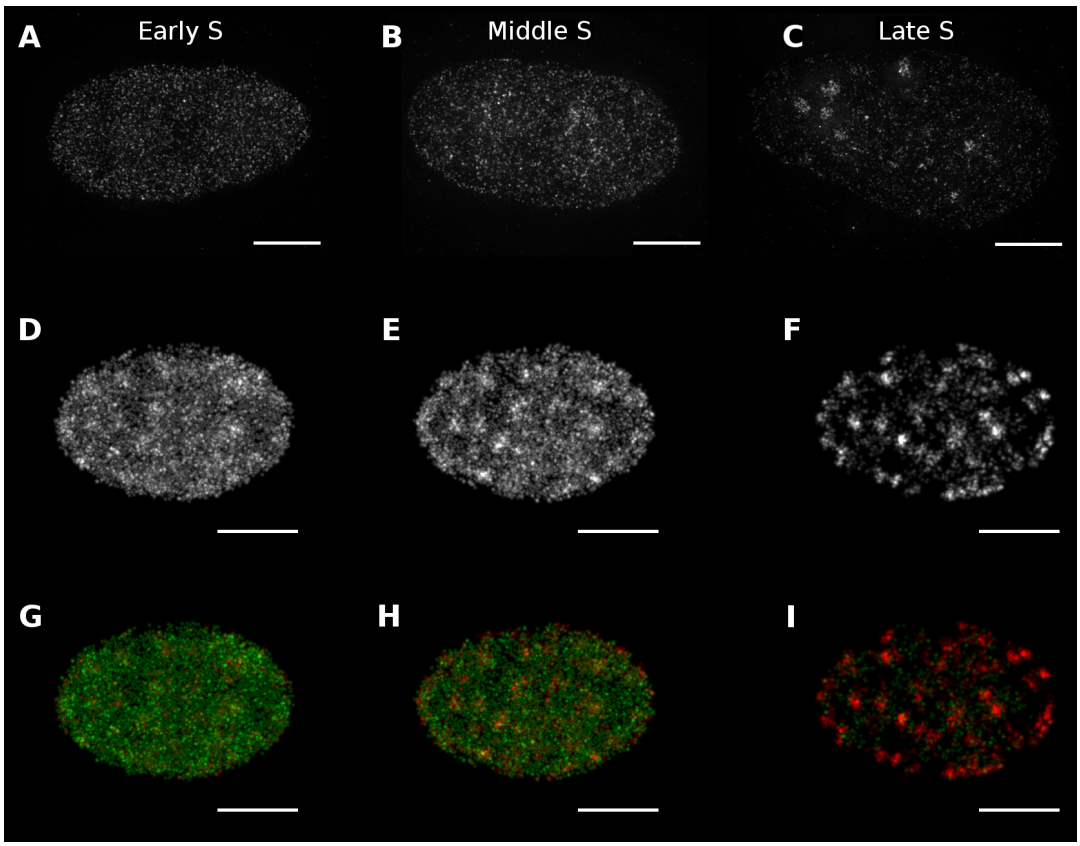

Figure 5.9: **Comparison between the microscopy pattern of GFP-tagged PCNA during replication in experiment and model.** Figures **A-C** show microscopy patterns of replication in HeLa cells and **D-I** the corresponding pattern of the replication model results on a 3D DNA conformation calculated using the random loop model. All simulation figures were generated from the same 3D DNA conformation and the S-phase times for the simulation images are 2 h 30 min (early S), 5 h (middle S) and 9 h 45 min (late S). In this Monte Carlo calculation the bead size was 10^5 bp per bead, the spring constants were 10^{-8} and $5 \cdot 10^{-7}$ for eu- and heterochromatin, the pseudo-gravity parameter was $2 \cdot 10^{-4}$, the equilibration temperature was 290 K and there were 5000 loop connections (i.e the average loop size was 2 Mbp). An ellipsoid nucleus with the axis dimensions 15 µm, 10 µm and 7 µm was used (estimation based on maximum thickness measurement, see figure 3 in appendix 2). Experimental figures (**A-C**) courtesy of Vadim Chagin, with post-processing by Corella Casas Delucchi.

resolved in the Monte Carlo model results and should not be confused with the loops of the random loop model. These latter loops, which have an average size of 2 Mbp, participate in the higher order chromatin organization.

In a previous study of human DNA by Mateos-Langerak et al. [10], different linking probabilities were used to model differences in displacement for transcriptionally active and inactive regions. However, in using such linking probability variations for eu- and heterochromatin, the author noticed that the distribution of both chromatin types over the nuclear volume was highly uneven (see figure 2 in appendix 2), meaning that euchromatin also strongly co-localized with the densely connected heterochromatin. Therefore, reflecting the different degrees of compaction for the two chromatin types, different spring constants for both chromatin types and random linking were used in the model presented here instead.

Figures 5.9 (**A-C**) show fluorescence microscopy images for the three sub-phases; figures 5.9 (**D-F**) show fork positions generated by the replication model for early, middle and late S-phase, marked at their corresponding position in a 3D DNA conformation calculated using the random loop model.

Microscopy images of early S-phase show a large number of small and evenly distributed replication foci in the entire nuclear volume. In the simulations, in early S-phase most forks (and indeed most 1D fork clusters) are in euchromatin, which for each chromosome is spread out over several cubic micrometers. The typical size of a one-dimensional fork cluster in euchromatin is so large that it is not possible to connect its member forks visually, giving the fork distribution a seemingly random pattern which resembles that of early S-phase microscopy images.

Experimentally, a small number of large and bright foci are observed during late S-phase. In the replication model, forks during late S-phase are located primarily in heterochromatin, which in the random loop model is constrained to a small volume for each chromosome. The large foci in the simulation images therefore appear when the one-dimensional replication clusters are concentrated in the heterochromatin zones of a chromosome, thus effectively forming 3D clusters. The high density of replication forks within one-dimensional clusters during late S-phase amplifies this effect.

The arrangement of foci at the nuclear and nucleolar periphery observed during middle S-phase is not reproduced by the model. This is expected, since interactions with the cellular scaffolding and crowding out due to newly replicated DNA are not part of the random loop model.

Although a steady increase in the size of replication clusters was observed in the 1D model, these 3D results show that for the microscopy pattern of replication foci it is more important what chromatin type the forks are in. This is illustrated by figures 5.9 (**G-I**), which highlight the chromatin types at the fork positions in figures 5.9 (**D-F**). The size of 1D clusters also has an effect, but it only becomes relevant in the last hour of replication, when it increases the intensity and localization of replication foci.

5.12 Discussion

In this chapter, a model for DNA replication that reproduces the characteristics of replication in human cells was introduced. The model consists of a minimal set of components, all of which are biologically motivated and whose parameters are determined from experimental data in HeLa.

A central ingredient of the model is the presence of a limiting factor that limits the total number of replication forks. Other authors [22, 150] already established that a limiting factor is needed to obtain realistic origin activation profiles and synthesis rates in models of mammalian DNA replication. The initial increase of the limiting factor was deduced from confocal RFi measurements performed in the Cardoso group and implemented in the model using the measured time constant of 15 minutes. After the first hour, the limiting factor was kept at the constant value 12000, which agrees with a replicon count of 6000 to 7000 (see [21]). The same number of limiting factors can be obtained by calculating the total number of replication forks based on the duration of S-phase, the size of the genome, and the fork speed obtained from the experimental characterization of HeLa [21]. This means that the limiting factor is fixed by two consistent experimental measurements. Using a constant limiting factor has the advantage that it is simpler than other approaches, which require a growing limiting factor [22, 157] or a time-dependent firing rate [179, 180] to control the replication rate.

Unlike previous models [163, 179], the model presented here explicitly uses the specific chromatin layout (i.e. chromosome sizes and pattern of eu- and heterochromatin zones) of human cells by modeling each HeLa chromosome as an instance of the corresponding human chromosome. However, the results of section 5.7 show that not all details matter for the timing of eu- and heterochromatin replication and the number of 1D replication clusters, as long as the distribution of chromatin zone sizes has most of its weight between 1 and 6 Mbp. It follows that the model is consistent with the observation that the overall replication program is resilient to the presence of chromosomal abnormalities.

Another assumption of the model is that the initiation of replication at the beginning of S-phase happens in all euchromatin zones. While good accessibility of euchromatin suggests that replication should start there, there was no indication a-priori of whether it starts in each euchromatic region or at positions selected completely at random from all euchromatin. The computer simulations of the model show that this second scenario leads to a considerable amount of euchromatin being replicated during late S-phase, which has not been observed in nature [47, 168]. As additional confirmation, early firing of euchromatic zones independent of size was also observed in a genome wide replication timing study by Woodfine et al. [30]. One hypothesis that could explain this, is that these initial firing events in euchromatin happen at transcriptionally active sites.

While it is generally believed that the number of potential origins is approximately one order of magnitude larger than the number of fired origins, this number cannot be measured. In the calculations shown here, this value was fixed at 500000, based on the criteria that the number of potential origins should not be larger than the minimum required for consistency with experiments, and that the final stage of S-phase should last no longer than 1-2 hours. Such a duration is in agreement with results by Widrow et al. [53] on the persistence of replicative synthesis at the end of S-phase. This final stage is most sensitive to the number of potential

origins, since its beginning is identical to the point in time when the system runs out of origins, which is directly determined by origin density.

The centerpiece of the replication model is the domino-like effect of firing of origins occurring in the proximity of active replication forks. Recruitment of origin activation kinase Cdk2 by fork component Cdc45, which has been observed in Chinese hamster ovary cells [181], is considered a likely candidate for the underlying biological process of induced firing in mammalian cells [22]. In the model shown here, induced firing is characterized by two parameters, which correspond to the characteristic distance over which firing is induced and to the distance over which firing is inhibited near forks. These two parameters are fixed based on the experimental distribution of distances between fired origins (figure 5.6), which has a steep increase at small distances and a long tail extending over several 100 Kbp, and comprises the total number of about 49000 fired origins. This means that only one out of ten potential origins fires. The resulting value of 67 Kbp for the inhibition distance corresponds approximately to half a chromatin loop. A biological rationale for inhibition of firing within short distances of replication forks is that it saves the costs of assembling replication machinery at sites that would soon be visited by an already active replication fork. The process of induced firing leads to replication "fronts" moving along the DNA much faster than individual replication forks. It was shown in section 5.10 that the speed of these fronts obtained in the model agrees with the experimental literature on replication timing [30, 175].

The values of the fork speed and the distances between fired origins used in the model agree with the values reported by other researchers in the literature [2, 152, 172]. Interestingly however, there is some disagreement with the recent and quite extensive study of replication in HeLa cells by Guilbaud et al. [155]. They report fork speeds that are about half of the value used here and an inter-origin distance that is significantly lower than established values (only 40 Kbp). In their distribution of distances between fired origins, the drop off for small distances happens at a significantly lower value. Taken together, these differences also imply a significantly higher number of fired origins (about 250000). Considering the extent of these differences, clonal variation, as suggested by the authors of the study, is indicated as a likely cause. This observation also highlights the importance of cell line context for HeLa DNA replication data.

A final important ingredient of the model is a reduced initial fork speed. While the measured value of 28 bp/s was used for most of S-phase, the fork speed was lowered to half of this value during the first 2.8 hours, based on experimental evidence that less than 10 percent of the DNA is replicated during this time. This modification leads also to a smaller amount of heterochromatin replicated during early S-phase. Although one can expect that the change from slow to fast fork speed in reality is smoother than in the model, this simple rule is a reasonable approximation. There are two likely candidate processes that could cause such an initial slowdown in replication. Either there is temporary depletion of a replication component (such as nucleotide scarcity), or there are a significant number of forks that are stalled, possibly due to unfinished transcription processes. Macroscopically, the consequences of stalled forks are identical to those of an on average lower effective fork speed.

More generally, introducing stalled forks into the model would not lead to a qualitative change of the results, since the moving fronts of induced firing, which propagate much faster

than the fork speed, are not hindered by stalled forks. Upon encountering a stalled fork, a front will nevertheless induce firing events further ahead, thus "rolling over" the stalled fork and ensuring that its vicinity is replicated by other forks.

The random positioning of origins within euchromatin, as expected, resulted in a certain stochasticity of the replication order. While euchromatic regions are on average replicated early and heterochromatic regions are on average replicated late, the exact time at which a specific position on the chromatin is replicated varies between individual simulations. This agrees well with the observation that differences in the replication timing of otherwise identical cells are possible [151, 182]. It is also consistent with results of the "flexible replicon" model by Cayrou et al. [151], who showed that randomly activating origins and then silencing origins in the vicinity of an activation event are necessary to produce a realistic distribution of distances between fired origins. However, in contrast to the flexible replicon model, the model presented here does not need a-priori grouping of origins into clusters.

In order see how the results of the one-dimensional replication model relate to the characteristic foci patterns observed in fluorescence microscopy, fork positions derived from the replication model were arranged on a three-dimensional chromatin conformation that was obtained using the random loop polymer model [178]. It was thus shown that the model, which is based on the microscopic properties of DNA replication, is consistent with the macroscopic effects observed in microscopy. Unlike in the model of Mateos-Langerak et al. [10], the choice was made not to rely on higher order chromatin organization to distinguish eu- and heterochromatin in the context of the random loop model, but instead to base the analysis on the assumption of different chromatin compactions for different chromatin types. A realistic three-dimensional representation of metazoan chromatin must show a noticeably more compact distribution of heterochromatin (compared to euchromatin) while mostly filling out the nuclear regions with chromatin [183–185]. The reason why a different approach to the random loop model was used here, is that it is not possible to fulfill both these criteria with different looping probabilities alone.

Because the distance between adjacent beads in the random loop model is typically on the order of 0.1 Mbp, the spring constant of the "spring" connecting them must contain all information on chromatin packing below the 0.1 Mbp scale. It follows that, if heterochromatin is packed more densely than euchromatin, the spring constant in heterochromatin must be different from that in euchromatin (i.e. bigger). Combining different spring constants with truly random linking indeed results in chromosomes that consist of a dense heterochromatic region with a wider nuclear region containing primarily euchromatin.

In the comparison of fluorescent microscopy images with these simulated images (see figure 5.9), there is qualitative agreement between experiment and theory. Especially the distributed foci pattern of early S-phase and the concentrated large foci of late S-phase are reproduced well. In chapter 5.11, the compaction of late replicating chromatin was identified as the main driver behind the formation of large foci in late S-phase. Lacking explicit information on nuclear scaffolding, the model cannot reproduce the attachment of heterochromatin to the nuclear membrane observed in the middle of S-phase.

In conclusion, it was possible to achieve a good level of consistency with experimental data by using only a one-dimensional replication model that does not contain any three-dimensional

mechanisms of replication organization. A fully three-dimensional model of metazoan replication might produce further quantitative improvements. Nonetheless, it will be a challenge to construct such a model without sacrificing simplicity.

6 Conclusion and Outlook

From the theoretical analyses performed in the course of this thesis, several biologically significant results were obtained. One is the identification of an essential protein interaction network in the pathway-independent double strand break response. Experiments with heavy ions had shown that recruitment of the DSB response protein NBS1 qualitatively changes its dynamics beyond a certain DSB density. In the lower range of damage densities, the recruitment speed was proportional to the damage density, whereas it remained constant at high damage densities.

To analyze this behavior, a minimal protein recruitment model was composed from the key proteins MRN (a complex including NBS1), ATM, MDC1 and the histone H2AX. With this model, it was possible to fit all available NBS1 recruitment data sets, using only a single set of parameters and the respective damage densities. Binding of the MRN complex, both directly to the DSB ends and to MDC1 recruited in the wider DSB vicinity, proved to be central in understanding the aforementioned change in recruitment dynamics. While the number of binding sites directly at the DSB ends is proportional to the number of DSBs, the number of binding sites in the surrounding DNA is approximately constant. It was thus shown that the observed qualitative difference in NBS1 recruitment corresponds to a shift away from the dynamics of binding in the focus vicinity, which dominates at low DSB densities, towards the dynamics of binding at the damage site, which is more important at high damage densities.

The DSB response model that was thus obtained is an ideal baseline for future, more extensive models. It could be enlarged to include more pathway-dependent response proteins, such as the NHEJ proteins DNA-PKcs and KU70/80. Alternatively, depending on the investigative goal, end processing or downstream proteins might be added. A qualitatively different approach might be a mixed diffusion/reaction model that could serve to improve agreement with the data of slow-diffusing proteins, such as MDC1, beyond the results of the mass-action model presented here.

Generic protein interaction networks consisting only of association and dissociation reactions were analyzed in order to find the prerequisites for such networks to show complex dynamics. Based on mass-action kinetics and the conservation of a set of elementary proteins, it was found that multistability and oscillations can only appear if at least four elementary proteins are present in the system. Starting from the general requirement of a positive feedback loop for multistability [84,130], a minimal multistable protein interaction network was constructed. It was necessary to introduce buffer species in order to maintain species conservation while allowing the concentration changes required by the feedback loop. Only with four elementary protein species was the total number of protein species sufficiently large to make this possible. In addition, it was possible to achieve an oscillation in a closely related reaction system.

A more mathematically rigorous proof that systems with only three elementary species cannot show multistability was derived using the deficiency one algorithm of chemical reaction network theory [135].

In light of the significance of bistability for cell-biological decision making and the ubiquity of protein aggregation in cellular systems, the identification of prerequisites for bistability in such systems is an important result. Furthermore, it has to be noted that many of the typical reaction patterns that give rise to multistability and switching are just more complicated association/dissociation networks. One example is the Michaelis-Menten function term, a typical building block of complex reaction dynamics. An instance of this function term was reproduced in a reaction network with three elementary species and it was discussed how the related highly nonlinear Hill-terms could be constructed in an association/dissociation network with more species. In summary, this analysis showed that many building blocks of protein dynamics theory are rooted in networks of the simplest chemical reactions and that the biologically important phenomena multistability and oscillation can emerge from a sufficiently large but still small system of these simple reactions.

While the DNA damage response serves to maintain the genome integrity to such a degree that the (somatic) cell produces viable daughter cells, DNA replication produces the second copy of the genome that is needed during cellular reproduction. It was thus another goal of this thesis to investigate the genome-scale program of DNA replication in human cells. To this end, a stochastic computer model of DNA replication was developed on the basis of several essential model ingredients that were identified. It was found that a model of human DNA replication which reliably reproduces the observed organization of replication on the chromosome scale but is microscopically stochastic requires the following ingredients: A fast-diffusing limiting factor, induced firing of origins close to forks, inhibition of origin firing in the immediate fork vicinity, reduced nucleotide incorporation (i.e., fork speed) in early S-phase and a restriction of initial firing events to euchromatic chromosome regions.

The replication model presented here improved upon the existing models in the literature not only by restricting the model ingredients to a small set of simple mechanisms, but also by showing qualitative and quantitative agreement with multiple independent experimental data sets. Good agreement was achieved for the measured distance distribution between adjacent fired origins and literature values of replication progression speeds, and the model showed consistency with replication band timing (early/late replication of eu/heterochromatin) data from the literature. Qualitative agreement with known replication microscopy patterns was achieved by mapping the one-dimensional replication model results on a three-dimensional DNA conformation obtained from a Monte Carlo model. These latter pseudo-microscopy images showed that the decisive factor for the formation of large replication clusters in late S-phase is the condensed state of the late-replicating heterochromatin.

Future prospects for the replication model could principally lie in further comparisons with new experimental data, for humans or also for other mammals. The author expects that especially the use of comprehensive data sets for other mammals could help to elucidate the differences and similarities between the replication timing program in the different species. For this task, the model can be directly applied, since the general organization of replication

is considered to be similar in humans and other mammals. However, the existing differences between species, such as genome size, S-phase duration and chromatin organization, must be reflected in the respective model parameters and results. The primary limitation of the replication model is that it does not include any three-dimensional structural elements that might be important during replication. Truly bringing replication modeling into the third dimension will, however, require much more complicated models.

The results and analysis presented in this thesis highlight that theoretical considerations, be they analytical or numerical, are essential to the development of reliable models of cell-biological processes. This is especially true in the realm of complex dynamical effects, where intuitive reasoning by experiment alone can be treacherous. For instance, the NBS1 recruitment data of the heavy ion experiments could easily mislead the observer into assuming that some sort of complex switching mechanism is involved, while the mathematical model unambiguously shows that the observed effect is caused by the different scaling behavior of two types of binding.

Theoretical models can also clarify the question of data consistency between measurements of experimentally independent variables that are connected through an underlying biological process. Often, in such cases, it is not obvious from the data alone, whether or not all experiments are consistent and can be explained by proposed mechanisms. As an example, this was the case with the existing replication progression speed measurements from the literature, where the replication fork speed was determined to be approximately 28 bp/s, but chromosome-scale timing measurements indicated replication speeds about five times larger. The simulations presented here showed that this was neither a contradiction nor caused by a new/unknown regulation mechanism, but that instead, the collective speed of clustered replicons is directly determined by the fork speed and the restrictions of the overall replication timing.

Given the scope of cell-biology as it presents itself today, the author of a PhD thesis can only ever hope to make a small incremental contribution to the overall progress in the field. It this spirit, the work presented here, as a part of an ongoing scientific endeavor, advanced the understanding of some important functional elements of the cell machinery. Even though it is clear that many more such efforts will be necessary, the author of this thesis hopes that a cohesive map of the essential functional structure of the cell will emerge at some point in the not too distant future.

Acknowledgements

First and foremost, I want to thank Barbara Drossel for taking me into her group and for the continuous support and guidance that made this dissertation possible. I am deeply grateful towards my wife Christine Jamin, my daughter Tarja and my parents Gisela and Harry for the support throughout the last years. I also want to thank the current and former members of Barbara Drossel's group, especially Laurin Lengert and Tiago Peixoto, for the many fruitful discussions and positive atmosphere during my time in the group. I also much appreciate the many productive discussions and delicious lunches with the members of the Cardoso lab.

I am grateful to all who contributed corrections and suggestions for improvement to the thesis text or parts thereof: Barbara Drossel, Laurin Lengert, Christopher Priester, Lotta Heckmann, Frank Tobias, Burkhard Jakob, Vadim Chagin, Corella Casas Delucchi and Nicor Lengert.

Finally, I want to thank Michael Ditter for many things, some of which have to do with the Roman Empire.

Bibliography

[1] Newman, M. E. J. Complex systems: A survey. *Arxiv Preprint* **arXiv:1112.1440** (2011).

[2] Alberts, B. *et al. Molecular Biology of the Cell* (Taylor & Francis, 2007), 5 edn.

[3] Gilbert, W. Origin of life: The rna world. *nature* **319**, 618 (1986).

[4] Watson, J. D. & Crick, F. H. C. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**, 737–738 (1953).

[5] Olins, A. L. & Olins, D. E. Spheroid chromatin units (v bodies). *Science* **183**, 330–332 (1974).

[6] Buongiorno-Nardelli, M., Micheli, G., Carri, M. T. & Marilley, M. A relationship between replicon size and supercoiled loop domains in the eukaryotic genome. *Nature* **298**, 100–102 (1982).

[7] Iarovaia, O. V., Bystritskiy, A., Ravcheev, D., Hancock, R. & Razin, S. V. Visualization of individual dna loops and a map of loop domains in the human dystrophin gene. *Nucleic Acids Res* **32**, 2079–2086 (2004). URL `http://dx.doi.org/10.1093/nar/gkh532`.

[8] Fajkus, J., Nicklas, J. A. & Hancock, R. Dna loop domains in a 1.4-mb region around the human hprt gene mapped by cleavage mediated by nuclear matrix-associated topoisomerase ii. *Mol Gen Genet* **260**, 410–416 (1998).

[9] Yokota, H., van den Engh, G., Hearst, J. E., Sachs, R. K. & Trask, B. J. Evidence for the organization of chromatin megabase pair-sized loops arranged along a random walk path in the human g0/g1 interphase nucleus. *The Journal of Cell Biology* **130**, 1239–1249 (1995).

[10] Mateos-Langerak, J. *et al.* Spatially confined folding of chromatin in the interphase nucleus. *Proc Natl Acad Sci U S A* **106**, 3812–3817 (2009). URL `http://dx.doi.org/10.1073/pnas.0809501106`.

[11] Lewin, B. *Genes IX* (Jones and Bartlett, 2008).

[12] Russo, V. E. A. *Epigenetic mechanisms of gene regulation* (Cold Spring Harbor Laboratory Press, Plainview, N.Y, 1996).

[13] Holliday, R. Mechanisms for the control of gene activity during development. *Biol Rev Camb Philos Soc* **65**, 431–471 (1990).

[14] Jablonka, E. & Raz, G. Transgenerational epigenetic inheritance: prevalence, mechanisms, and implications for the study of heredity and evolution. *Q Rev Biol* **84**, 131–176 (2009).

[15] Jablonka, E. Epigenetic inheritance and plasticity: The responsive germline. *Prog Biophys Mol Biol* (2012). URL http://dx.doi.org/10.1016/j.pbiomolbio.2012.08.014.

[16] Hathaway, N. A. *et al.* Dynamics and memory of heterochromatin in living cells. *Cell* **149**, 1447–1460 (2012). URL http://dx.doi.org/10.1016/j.cell.2012.03.052.

[17] Blow, J. J. & Dutta, A. Preventing re-replication of chromosomal dna. *Nat Rev Mol Cell Biol* **6**, 476–486 (2005). URL http://dx.doi.org/10.1038/nrm1663.

[18] Jacob, F., Brenner, S. & Cuzin, F. On the regulation of dna replication in bacteria. *Cold Spring Harbor Symposia on Quantitative Biology* **28**, 329–348 (1963). URL http://symposium.cshlp.org/content/28/329.short. http://symposium.cshlp.org/content/28/329.full.pdf+html.

[19] Sclafani, R. A. & Holzen, T. M. Cell cycle regulation of dna replication. *Annu Rev Genet* **41**, 237–280 (2007). URL http://dx.doi.org/10.1146/annurev.genet.41.110306.130308.

[20] Chagin, V. O., Stear, J. H. & Cardoso, M. C. Organization of dna replication. *Cold Spring Harb Perspect Biol* **2**, a000737 (2010). URL http://dx.doi.org/10.1101/cshperspect.a000737.

[21] Chagin, V. O., Casas-Delucchi, C. S., Reinhart, M. & Cardoso, M. C. To be published (2013).

[22] Goldar, A., Marsolier-Kergoat, M.-C. & Hyrien, O. Universal temporal profile of replication origin activation in eukaryotes. *PLoS ONE* **4**, e5899 (2009).

[23] Woodfine, K. *et al.* Replication timing of human chromosome 6. *Cell Cycle* **4**, 172–176 (2005).

[24] Farkash-Amar, S. *et al.* Global organization of replication time zones of the mouse genome. *Genome Res* **18**, 1562–1570 (2008). URL http://dx.doi.org/10.1101/gr.079566.108.

[25] Cayrou, C. *et al.* New insights into replication origin characteristics in metazoans. *Cell Cycle* **11**, 658–667 (2012). URL http://dx.doi.org/10.4161/cc.11.4.19097.

[26] Karnani, N., Taylor, C., Malhotra, A. & Dutta, A. Pan-s replication patterns and chromosomal domains defined by genome-tiling arrays of encode genomic areas. *Genome Res* **17**, 865–876 (2007). URL http://dx.doi.org/10.1101/gr.5427007.

[27] Lucas, I. *et al.* High-throughput mapping of origins of replication in human cells. *EMBO Rep* **8**, 770–777 (2007). URL http://dx.doi.org/10.1038/sj.embor.7401026.

[28] White, E. J. *et al.* Dna replication-timing analysis of human chromosome 22 at high resolution and different developmental states. *Proc Natl Acad Sci U S A* **101**, 17771–17776 (2004). URL http://dx.doi.org/10.1073/pnas.0408170101.

[29] Donaldson, A. D. Shaping time: chromatin structure and the dna replication programme. *Trends Genet* **21**, 444–449 (2005). URL http://dx.doi.org/10.1016/j.tig.2005.05.012.

[30] Woodfine, K. *et al.* Replication timing of the human genome. *Human Molecular Genetics* **13**, 191–202 (2004).

[31] Hiratani, I. *et al.* Global reorganization of replication domains during embryonic stem cell differentiation. *PLoS Biol* **6**, e245 (2008). URL http://dx.doi.org/10.1371/journal.pbio.0060245.

[32] Sequeira-Mendes, J. *et al.* Transcription initiation activity sets replication origin efficiency in mammalian cells. *PLoS Genet* **5**, e1000446 (2009). URL http://dx.doi.org/10.1371/journal.pgen.1000446.

[33] DePamphilis, M. L. Replication origins in metazoan chromosomes: fact or fiction? *Bioessays* **21**, 5–16 (1999). URL http://dx.doi.org/3.0.CO;2-6.

[34] Machida, Y. J., Hamlin, J. L. & Dutta, A. Right place, right time, and only once: replication initiation in metazoans. *Cell* **123**, 13–24 (2005). URL http://dx.doi.org/10.1016/j.cell.2005.09.019.

[35] Méchali, M. Eukaryotic dna replication origins: many choices for appropriate answers. *Nat Rev Mol Cell Biol* **11**, 728–738 (2010). URL http://dx.doi.org/10.1038/nrm2976.

[36] Lygeros, J. *et al.* Stochastic hybrid modeling of dna replication across a complete genome. *Proc Natl Acad Sci U S A* **105**, 12295–12300 (2008). URL http://dx.doi.org/10.1073/pnas.0805549105.

[37] Spiesser, T. W., Klipp, E. & Barberis, M. A model for the spatiotemporal organization of dna replication in saccharomyces cerevisiae. *Mol Genet Genomics* **282**, 25–35 (2009). URL http://dx.doi.org/10.1007/s00438-009-0443-9.

[38] Yang, S. C.-H., Rhind, N. & Bechhoefer, J. Modeling genome-wide replication kinetics reveals a mechanism for regulation of replication timing. *Mol Syst Biol* **6**, 404 (2010). URL http://dx.doi.org/10.1038/msb.2010.61.

[39] Lebofsky, R., Heilig, R., Sonnleitner, M., Weissenbach, J. & Bensimon, A. Dna replication origin interference increases the spacing between initiation events in human cells. *Mol Biol Cell* **17**, 5337–5345 (2006). URL http://dx.doi.org/10.1091/mbc.E06-04-0298.

[40] Patel, P. K., Arcangioli, B., Baker, S. P., Bensimon, A. & Rhind, N. Dna replication origins fire stochastically in fission yeast. *Mol Biol Cell* **17**, 308–316 (2006). URL http://dx.doi.org/10.1091/mbc.E05-07-0657.

[41] Rhind, N. Dna replication timing: random thoughts about origin firing. *Nat Cell Biol* **8**, 1313–1316 (2006). URL http://dx.doi.org/10.1038/ncb1206-1313.

[42] Czajkowsky, D. M., Liu, J., Hamlin, J. L. & Shao, Z. Dna combing reveals intrinsic temporal disorder in the replication of yeast chromosome vi. *J Mol Biol* **375**, 12–19 (2008). URL http://dx.doi.org/10.1016/j.jmb.2007.10.046.

[43] Takebayashi, S.-I. *et al.* Regulation of replication at the r/g chromosomal band boundary and pericentromeric heterochromatin of mammalian cells. *Exp Cell Res* **304**, 162–174 (2005). URL http://dx.doi.org/10.1016/j.yexcr.2004.10.024.

[44] Goren, A. & Cedar, H. Replicating by the clock. *Nat Rev Mol Cell Biol* **4**, 25–32 (2003). URL http://dx.doi.org/10.1038/nrm1008.

[45] Shopland, L. S., Johnson, C. V., Byron, M., McNeil, J. & Lawrence, J. B. Clustering of multiple specific genes and gene-rich r-bands around sc-35 domains: evidence for local euchromatic neighborhoods. *J Cell Biol* **162**, 981–990 (2003). URL http://dx.doi.org/10.1083/jcb.200303131.

[46] O'Keefe, R. T., Henderson, S. C. & Spector, D. L. Dynamic organization of dna replication in mammalian cell nuclei: spatially and temporally defined replication of chromosome-specific alpha-satellite dna sequences. *J Cell Biol* **116**, 1095–1110 (1992).

[47] Berezney, R., Dubey, D. D. & Huberman, J. A. Heterogeneity of eukaryotic replicons, replicon clusters, and replication foci. *Chromosoma* **108**, 471–484 (2000). URL http://dx.doi.org/10.1007/s004120050399. 10.1007/s004120050399.

[48] Aladjem, M. I. & Fanning, E. The replicon revisited: an old model learns new tricks in metazoan chromosomes. *EMBO Rep* **5**, 686–691 (2004). URL http://dx.doi.org/10.1038/sj.embor.7400185.

[49] Zink, D. The temporal program of dna replication: new insights into old questions. *Chromosoma* **115**, 273–287 (2006). URL http://dx.doi.org/10.1007/s00412-006-0062-8.

[50] Gilbert, D. M. Replication origin plasticity, taylor-made: inhibition vs recruitment of origins under conditions of replication stress. *Chromosoma* **116**, 341–347 (2007). URL http://dx.doi.org/10.1007/s00412-007-0105-9.

[51] Willis, N. & Rhind, N. Regulation of dna replication by the s-phase dna damage checkpoint. *Cell Div* **4**, 13 (2009). URL http://dx.doi.org/10.1186/1747-1028-4-13.

[52] Nordman, J. & Orr-Weaver, T. L. Regulation of dna replication during development. *Development* **139**, 455–464 (2012). URL http://dx.doi.org/10.1242/dev.061838.

[53] Widrow, R. J., Hansen, R. S., Kawame, H., Gartler, S. M. & Laird, C. D. Very late dna replication in the human cell cycle. *PNAS* **95**, 11246–11250 (1998).

[54] Friedberg, E. C. *et al. DNA repair and mutagenesis* (American Soc. for Microbiology Press, 2006), 2 edn.

[55] Memisoglu, A. & Samson, L. Base excision repair in yeast and mammals. *Mutat Res* **451**, 39–51 (2000).

[56] Christmann, M., Tomicic, M. T., Roos, W. P. & Kaina, B. Mechanisms of human dna repair: an update. *Toxicology* **193**, 3–34 (2003).

[57] Bennardo, N., Cheng, A., Huang, N. & Stark, J. M. Alternative-nhej is a mechanistically distinct pathway of mammalian chromosome break repair. *PLoS Genet* **4**, e1000110 (2008). URL http://dx.doi.org/10.1371/journal.pgen.1000110.

[58] McVey, M. & Lee, S. E. Mmej repair of double-strand breaks (director's cut): deleted sequences and alternative endings. *Trends Genet* **24**, 529–538 (2008). URL http://dx.doi.org/10.1016/j.tig.2008.08.007.

[59] Hopfield, J. J. Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proc Natl Acad Sci U S A* **71**, 4135–4139 (1974).

[60] SHIMOMURA, O., JOHNSON, F. H. & SAIGA, Y. Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, aequorea. *J Cell Comp Physiol* **59**, 223–239 (1962).

[61] Shimomura, O. Structure of the chromophore of aequorea green fluorescent protein. {*FEBS*} *Letters* **104**, 220 – 222 (1979). URL http://www.sciencedirect.com/science/article/pii/0014579379808182.

[62] Prasher, D. C., Eckenrode, V. K., Ward, W. W., Prendergast, F. G. & Cormier, M. J. Primary structure of the aequorea victoria green-fluorescent protein. *Gene* **111**, 229–233 (1992).

[63] Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W. & Prasher, D. C. Green fluorescent protein as a marker for gene expression. *Science* **263**, 802–805 (1994).

[64] Inouye, S. & Tsuji, F. I. Aequorea green fluorescent protein. expression of the gene and fluorescence characteristics of the recombinant protein. *FEBS Lett* **341**, 277–280 (1994).

[65] Sadegh Zadeh, K., Montas, H. J. & Shirmohammadi, A. Identification of biomolecule mass transport and binding rate parameters in living cells by inverse modeling. *Theor Biol Med Model* **3**, 36 (2006). URL http://dx.doi.org/10.1186/1742-4682-3-36.

[66] Sprague, B. L. *et al.* Analysis of binding at a single spatially localized cluster of binding sites by fluorescence recovery after photobleaching. *Biophys J* **91**, 1169–1191 (2006). URL http://dx.doi.org/10.1529/biophysj.105.073676.

[67] Hargrove, J. L. Microcomputer-assisted kinetic modeling of mammalian gene expression. *FASEB J* **7**, 1163–1170 (1993).

[68] Drossel, B. Komplexe dynamische systeme. Tech. Rep., Technische Universität Darmstadt (2010).

[69] Wutz, A. Epigenetic regulation of stem cells : the role of chromatin in cell differentiation. *Adv Exp Med Biol* **786**, 307–328 (2013). URL http://dx.doi.org/10.1007/978-94-007-6621-1_17.

[70] Cook, D. & Genever, P. Regulation of mesenchymal stem cell differentiation. *Adv Exp Med Biol* **786**, 213–229 (2013). URL http://dx.doi.org/10.1007/978-94-007-6621-1_12.

[71] Ferrell, J. E., Jr. Bistability, bifurcations, and waddington's epigenetic landscape. *Curr Biol* **22**, R458–R466 (2012). URL http://dx.doi.org/10.1016/j.cub.2012.03.045.

[72] Cepko, C. L., Austin, C. P., Yang, X., Alexiades, M. & Ezzeddine, D. Cell fate determination in the vertebrate retina. *Proc Natl Acad Sci U S A* **93**, 589–595 (1996).

[73] Selivanov, V. A. *et al.* Multistationary and oscillatory modes of free radicals generation by the mitochondrial respiratory chain revealed by a bifurcation analysis. *PLoS Comput Biol* **8**, e1002700 (2012). URL http://dx.doi.org/10.1371/journal.pcbi.1002700.

[74] Santos, S. D. M. & Ferrell, J. E. Systems biology: On the cell cycle and its switches. *Nature* **454**, 288–289 (2008). URL http://dx.doi.org/10.1038/454288a.

[75] Santos, S. D. M., Wollman, R., Meyer, T. & Ferrell, J. E., Jr. Spatial positive feedback at the onset of mitosis. *Cell* **149**, 1500–1513 (2012). URL http://dx.doi.org/10.1016/j.cell.2012.05.028.

[76] Roos, W. P. & Kaina, B. Dna damage-induced cell death: from specific dna lesions to the dna damage response and apoptosis. *Cancer Lett* **332**, 237–248 (2013). URL http://dx.doi.org/10.1016/j.canlet.2012.01.007.

[77] Pritchard, L. & Kell, D. B. Schemes of flux control in a model of saccharomyces cerevisiae glycolysis. *Eur J Biochem* **269**, 3894–3904 (2002).

[78] Gehrmann, E. *et al.* Robustness of glycolysis in yeast to internal and external noise. *Phys Rev E Stat Nonlin Soft Matter Phys* **84**, 021913 (2011).

[79] Rust, M. J., Golden, S. S. & O'Shea, E. K. Light-driven changes in energy metabolism directly entrain the cyanobacterial circadian oscillator. *Science* **331**, 220–223 (2011). URL http://dx.doi.org/10.1126/science.1197243.

[80] Angeli, D., Ferrell, J. E., Jr & Sontag, E. D. Detection of multistability, bifurcations, and hysteresis in a large class of biological positive-feedback systems. *Proc Natl Acad Sci U S A* **101**, 1822–1827 (2004). URL http://dx.doi.org/10.1073/pnas.0308265100.

[81] Novák, B. & Tyson, J. J. Design principles of biochemical oscillators. *Nat Rev Mol Cell Biol* **9**, 981–991 (2008). URL http://dx.doi.org/10.1038/nrm2530.

[82] Pigolotti, S., Krishna, S. & Jensen, M. H. Oscillation patterns in negative feedback loops. *Proc Natl Acad Sci U S A* **104**, 6533–6537 (2007). URL http://dx.doi.org/10.1073/pnas.0610759104.

[83] Ackermann, E., Weiel, E. M., Pfaff, T. & Drossel, B. Boolean versus continuous dynamics in modules with two feedback loops. *Eur Phys J E Soft Matter* **35**, 107 (2012). URL http://dx.doi.org/10.1140/epje/i2012-12107-9.

[84] Soulé, C. Graphic requirements for multistationarity. *ComPlexUs* **1**, 123–133 (2003).

[85] Amin, M., Porter, S. L. & Soyer, O. S. Split histidine kinases enable ultrasensitivity and bistability in two-component signaling networks. *PLoS Comput Biol* **9**, e1002949 (2013). URL http://dx.doi.org/10.1371/journal.pcbi.1002949.

[86] Mouri, K., Nacher, J. C. & Akutsu, T. A mathematical model for the detection mechanism of dna double-strand breaks depending on autophosphorylation of atm. *PLoS One* **4**, e5131 (2009). URL http://dx.doi.org/10.1371/journal.pone.0005131.

[87] Cedar, H. & Bergman, Y. Programming of dna methylation patterns. *Annu Rev Biochem* **81**, 97–117 (2012). URL http://dx.doi.org/10.1146/annurev-biochem-052610-091920.

[88] Strogatz, S. H. *Nonlinear Dynamics and Chaos* (Westview Press, 2000).

[89] Ferrell, J. E., Jr *et al.* Simple, realistic models of complex biological processes: positive feedback and bistability in a cell fate switch and a cell cycle oscillator. *FEBS Lett* **583**, 3999–4005 (2009). URL http://dx.doi.org/10.1016/j.febslet.2009.10.068.

[90] Tobias, F. *et al.* Spatiotemporal dynamics of early dna damage response proteins on complex dna lesions. *PLoS One* **8**, e57953 (2013). URL http://dx.doi.org/10.1371/journal.pone.0057953.

[91] Petrini, J. H. J. & Stracker, T. H. The cellular response to dna double-strand breaks: defining the sensors and mediators. *Trends Cell Biol* **13**, 458–462 (2003).

[92] Difilippantonio, S. & Nussenzweig, A. The nbs1-atm connection revisited. *Cell Cycle* **6**, 2366–2370 (2007).

[93] Hopfner, K.-P. *et al.* The rad50 zinc-hook is a structure joining mre11 complexes in dna recombination and repair. *Nature* **418**, 562–566 (2002). URL http://dx.doi.org/10.1038/nature00922.

[94] Uziel, T. *et al.* Requirement of the mrn complex for atm activation by dna damage. *EMBO J* **22**, 5612–5621 (2003). URL http://dx.doi.org/10.1093/emboj/cdg541.

[95] Lee, J.-H. & Paull, T. T. Atm activation by dna double-strand breaks through the mre11-rad50-nbs1 complex. *Science* **308**, 551–554 (2005). URL http://dx.doi.org/10.1126/science.1108297.

[96] Cerosaletti, K., Wright, J. & Concannon, P. Active role for nibrin in the kinetics of atm activation. *Mol Cell Biol* **26**, 1691–1699 (2006). URL http://dx.doi.org/10.1128/MCB.26.5.1691-1699.2006.

[97] Lee, J.-H. & Paull, T. T. Activation and regulation of atm kinase activity in response to dna double-strand breaks. *Oncogene* **26**, 7741–7748 (2007). URL http://dx.doi.org/10.1038/sj.onc.1210872.

[98] You, Z., Chahwan, C., Bailis, J., Hunter, T. & Russell, P. Atm activation and its recruitment to damaged dna require binding to the c terminus of nbs1. *Mol Cell Biol* **25**, 5363–5379 (2005). URL http://dx.doi.org/10.1128/MCB.25.13.5363-5379.2005.

[99] Rogakou, E. P., Pilch, D. R., Orr, A. H., Ivanova, V. S. & Bonner, W. M. Dna double-stranded breaks induce histone h2ax phosphorylation on serine 139. *J Biol Chem* **273**, 5858–5868 (1998).

[100] Bhatti, S. *et al.* Atm protein kinase: the linchpin of cellular defenses to stress. *Cell Mol Life Sci* **68**, 2977–3006 (2011). URL http://dx.doi.org/10.1007/s00018-011-0683-9.

[101] Bakkenist, C. J. & Kastan, M. B. Dna damage activates atm through intermolecular autophosphorylation and dimer dissociation. *Nature* **421**, 499–506 (2003). URL http://dx.doi.org/10.1038/nature01368.

[102] Kozlov, S. V. *et al.* Autophosphorylation and atm activation: additional sites add to the complexity. *J Biol Chem* **286**, 9107–9119 (2011). URL http://dx.doi.org/10.1074/jbc.M110.204065.

[103] Stewart, G. S., Wang, B., Bignell, C. R., Taylor, A. M. R. & Elledge, S. J. Mdc1 is a mediator of the mammalian dna damage checkpoint. *Nature* **421**, 961–966 (2003). URL http://dx.doi.org/10.1038/nature01446.

[104] Stucki, M. *et al.* Mdc1 directly binds phosphorylated histone h2ax to regulate cellular responses to dna double-strand breaks. *Cell* **123**, 1213–1226 (2005). URL http://dx.doi.org/10.1016/j.cell.2005.09.038.

[105] Stucki, M. & Jackson, S. P. gammah2ax and mdc1: anchoring the dna-damage-response machinery to broken chromosomes. *DNA Repair (Amst)* **5**, 534–543 (2006). URL http://dx.doi.org/10.1016/j.dnarep.2006.01.012.

[106] Spycher, C. *et al.* Constitutive phosphorylation of mdc1 physically links the mre11-rad50-nbs1 complex to damaged chromatin. *J Cell Biol* **181**, 227–240 (2008). URL http://dx.doi.org/10.1083/jcb.200709008.

[107] Jungmichel, S. & Stucki, M. Mdc1: The art of keeping things in focus. *Chromosoma* **119**, 337–349 (2010). URL http://dx.doi.org/10.1007/s00412-010-0266-9.

[108] Lukas, C., Falck, J., Bartkova, J., Bartek, J. & Lukas, J. Distinct spatiotemporal dynamics of mammalian checkpoint regulators induced by dna damage. *Nat Cell Biol* **5**, 255–260 (2003). URL http://dx.doi.org/10.1038/ncb945.

[109] Lukas, C. *et al.* Mdc1 couples dna double-strand break recognition by nbs1 with its h2ax-dependent chromatin retention. *EMBO J* **23**, 2674–2683 (2004). URL http://dx.doi.org/10.1038/sj.emboj.7600269.

[110] Bekker-Jensen, S., Lukas, C., Melander, F., Bartek, J. & Lukas, J. Dynamic assembly and sustained retention of 53bp1 at the sites of dna damage are controlled by mdc1/nfbd1. *J Cell Biol* **170**, 201–211 (2005). URL http://dx.doi.org/10.1083/jcb.200503043.

[111] Jakob, B., Rudolph, J. H., Gueven, N., Lavin, M. F. & Taucher-Scholz, G. Live cell imaging of heavy-ion-induced radiation responses by beamline microscopy. *Radiat Res* **163**, 681–690 (2005).

[112] Jakob, B. *et al.* Dna double-strand breaks in heterochromatin elicit fast repair protein recruitment, histone h2ax phosphorylation and relocation to euchromatin. *Nucleic Acids Res* **39**, 6489–6499 (2011). URL http://dx.doi.org/10.1093/nar/gkr230.

[113] Bekker-Jensen, S. *et al.* Spatial organization of the mammalian genome surveillance machinery in response to dna strand breaks. *J Cell Biol* **173**, 195–206 (2006). URL http://dx.doi.org/10.1083/jcb.200510130.

[114] Tobias, F. *Analyse von schnellen dynamischen Prozessen und Proteininteraktionen nach dicht ionisierender Bestrahlung*. Ph.D. thesis, TU Darmstadt (2012).

[115] Wu, L., Luo, K., Lou, Z. & Chen, J. Mdc1 regulates intra-s-phase checkpoint by targeting nbs1 to dna double-strand breaks. *Proc Natl Acad Sci U S A* **105**, 11200–11205 (2008). URL http://dx.doi.org/10.1073/pnas.0802885105.

[116] Chapman, J. R. & Jackson, S. P. Phospho-dependent interactions between nbs1 and mdc1 mediate chromatin retention of the mrn complex at sites of dna damage. *EMBO Rep* **9**, 795–801 (2008). URL http://dx.doi.org/10.1038/embor.2008.103.

[117] Löbrich, M., Rydberg, B. & Cooper, P. K. Dna double-strand breaks induced by high-energy neon and iron ions in human fibroblasts. ii. probing individual noti fragments by hybridization. *Radiat Res* **139**, 142–151 (1994).

[118] Prise, K. M. *et al.* A review of dsb induction data for varying quality radiations. *Int J Radiat Biol* **74**, 173–184 (1998).

[119] Cash, J. R. & Karp, A. H. A variable order runge-kutta method for initial value problems with rapidly varying right-hand sides. *ACM Trans. Math. Softw.* **16**, 201–222 (1990). URL http://doi.acm.org/10.1145/79505.79507.

[120] Nelder, J. A. & Mead, R. A simplex method for function minimization. *The Computer Journal* **7**, 308–313 (1965). URL http://dx.doi.org/10.1093/comjnl/7.4.308.

[121] Lou, Z. *et al.* Mdc1 maintains genomic stability by participating in the amplification of atm-dependent dna damage signals. *Mol Cell* **21**, 187–200 (2006). URL http://dx.doi.org/10.1016/j.molcel.2005.11.025.

[122] Kozlov, S. V. *et al.* Involvement of novel autophosphorylation sites in atm activation. *EMBO J* **25**, 3504–3514 (2006). URL http://dx.doi.org/10.1038/sj.emboj.7601231.

[123] Politi, A. *et al.* Mathematical modeling of nucleotide excision repair reveals efficiency of sequential assembly strategies. *Mol Cell* **19**, 679–690 (2005). URL http://dx.doi.org/10.1016/j.molcel.2005.06.036.

[124] Cucinotta, F. A., Pluth, J. M., Anderson, J. A., Harper, J. V. & O'Neill, P. Biochemical kinetics model of dsb repair and induction of gamma-h2ax foci by non-homologous end joining. *Radiat Res* **169**, 214–222 (2008). URL http://dx.doi.org/10.1667/RR1035.1.

[125] Meyer, B. *Kernweite H2AX-Phosphorylierung nach Schwerionenbestrahlung*. Ph.D. thesis, TU Darmstadt (2012).

[126] Hable, V. *et al.* Recruitment kinetics of dna repair proteins mdc1 and rad52 but not 53bp1 depend on damage complexity. *PLoS One* **7**, e41943 (2012). URL http://dx.doi.org/10.1371/journal.pone.0041943.

[127] Gapud, E. J. *et al.* Ataxia telangiectasia mutated (atm) and dna-pkcs kinases have overlapping activities during chromosomal signal joint formation. *Proc Natl Acad Sci U S A* **108**, 2022–2027 (2011). URL http://dx.doi.org/10.1073/pnas.1013295108.

[128] Friedland, W., Jacob, P. & Kundrát, P. Stochastic simulation of dna double-strand break repair by non-homologous end joining based on track structure calculations. *Radiat Res* **173**, 677–688 (2010). URL http://dx.doi.org/10.1667/RR1965.1.

[129] Friedland, W., Kundrát, P. & Jacob, P. Stochastic modelling of dsb repair after photon and ion irradiation. *Int J Radiat Biol* **88**, 129–136 (2012). URL http://dx.doi.org/10.3109/09553002.2011.611404.

[130] Thomas, R. On the relation between the logical structure of systems and their ability to generate multiple steady states or sustained oscillations. *Springer series in Synergetics* **9**, 180–193 (1981).

[131] Doedel, E. J. & Oldeman, B. E. Continuation and bifurcation software for ordinary differential equations. Tech. Rep., Concordia University, Montreal (2012).

[132] Priester, C. *Suche nach Bistabilität bei der Bildung von Proteinclustern*. Master's thesis, Technische Universität Darmstadt (2012).

[133] Feinberg, M. Chemical reaction network structure and the stability of complex isothermal reactors - i. the deficiency zero and deficiency one theorems. *Chemical Engineering Science* **42**, 2229–2268 (1987).

[134] Feinberg, M. The existence and uniqueness of steady states for a class of chemical reaction networks. *Arch. Rational Mech. Anal.* **132**, 311–370 (1995).

[135] Feinberg, M. Multiple steady states for chemical reaction networks of deficiency one. *Archive for Rational Mechanics and Analysis* **132**, 371–406 (1995). URL http://dx.doi.org/10.1007/BF00375615. 10.1007/BF00375615.

[136] Ellison, P. R. *The Advanced Deficiency Algorithm and Its Applications to Mechanism Discrimination*. Ph.D. thesis, University of Rochester (1998).

[137] Gunawardena, J. Chemical reaction network theory for in-silico biologists. Tech. Rep., Bauer Center for Genomics Research, Harvard University (2003).

[138] Craciun, G. & Feinberg, M. Multiple equilibria in complex chemical reaction networks: extensions to entrapped species models. *Syst Biol (Stevenage)* **153**, 179–186 (2006).

[139] Shinar, G. & Feinberg, M. Concordant chemical reaction networks. *Math Biosci* **240**, 92–113 (2012). URL http://dx.doi.org/10.1016/j.mbs.2012.05.004.

[140] Shinar, G. & Feinberg, M. Concordant chemical reaction networks and the species-reaction graph. *Math Biosci* **241**, 1–23 (2013). URL http://dx.doi.org/10.1016/j.mbs.2012.08.002.

[141] Lotka, A. J. Undamped oscillations derived from the law of mass action. *J. Am. Chem. Soc* **42**, 1595–1599 (1920).

[142] Prigogine, I. & Lefever, R. Symmetry breaking instabilities in dissipative systems. ii. *J. Chem. Phys* **48**, 1695–1700 (1968).

[143] Bray, W. A periodic reaction in homogeneous solution and its relation to catalysis. *J. Am. Chem. Soc.* **43**, 1262–1267 (1921).

[144] Belousov, B. P. A periodic reaction and its mechanism. *Collection of short papers on radiation medicine* **147**, 145 (1958).

[145] Zhabotinsky, A. M. Periodical oxidation of malonic acid in solution (a study of the belousov reaction kinetics). *Biofizika* **9**, 306–311 (1964).

[146] Griffith, J. S. Mathematics of cellular control processes. i. negative feedback to one gene. *J Theor Biol* **20**, 202–208 (1968).

[147] Michaelis, L. & Menten, M. L. Die kinetik der invertinwirkung. *Biochem. z* **49**, 352 (1913).

[148] de Moura, A. P. S., Retkute, R., Hawkins, M. & Nieduszynski, C. A. Mathematical modelling of whole chromosome replication. *Nucleic Acids Res* **38**, 5623–5633 (2010). URL http://dx.doi.org/10.1093/nar/gkq343.

[149] Blow, J. J. & Ge, X. Q. A model for dna replication showing how dormant origins safeguard against replication fork failure. *EMBO Rep* **10**, 406–412 (2009). URL http://dx.doi.org/10.1038/embor.2009.5.

[150] Shaw, A., Olivares-Chauvet, P., Maya-Mendoza, A. & Jackson, D. A. S-phase progression in mammalian cells: modelling the influence of nuclear organization. *Chromosome Research* 163–178 (2010).

[151] Cayrou, C. *et al.* Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features. *Genome Res* **21**, 1438–1449 (2011). URL http://dx.doi.org/10.1101/gr.121830.111.

[152] Conti, C. *et al.* Replication fork velocities at adjacent replication origins are coordinately modified during dna replication in human cells. *Molecular Biology of the Cell* **18**, 3059–3067 (2007). URL http://www.molbiolcell.org/cgi/content/abstract/18/8/3059.

[153] Hyrien, O., Marheineke, K. & Goldar, A. Paradoxes of eukaryotic dna replication: Mcm proteins and the random completion problem. *Bioessays* **25**, 116–125 (2003). URL http://dx.doi.org/10.1002/bies.10208.

[154] Jackson, D. A. & Pombo, A. Replicon clusters are stable units of chromosome structure: evidence that nuclear organization contributes to the efficient activation and propagation of s phase in human cells. *J Cell Biol* **140**, 1285–1295 (1998).

[155] Guilbaud, G. *et al.* Evidence for sequential and increasing activation of replication origins along replication timing gradients in the human genome. *PLoS Comput Biol* **7**, e1002322 (2011). URL http://dx.doi.org/10.1371/journal.pcbi.1002322.

[156] Bechhoefer, J. & Rhind, N. Replication timing and its emergence from stochastic processes. *Trends Genet* (2012). URL http://dx.doi.org/10.1016/j.tig.2012.03.011.

[157] Goldar, A., Labit, H., Marheineke, K. & Hyrien, O. A dynamic stochastic model for dna replication initiation in early embryos. *PLoS ONE* **3**, e2919 (2008).

[158] Baker, A. *et al.* Replication fork polarity gradients revealed by megabase-sized u-shaped replication timing domains in human cell lines. *PLoS Comput Biol* **8**, e1002443 (2012). URL http://dx.doi.org/10.1371/journal.pcbi.1002443.

[159] Dreszer, T. R. *et al.* The ucsc genome browser database: extensions and updates 2011. *Nucleic Acids Res* **40**, D918–D923 (2012). URL http://dx.doi.org/10.1093/nar/gkr1055.

[160] Drouin, R., Lemieux, N. & Richer, C. L. Analysis of dna replication during s-phase by means of dynamic chromosome banding at high resolution. *Chromosoma* **99**, 273–280 (1990).

[161] Shaffer, L. G., Slovak, M. L. & J., C. L. *ISCN 2009: An International System for Human Cytogenetic Nomenclature (2009): Recommendations of the International Standing Committee on Human Cytogenetic Nomenclature* (S. Karger Publishing, 2009), 1 edn.

[162] Macville, M. *et al.* Comprehensive and definitive molecular cytogenetic characterization of hela cells by spectral karyotyping. *Cancer Research* **59**, 141–150 (1999).

[163] Takahashi, M. A model for the spatio-temporal organization of dna replication in mammalian cells. *J. theor. Biol.* **129**, 91–115 (1987).

[164] Cremer, T. & Cremer, C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet* **2**, 292–301 (2001). URL http://dx.doi.org/10.1038/35066075.

[165] Courbet, S. *et al.* Replication fork movement sets chromatin loop size and origin choice in mammalian cells. *Nature* **455**, 557–560 (2008). URL http://dx.doi.org/10.1038/nature07233.

[166] Guillou, E. *et al.* Cohesin organizes chromatin loops at dna replication factories. *Genes Dev* **24**, 2812–2822 (2010). URL http://dx.doi.org/10.1101/gad.608210.

[167] Nakamura, H., Morita, T. & Sato, C. Structural organizations of replicon domains during dna synthetic phase in the mammalian nucleus. *Exp Cell Res* **165**, 291–297 (1986).

[168] Ferreira, J., Paolella, G., Ramos, C. & Lamond, A. I. Spatial organization of large-scale chromatin domains in the nucleus: a magnified view of single chromosome territories. *J Cell Biol* **139**, 1597–1610 (1997).

[169] Dross, N. *et al.* Mapping egfp oligomer mobility in living cell nuclei. *PLoS ONE* **4**, e5041 (2009). URL http://dx.doi.org/10.1371%2Fjournal.pone.0005041.

[170] Landry, J. J. M. *et al.* The genomic and transcriptomic landscape of a hela cell line. *G3 (Bethesda)* (2013). URL http://dx.doi.org/10.1534/g3.113.005777.

[171] Salic, A. & Mitchison, T. J. A chemical method for fast and sensitive detection of dna synthesis in vivo. *Proc Natl Acad Sci U S A* **105**, 2415–2420 (2008). URL http://dx.doi.org/10.1073/pnas.0712168105.

[172] Huberman, J. A. & Riggs, A. D. Autoradiography of chromosomal dna fibers from chinese hamster cells. *Proc Natl Acad Sci U S A* **55**, 599–606 (1966).

[173] Dershowitz, A. & Newlon, C. S. The effect on chromosome stability of deleting replication origins. *Mol Cell Biol* **13**, 391–398 (1993).

[174] Dershowitz, A. *et al.* Linear derivatives of saccharomyces cerevisiae chromosome iii can be maintained in the absence of autonomously replicating sequence elements. *Mol Cell Biol* **27**, 4652–4663 (2007). URL http://dx.doi.org/10.1128/MCB.01246-06.

[175] Watanabe, Y. *et al.* Chromosome-wide assessment of replication timing for human chromosomes 11q and 21q: disease-related genes in timing-switch regions. *Human Molecular Genetics* **11**, 13–21 (2002).

[176] Hansen, R. S. *et al.* Sequencing newly replicated dna reveals widespread plasticity in human replication timing. *PNAS* **107**, 139–144 (2010).

[177] Manders, E. M., Stap, J., Brakenhoff, G. J., van Driel, R. & Aten, J. A. Dynamics of three-dimensional replication patterns during the s-phase, analysed by double labelling of dna and confocal microscopy. *J Cell Sci* **103 (Pt 3)**, 857–862 (1992).

[178] Bohn, M., Heermann, D. W. & van Driel, R. Random loop model for long polymers. *Phys Rev E Stat Nonlin Soft Matter Phys* **76**, 051805 (2007).

[179] Jun, S., Zhang, H. & Bechhoefer, J. Nucleation and growth in one dimension. i. the generalized kolmogorov-johnson-mehl-avrami model. *Phys. Rev. E* **71**, 011908 (2005).

[180] Bechhoefer, J. & Marshall, B. How xenopus laevis replicates dna reliably even though its origins of replication are located and initiated stochastically. *Phys Rev Lett* **98**, 098105 (2007).

[181] Alexandrow, M. G. & Hamlin, J. L. Chromatin decondensation in s-phase involves recruitment of cdk2 by cdc45 and histone h1 phosphorylation. *The Journal of Cell Biology* **168**, 875–886 (2005).

[182] Bickmore, W. A. & Carothers, A. D. Factors affecting the timing and imprinting of replication on a mammalian chromosome. *J Cell Sci* **108 (Pt 8)**, 2801–2809 (1995).

[183] Parada, L. & Misteli, T. Chromosome positioning in the interphase nucleus. *Trends Cell Biol* **12**, 425–432 (2002).

[184] Bolzer, A. *et al.* Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS Biol* **3**, e157 (2005). URL `http://dx.doi.org/10.1371/journal.pbio.0030157`.

[185] Casas-Delucchi, C. S. *et al.* Histone acetylation controls the inactive x chromosome replication dynamics. *nature communications* **2** (2010).

Appendix

1 U, M, L Inequalities for Deficiency One Algorithm

As explained in section 4, the symmetries of network (4.18-4.21) effectively reduce its deficiency to $\delta = 1$. Consequently, only one of the possible variants of confluence vector (4.27) has to be investigated. The author chooses $\alpha > 0, \beta > 0$, so that the confluence vector has the following signs (example vector with $\alpha = 1, \beta = 1$):

$$g = (-1, 1, -1, 1, 2, -2, -1, -1, 2, 0)\,.$$

As mentioned before, $\{\mu_A, \mu_{AB}, \mu_{AC}, \mu_{ABC}\}$ and analogue sets for B and C have to contain at least one positive and one negative element due to the conservation of each elementary species. Several systems to be investigated contain the contradiction $\mu_A + \mu_B + \mu_C > \mu_A + \mu_B + \mu_C$ which allows to skip detailed investigation. With X being either U, M or L, these systems are:

(i) $\{\Lambda_1, \Lambda_2, \Lambda_3, \Lambda_4\} = \{X, X, U, U\}$ except $\{L, L, U, U\}$

(ii) $\{\Lambda_1, \Lambda_2, \Lambda_3, \Lambda_4\} = \{X, X, M, U\}$ except $\{L, L, M, U\}$

(iii) $\{\Lambda_1, \Lambda_2, \Lambda_3, \Lambda_4\} = \{X, X, U, M\}$ except $\{L, L, U, M\}$

(iv) $\{\Lambda_1, \Lambda_2, \Lambda_3, \Lambda_4\} = \{X, X, M, M\}$ except $\{M, M, M, M\}$

The listed exceptions must be investigated directly, which is done in the remainder of this section.

$\{\Lambda_1, \Lambda_2, \Lambda_3, \Lambda_4\} = \{L, L, U, U\}$

$$\begin{gathered}
\mu_A + \mu_B > \mu_{AB}, \\
\mu_A + \mu_C > \mu_{AC}, \\
\mu_B + \mu_C > \mu_{BC}, \\
\{\mu_{AB} + \mu_C, \mu_{AC} + \mu_B\} < \mu_{ABC} < \mu_A + \mu_{BC}, \\
\{\mu_B + \mu_C, \mu_{BC}, \mu_{AB} + \mu_C, \mu_{AC} + \mu_B, \mu_{ABC}, \mu_A + \mu_{BC}\} > \{\mu_A + \mu_B, \mu_{AB}, \mu_A + \mu_C, \mu_{AC}\}
\end{gathered}$$

It follows from the last row that $B > 0$, $AC > A$ and $BC > C > 0$. Because B and BC are positive, the first conservation exclusion rule demands that either A or AC should be positive. Since $AC > A$, this means that AC is always positive. Consequently, with $\mu_{AC} + \mu_B < \mu_{ABC}$, all μ components containing B are positive, which conflicts with the conservation of B. The same argument holds for $\{L, L, M, U\}$ and $\{L, L, U, M\}$.

$\{\Lambda_1, \Lambda_2, \Lambda_3, \Lambda_4\} = \{M, M, M, M\}$

$$\mu_A + \mu_B = \mu_{AB},$$
$$\mu_A + \mu_C = \mu_{AC},$$
$$\mu_B + \mu_C = \mu_{BC},$$
$$\{\mu_{AB} + \mu_C, \mu_{AC} + \mu_B\} = \mu_{ABC} = \mu_A + \mu_{BC}$$

For this system, several distinct cases have to be investigated:

$\mu_A, \mu_B > 0$**:** The first row states that $\mu_{AB} > 0$. If $\mu_{AC} > 0$, then it follows that $\mu_{AC} + \mu_B = \mu_{ABC} > 0$, which contradicts the conservation of A. If, on the other hand, $\mu_{AC} < 0$, then, because of $\mu_A + \mu_C = \mu_{AC}$, μ_C must be negative. This, however, is contradicted by the third conservation exclusion rule.

$\mu_A, \mu_B < 0$**:** This combination is just the sign reversed previous combination and thus already contradicted.

$\mu_A > 0, \mu_B < 0, \mu_{AB} > 0$**:** Whenever $\mu_{ABC} > 0$, the last equation leads to $\mu_{AC} > 0$, since $\mu_B < 0$. This is in conflict with conservation of A, as μ_A, μ_{AB}, μ_{AC} and μ_{ABC} cannot all be positive. If $\mu_{ABC} < 0$, then from $\mu_{AB} + \mu_C = \mu_{ABC}$ follows that $\mu_C < 0$ and from $\mu_{BC} + \mu_A$ follows that $\mu_{BC} < 0$. The second conservation exclusion rule, however, states that μ_C and μ_{BC} cannot both be negative if μ_A and μ_{AB} are positive.

$\mu_A > 0, \mu_B < 0, \mu_{AB} < 0$**:** Because the UML partition $\{\Lambda_1, \Lambda_2, \Lambda_3, \Lambda_4\} = \{M, M, M, M\}$ is symmetrical with regard to the permutation of the elementary species, this case is identical with the previous case $\mu_A > 0, \mu_B < 0, \mu_{AB} > 0$ (it can be transformed into this case by exchanging A with B and reversing all signs).

$\mu_A < 0, \mu_B > 0$ All combinations possible here are sign-reversed to $\mu_A > 0, \mu_B < 0$ and thus already covered by the above contradictions.

$\{\Lambda_1, \Lambda_2, \Lambda_3, \Lambda_4\} = \{U, U, L, U\}$

$$\mu_A + \mu_B < \mu_{AB},$$
$$\mu_A + \mu_C < \mu_{AC},$$
$$\mu_B + \mu_C < \mu_{BC},$$
$$\{\mu_{AB} + \mu_C, \mu_{AC} + \mu_B\} < \mu_{ABC} < \mu_A + \mu_{BC},$$
$$\{\mu_A + \mu_B, \mu_{AB}, \mu_A + \mu_C, \mu_{AC}, \mu_{AB} + \mu_C, \mu_{AC} + \mu_B, \mu_{ABC}, \mu_A + \mu_{BC}\} > \{\mu_B + \mu_C, \mu_{BC}\}$$

The last row contains the following relations: $\mu_A > 0$, $\mu_A > \mu_C$, $\mu_A > \mu_B$, $\mu_{AC} > \mu_C$, $\mu_{AB} > \mu_B$, $\mu_{AC} > \mu_{BC}$, $\mu_{AB} > \mu_{BC}$ and $\mu_{ABC} > \mu_{BC}$. Whenever $\mu_{BC} > 0$, all entries in μ containing A are positive, in violation of the conservation of A. There remain three cases with $\mu_{BC} < 0$:

$\mu_{BC} < 0$**,**$\mu_B > 0$**:** In this case, $\mu_A > 0$, $\mu_{AB} > 0$ (because of $\mu_{AB} > \mu_B$), $\mu_C < 0$ (because of $\mu_B + \mu_C < \mu_{BC}$) and $\mu_{BC} < 0$, which is a combination that is forbidden by the second conservation exclusion rule.

$\mu_{BC} < 0$,$\mu_C > 0$**:** In this case, $\mu_A > 0$, $\mu_{AC} > 0$ (because of $\mu_{AC} > \mu_C$), $\mu_B < 0$ (because of $\mu_B + \mu_C < \mu_{BC}$) and $\mu_{BC} < 0$, which is a combination that is forbidden by the first conservation exclusion rule.

$\mu_B < 0$,$\mu_C < 0$**:** Because $\mu_C < 0$, $\mu_{BC} < 0$ and $\mu_A > 0$, due to the second conservation exclusion rule, μ_{AB} must be negative. With $\mu_{BC} < \mu_{AC} + \mu_B$ and $\mu_{AB} > \mu_A + \mu_B$, it follows that $\mu_{ABC} < \mu_A + \mu_{BC} < \mu_A + \mu_{AC} + \mu_B < 0$. However, if $\mu_{ABC} < 0$, then all μ components containing B are negative, in conflict with the conservation of B.

$\{\Lambda_1, \Lambda_2, \Lambda_3, \Lambda_4\} = \{M, U, L, U\}$

$$\mu_A + \mu_B = \mu_{AB},$$
$$\mu_A + \mu_C < \mu_{AC},$$
$$\mu_B + \mu_C < \mu_{BC},$$
$$\{\mu_{AB} + \mu_C, \mu_{AC} + \mu_B\} < \mu_{ABC} < \mu_A + \mu_{BC},$$
$$\{\mu_A + \mu_C, \mu_{AC}, \mu_{AB} + \mu_C, \mu_{AC} + \mu_B, \mu_{ABC}, \mu_A + \mu_{BC}\} > \{\mu_A + \mu_B, \mu_{AB}\} > \{\mu_B + \mu_C, \mu_{BC}\}$$

The last row gives the relations $\mu_{AC} > \mu_A > \mu_C > 0$ and $\mu_{ABC} > \mu_{AB} > \mu_{BC} > 0 > \mu_B$. If μ_{AB} or μ_{BC} (and thus also μ_{AB}) are positive, all μ entries for species containing A are positive, which is in conflict with the conservation of A. In the remaining case ($\mu_{AB} < 0$, $\mu_{BC} < 0$), μ_A and μ_{AC} are positive, while μ_B and μ_{BC} are negative, in violation of the first conservation exclusion rule.

$\{\Lambda_1, \Lambda_2, \Lambda_3, \Lambda_4\} = \{L, U, L, U\}$

$$\mu_A + \mu_B > \mu_{AB},$$
$$\mu_A + \mu_C < \mu_{AC},$$
$$\mu_B + \mu_C < \mu_{BC},$$
$$\{\mu_{AB} + \mu_C, \mu_{AC} + \mu_B\} < \mu_{ABC} < \mu_A + \mu_{BC},$$
$$\{\mu_A + \mu_C, \mu_{AC}, \mu_{AB} + \mu_C, \mu_{AC} + \mu_B, \mu_{ABC}, \mu_A + \mu_{BC}\} > \{\mu_A + \mu_B, \mu_{AB}, \mu_B + \mu_C, \mu_{BC}\}$$

The last row contains the relations $\mu_C > 0$ and $\mu_{AC} > 0$, from which follows that, due to the third conservation exclusion rule, either μ_B or μ_{AB} must be positive. If $\mu_B > 0$, then, because $\mu_C > 0$, μ_{BC} is also positive. Since the last row contains $\mu_{ABC} > \mu_{BC}$, all species containing C have positive μ components, which is contradicted by the conservation of C.

If, on the other hand, $\mu_{AB} > 0$, then it follows from $\mu_{ABC} > \mu_{AB} + \mu_C$ that all μ components containing A are positive, which is prohibited by the conservation of A.

$\{\Lambda_1, \Lambda_2, \Lambda_3, \Lambda_4\} = \{M, M, L, U\}$

$$\mu_A + \mu_B = \mu_{AB},$$
$$\mu_A + \mu_C = \mu_{AC},$$
$$\mu_B + \mu_C < \mu_{BC},$$
$$\{\mu_{AB} + \mu_C, \mu_{AC} + \mu_B\} < \mu_{ABC} < \mu_A + \mu_{BC},$$
$$\{\mu_{AB} + \mu_C, \mu_{AC} + \mu_B, \mu_{ABC}, \mu_A + \mu_{BC}\} > \{\mu_A + \mu_B, \mu_{AB}, \mu_A + \mu_C, \mu_{AC}\} > \{\mu_B + \mu_C, \mu_{BC}\}$$

The last row directly leads to: $\{\mu_A, \mu_B, \mu_C\} > 0, \mu_{AB} > \mu_A, \mu_{AC} > \mu_A$ and $\mu_{BC} > \mu_B$. Using this in the third row results in all components of μ being positive, which contradicts conservation.

$\{\Lambda_1, \Lambda_2, \Lambda_3, \Lambda_4\} = \{M, L, L, U\}$

$$\begin{aligned}
\mu_A + \mu_B &= \mu_{AB}, \\
\mu_A + \mu_C &> \mu_{AC}, \\
\mu_B + \mu_C &< \mu_{BC}, \\
\{\mu_{AB} + \mu_C, \mu_{AC} + \mu_B\} &< \mu_{ABC} < \mu_A + \mu_{BC}, \\
\{\mu_{AB} + \mu_C, \mu_{AC} + \mu_B, \mu_{ABC}, \mu_A + \mu_{BC}\} &> \{\mu_A + \mu_B, \mu_{AB}\} > \{\mu_B + \mu_C, \mu_{BC}, \mu_A + \mu_C, \mu_{AC}\}
\end{aligned}$$

The same steps as in $\{M, M, L, U\}$ lead to the same contradiction.

$\{\Lambda_1, \Lambda_2, \Lambda_3, \Lambda_4\} = \{L, M, L, U\}$

$$\begin{aligned}
\mu_A + \mu_B &> \mu_{AB}, \\
\mu_A + \mu_C &= \mu_{AC}, \\
\mu_B + \mu_C &< \mu_{BC}, \\
\{\mu_{AB} + \mu_C, \mu_{AC} + \mu_B\} &< \mu_{ABC} < \mu_A + \mu_{BC}, \\
\{\mu_{AB} + \mu_C, \mu_{AC} + \mu_B, \mu_{ABC}, \mu_A + \mu_{BC}\} &> \{\mu_A + \mu_C, \mu_{AC}\} > \{\mu_B + \mu_C, \mu_{BC}, \mu_A + \mu_B, \mu_{AB}\}
\end{aligned}$$

The same steps as in $\{M, M, L, U\}$ lead to the same contradiction.

$\{\Lambda_1, \Lambda_2, \Lambda_3, \Lambda_4\} = \{L, L, L, U\}$

$$\begin{aligned}
\mu_A + \mu_B &> \mu_{AB}, \\
\mu_A + \mu_C &> \mu_{AC}, \\
\mu_B + \mu_C &< \mu_{BC}, \\
\{\mu_{AB} + \mu_C, \mu_{AC} + \mu_B\} &< \mu_{ABC} < \mu_A + \mu_{BC}, \\
\{\mu_{AB} + \mu_C, \mu_{AC} + \mu_B, \mu_{ABC}, \mu_A + \mu_{BC}\} &> \{\mu_A + \mu_B, \mu_{AB}, \mu_B + \mu_C, \mu_{BC}, \mu_A + \mu_C, \mu_{AC}\}
\end{aligned}$$

The same steps as in $\{M, M, L, U\}$ lead to the same contradiction.

The remaining variants of U, M and L all are inversions of the above systems. Since in each inversion all inequalities are reversed compared to the original system, the solution (if existent) would therefore be $-\mu$ (if μ is a solution of the original system). However, if μ is not compatible to the stochiometric space of the network, then $-\mu$ is neither. In conclusion, it was shown that none of the possible set combinations yields an inequality system that has a solution which is compatible to the stochiometric subspace. Thus, there cannot be multiple steady states in network (4.18-4.21).

2 Additional Replication Data and Timing Comparisons

The comparison of replication timing data by Woodfine et al. [30] and the replication model presented here was repeated for 23 human chromosomes. The same level of agreement as for chromosome 6 was found for all chromosomes except chromosomes 9, 16 and the X chromosome. In chromosomes 9 and 16, the experimental data shows early replication of larger heterochromatic regions, whereas in the X chromosome data, the overall differences in replication timing for euchromatin and heterochromatin are much less pronounced. Possible reasons for this could be either epigenetic modifications regulating the replication of these chromosomes or the experiment statistics. The Pearson's correlation coefficients for all chromosomes and figures analogous to figure 5.8 for all chromosomes are shown in Table 1 and the comparisons are shown in Figures 4-7.

Chromosome	Correlation Coefficient
1	0.34
2	0.13
3	0.13
4	0.31
5	0.18
6	0.36
7	0.30
8	0.24
9	−0.01
10	0.26
11	0.21
12	0.28
13	0.24
14	0.19
15	0.11
16	0.03
17	0.36
18	0.19
19	0.47
20	0.14
21	0.27
22	0.26
X	−0.02

Table 1: **Pearson's correlation coefficients for all chromosomes.** Coefficients are calculated between the timing of sample positions in the model and in micro-array measurements performed by Woodfine et al. [30] for 23 human chromosomes. The theoretical values used were averaged over 100 simulations, the experimental data is averaged over 4 data sets.

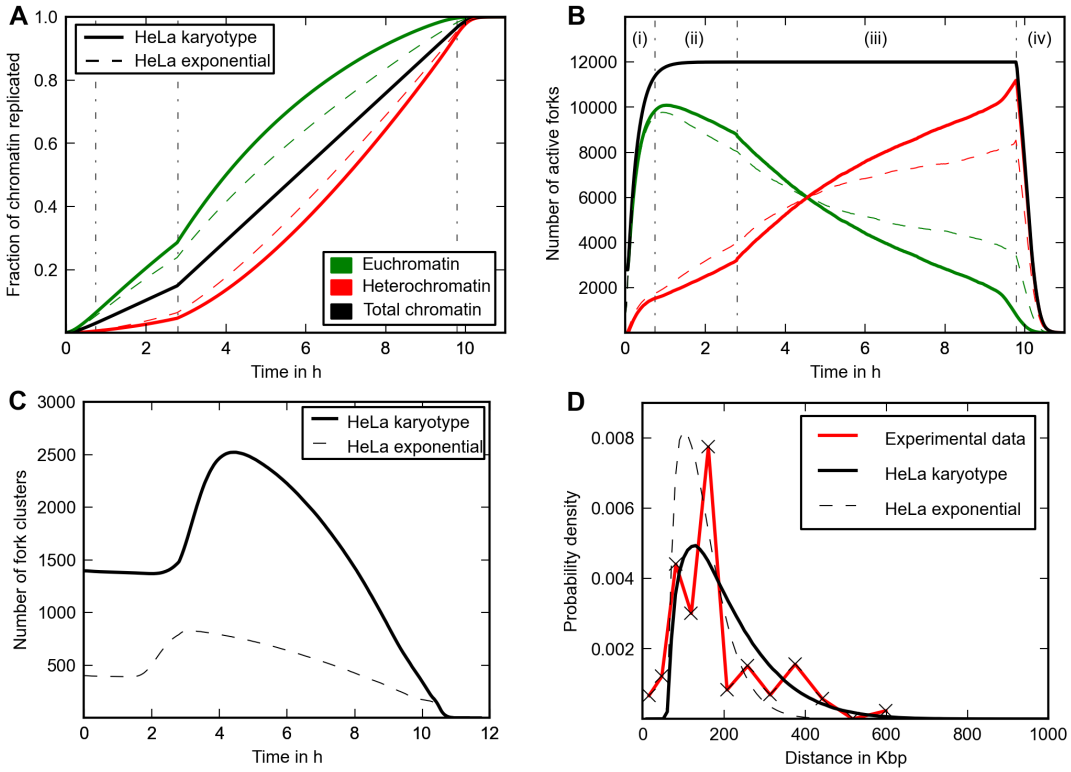

Figure 1: **Comparison of results for the chromatin pattern modeled on human cells and an exponential chromatin zone size distribution. A** Total fraction of the genome replicated over time, analogous to figure 5.5 A. **B** Number of active forks in the different chromatin types over time, analogous to figure 5.5 B. **C** The number of replication clusters over time for both chromatin distributions, analogous to figure 5.7 A. For the exponential distribution, the total number of clusters is significantly reduced from the outset, since many initial firing positions (due to the large number of very small chromatin zones) are closer than 1 Mbp together. **D** Distribution of distances between adjacent fired origins, analogous to figure 5.7 B. Because lower cluster counts lead to clusters that are larger and more dense, the distance distribution for the exponential chromatin zone size distribution is shifted towards low distances.

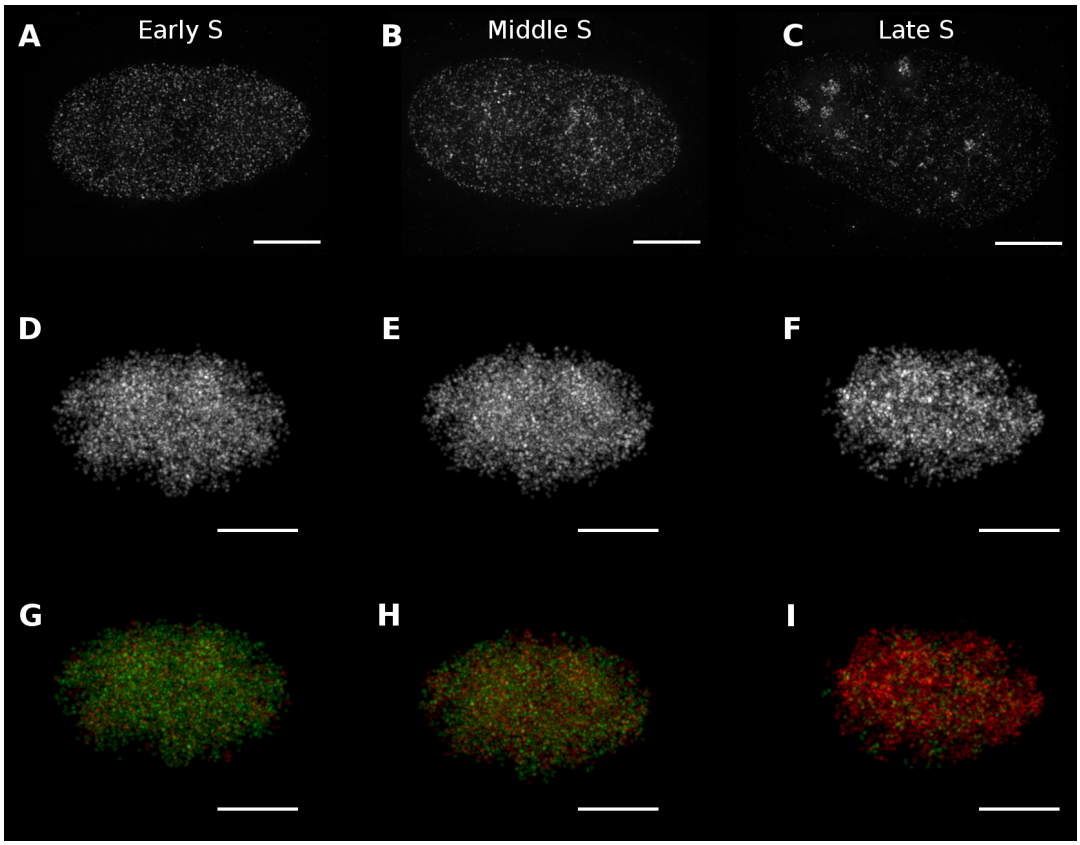

Figure 2: **Microscopy-like 3D projection for the random loop model parameters used by Mateos-Langerak [10]**. Both chromatin types have the same spring constant but the number of connections within them is different. Consistent with that publication, the total number of connections is 5000, and the relative connection portions are 7/11 for Heterochromatin, 3/11 for Euchromatin and 1/11 for inter-chromatin connections. Using these parameters, no clearly discernible formation of 3D foci is observed.

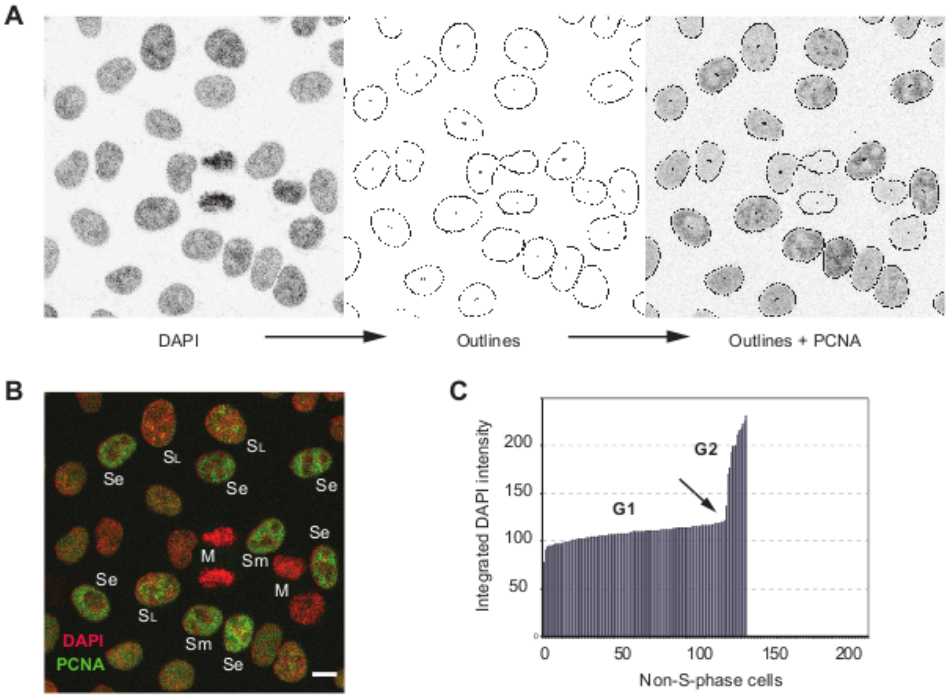

Figure 3: **Quantification of DNA content corresponding to the three major S-phase patterns.** Nuclei of HeLa Kyoto cells expressing mCherry PCNA are stained with DAPI.**A** Procedure for assigning DNA contents of individual cells with S-phase patterns. **B** Cells in early, middle and late S-phase were classified based on characteristic features of the PCNA distribution: Uniform nucleoplasmic foci in early S-phase, perinucleoar foci rings in middle S-phase and bright foci clusters in late S-phase, respectively. Mitotic nuclei can be recognized by their relatively high DAPI intensity and absence of PCNA signal. **C** Classification of non-S-phase cells into G1 and G2 populations based on the sharp increase in DAPI intensity. **A** and **B** show a small area of the field used for the analysis in **C**. The scale bar is 10 µm, this data was also used to estimate the nuclear sizes and shapes for the 3D model. Figure courtesy of Vadim Chagin.

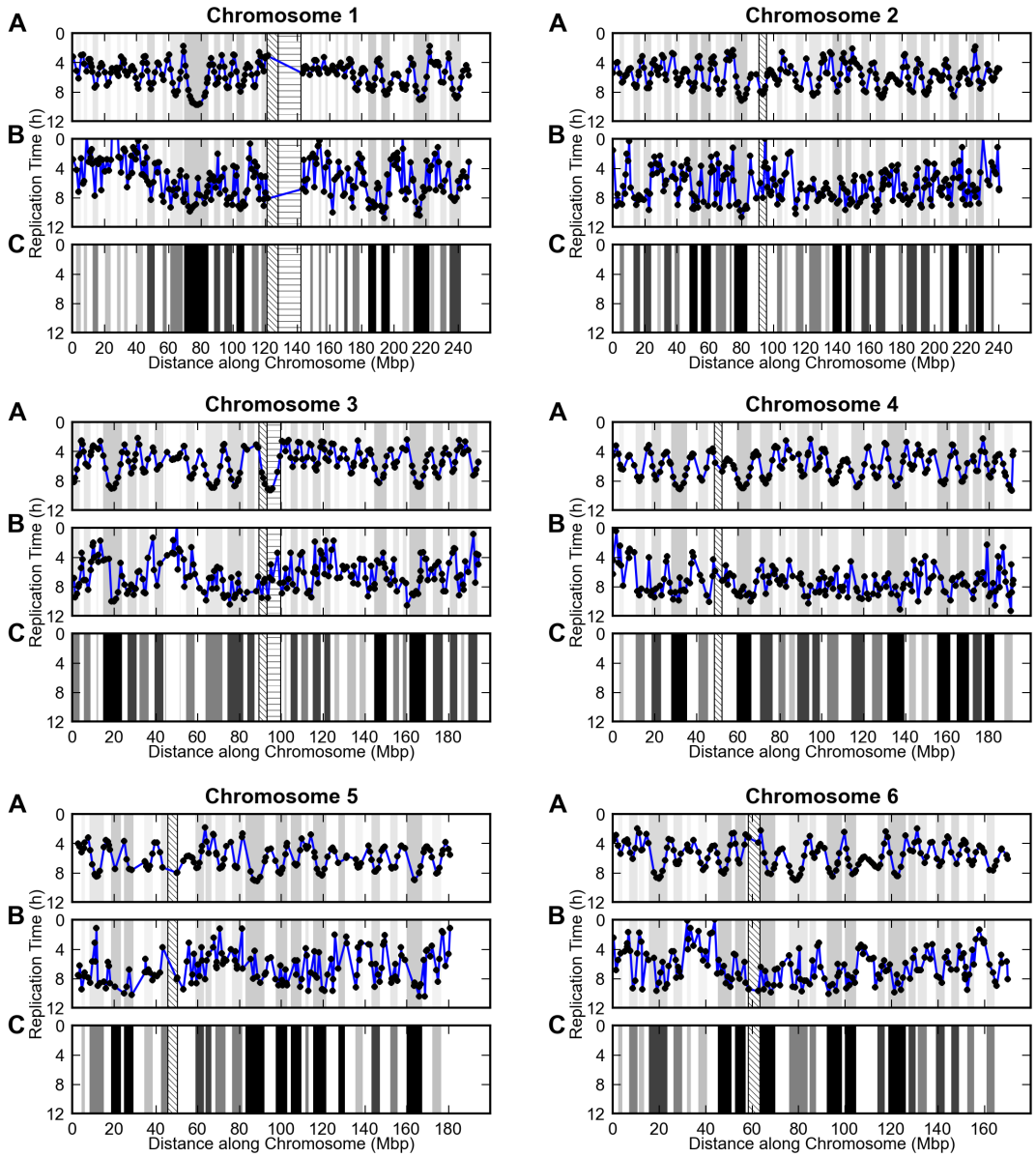

Figure 4: **Replication timing comparison for chromosomes 1-6.**

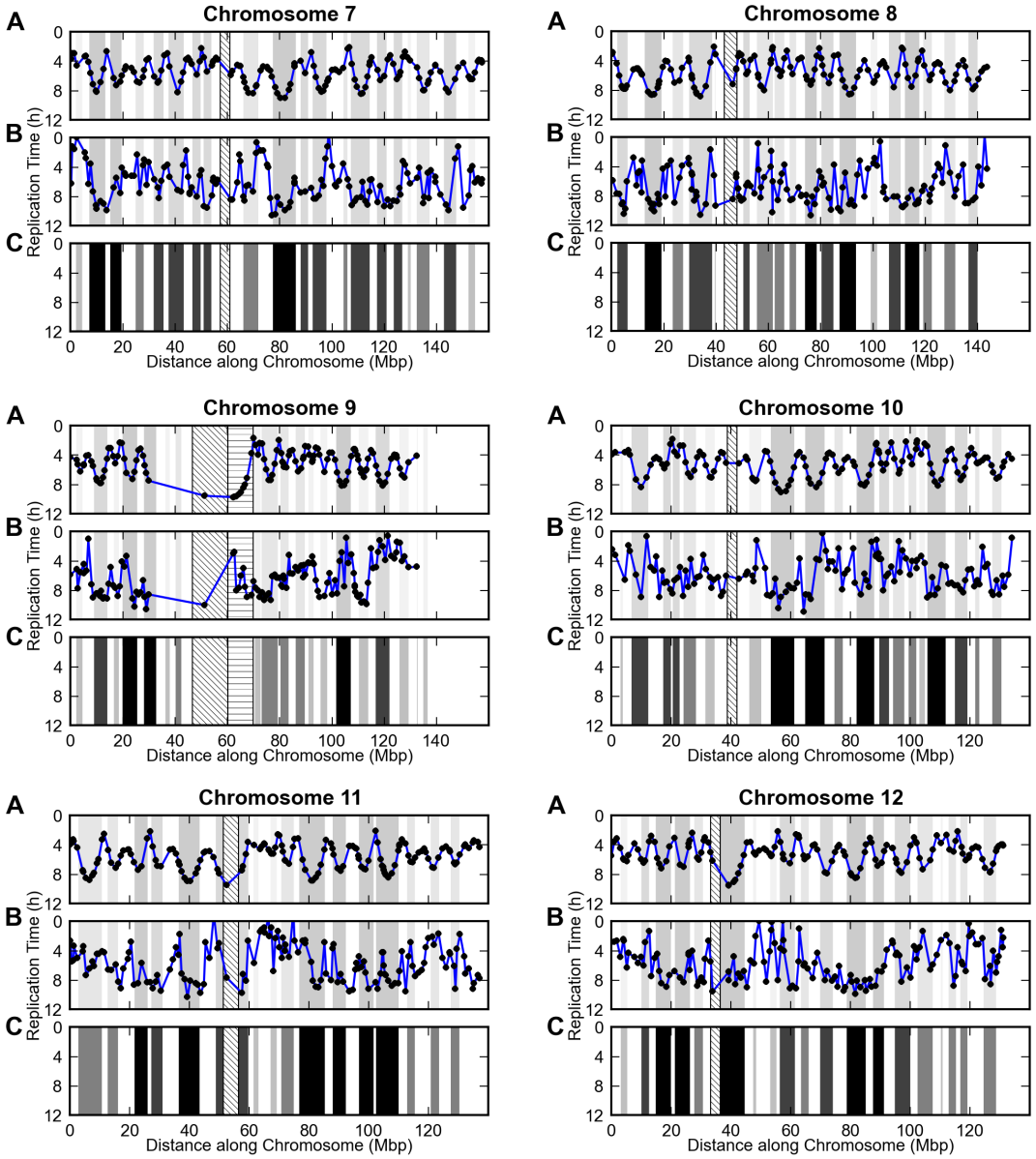

Figure 5: **Replication timing comparison for chromosomes 7-12.**

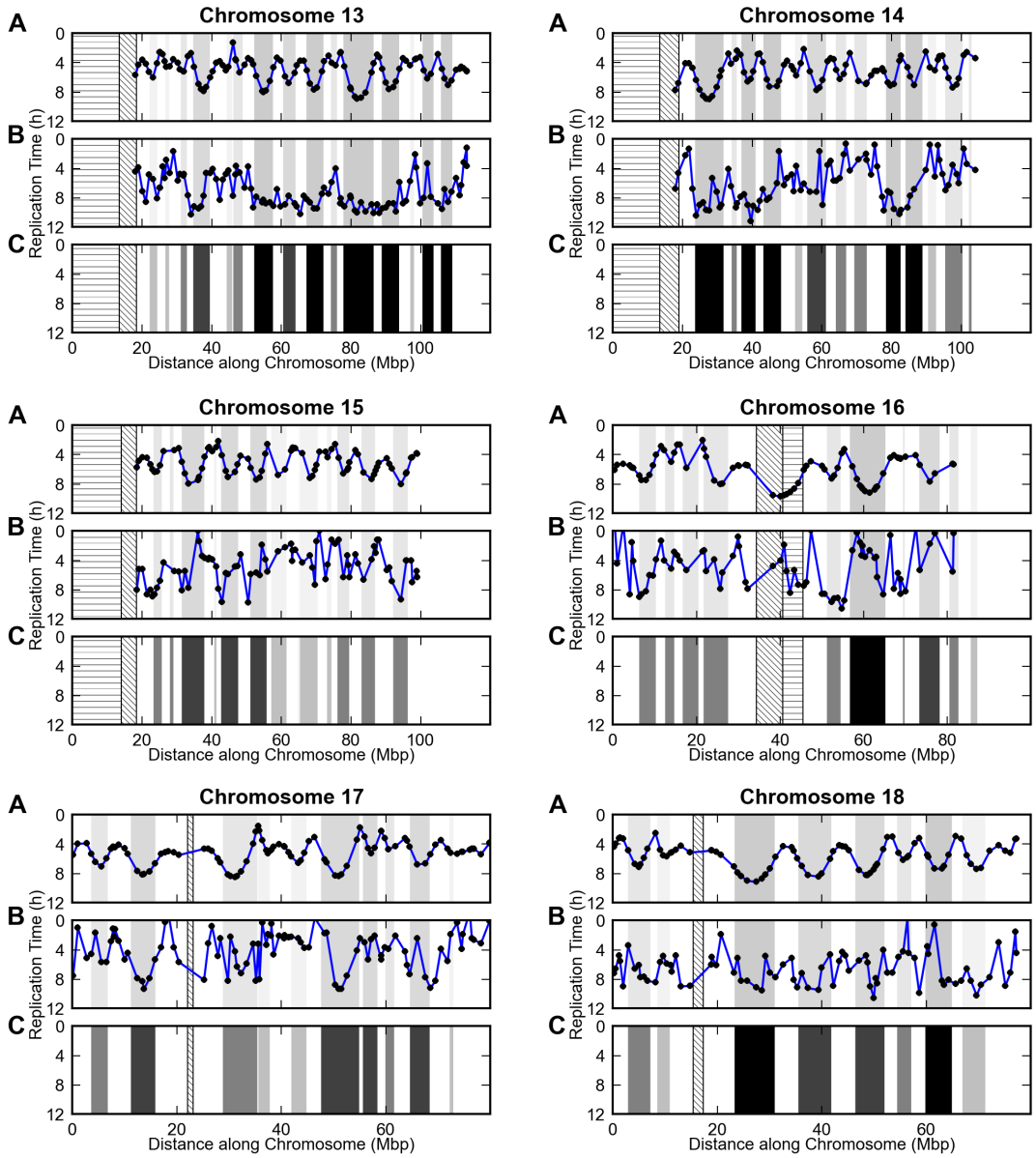

Figure 6: **Replication timing comparison for chromosomes 13-18.**

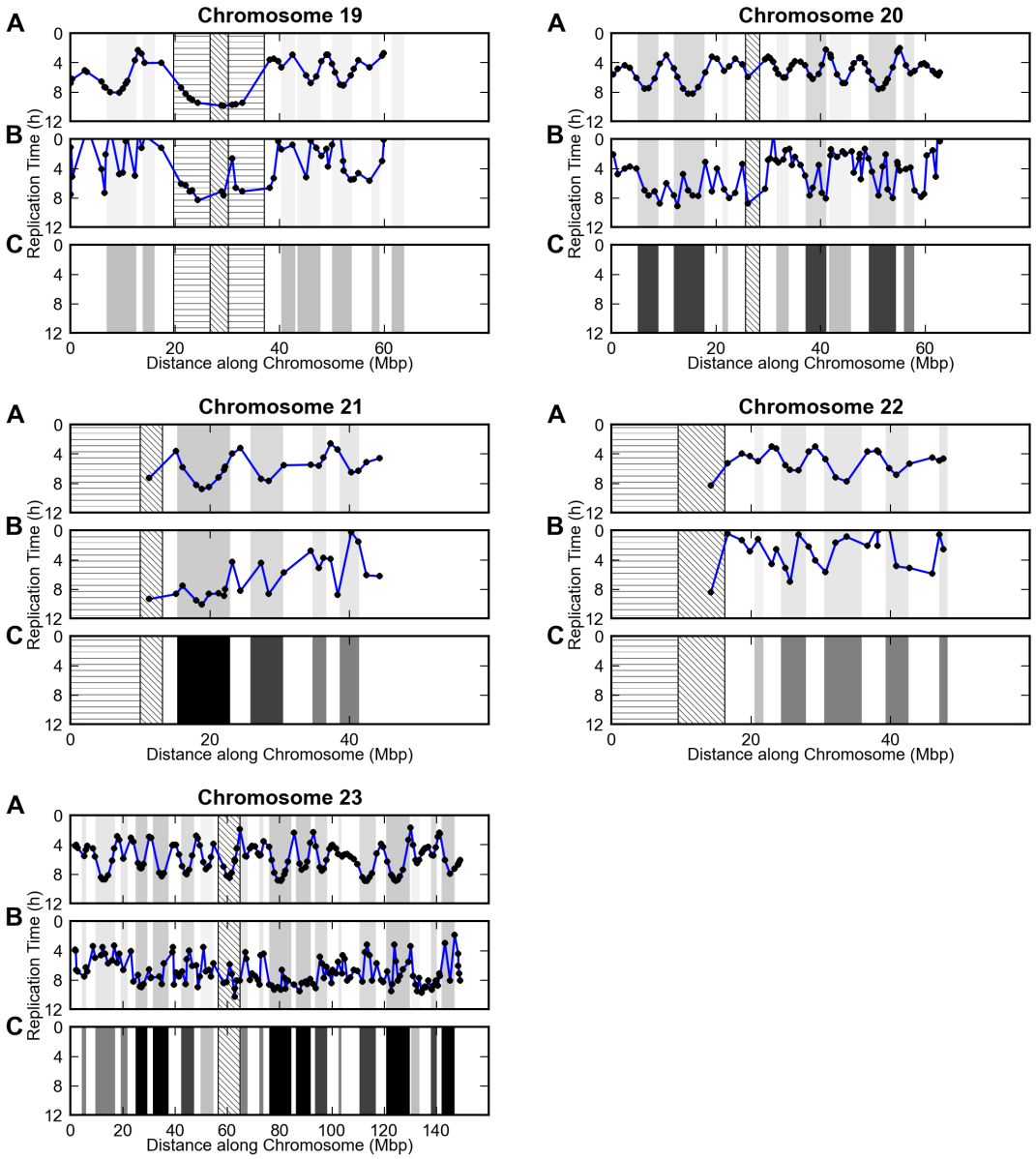

Figure 7: **Replication timing comparison for chromosomes 19-22 and the X chromosome.**

3 Full Optimization Results for the DSB Response Model

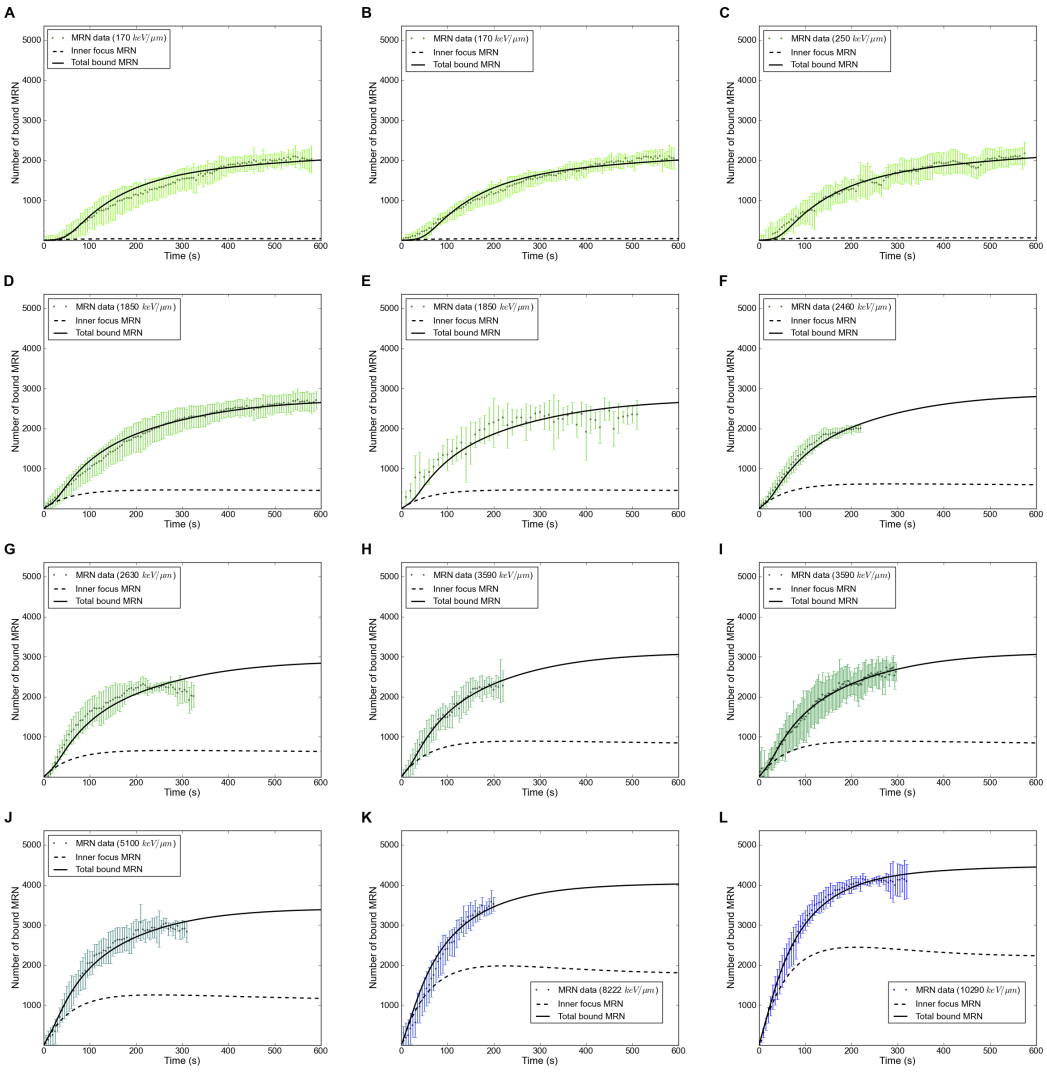

Figure 8: **Comparison of the final optimization result with all NBS1 recruitment data sets.** Protein concentrations and rate constants were identical for all model calculations. Only the number of DSBs was set calculated from LET for each simulation. Curves **B**, **I** and **L** are shown in section 3.3. Experimental data courtesy of Frank Tobias and Burkhard Jakob.

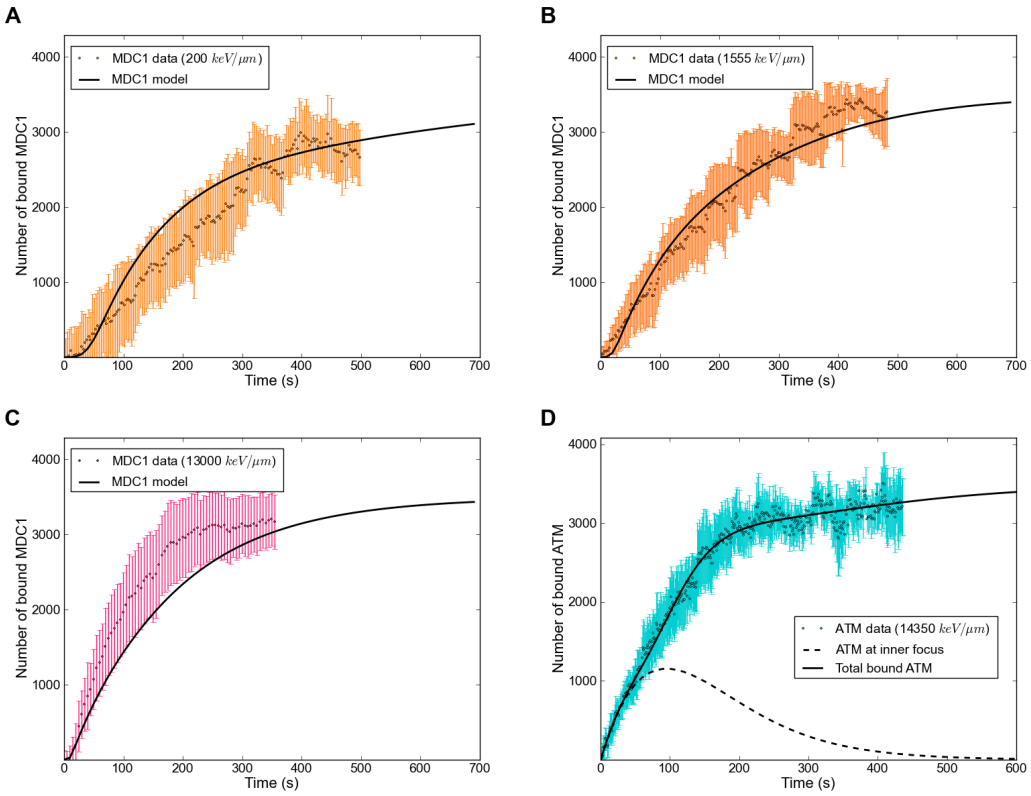

Figure 9: **Comparison of the final optimization result with all MDC1 and ATM recruitment data sets.** MDC1 data sets were not used in the model parameter optimization, so the absolute values shown here are chosen for best experiment/simulation compatibility. The scaling parameter for the ATM data set is 3401.57. Experimental data courtesy of Frank Tobias and Burkhard Jakob.

Reaction	Rate constant
A	2067.24
B	2184.11
C	2097.18
D	2716.53
E	2331.74
F	1997.80
G	2238.05
H	2208.15
I	2738.65
J	2904.45
K	3483.53
L	4102.78

Table 2: **Optimization results for the NBS1 data set scaling parameters of the DSB response model.** Introduction of these parameters was necessary because the relation between the different signal strengths of the experimental recruitment data sets could not be determined.

Curriculum Vitae

Daniel Löb

Mittelweg 8
64720 Michelstadt, Germany
+49 160 91236780
daniel@danielloeb.eu

Education

Doctoral program at Darmstadt University of Technology (TUD) Theory of Complex Systems Group, Prof. Barbara Drossel.	since Dec. 2008
Graduation as Master of Science, Physics at TUD	Nov. 2008
Master Thesis: "Radiation-Hydrodynamics Simulation of Laser-Plasma Interactions" in the Laser and Plasma Physics Group of Prof. Markus Roth at TUD and the Society for Heavy Ion Research (GSI), Darmstadt.	Nov. 2007 - Nov. 2008
Completion of Master classes at TUD	Apr. 2007 - Oct. 2007
ERASMUS stay with studies in applied Mathematics and Physics at Escuela Técnica Superior de Ingenieros Industriales, Universidad Politécnica de Madrid in Madrid, Spain	Sep. 2006 - Mar. 2007
Bachelor of Science, Physics at TUD, thesis topic: "Investigations on the Ablation of Laser-heated Carbon Foils"	Okt. 2003 - Aug. 2006

Publications

Frank Tobias, Daniel Löb, Nicor Lengert, Marco Durante, Barbara Drossel, Gisela Taucher-Scholz, Burkhard Jakob, *Spatiotemporal Dynamics of Early DNA Damage Response Proteins on Complex DNA Lesions*, PLOS ONE **8**, e57953 (2013) http://dx.doi.org/10.137/journal.pone.0057953

Conference Contributions

D. Löb, M. Ditter (2011) *Modelling the Antonine Plague - Epidemics on the Road and Sea Network of the Roman Empire* European Conference on Complex Systems, September 12.-16. 2011 in Vienna (Talk)

D. Löb, B. Drossel (2011) *Modelling the focus formation after double-strand breaks caused by ionizing radiation* European Conference on Complex Systems, September 12.-16. 2011 in Vienna (Poster)

D. Löb, B. Drossel (2011) *A Stochastic Model of DNA Replication Dynamics* Verhandlungen der Deutschen Physikalischen Gesellschaft e.V. BP 4.4, German Physical Society SMK, KR, ST, SOE and AIW Divisions Spring Meeting, March 13.-18. 2011 in Dresden (Talk)

D. Löb, M. Ditter (2011) *Disease Dynamics on the Road and Sea Network of the Roman Empire* Verhandlungen der Deutschen Physikalischen Gesellschaft e.V. SOE 22.2, German Physical Society SMK, KR, ST, SOE and AIW Divisions Spring Meeting, March 13.-18. 2011 in Dresden (Talk)

G. Weiss, D. Löb, B. Drossel (2011) *Modelling the recruitment of DNA repair enzymes* Verhandlungen der Deutschen Physikalischen Gesellschaft e.V. BP 9.6, German Physical Society SMK, KR, ST, SOE and AIW Divisions Spring Meeting, March 13.-18. 2011 in Dresden (Poster)

D. Löb, B. Drossel (2010) *Modelling protein accumulation at DNA damage sites* Verhandlungen der Deutschen Physikalischen Gesellschaft e.V. BP 32.37, German Physical Society SMK, KR, ST, SOE and AIW Divisions Spring Meeting, March 21.-26. 2010 in Regensburg (Poster)

D. Löb, M. Roth (2008) *Simulation-based Analysis of Interferometry Data on Laser-heated Plasmas* Proceedings of the 30. European Conference on Laser Interaction with Matter, P015, TUD, GSI, EMMI, IAEA, Laser and Particle Beams, August 31. - September 5. 2008 in Darmstadt (Poster)

D. Löb, A. Tauschwitz, A. Pelka, M. Roth (2008) *Simulationen lasererzeugter Plasmen mit MULTI* Verhandlungen der Deutschen Physikalischen Gesellschaft e.V., PIII.27, German Physical Society Atoms, Molecules, Quantum-Optics and Plasma Division Spring Meeting, March 10.-14. 2008 in Darmstadt (Poster)

Research and Teaching Experience

Research assistant in the Theory of Complex Systems Group at TUD as a part of the doctoral program.	since Dec. 2008
Held a two-week preparational course in applied mathematics for physics freshmen.	Apr. 2009
Administrator of the distributed computer cluster (30 PCs) at the theory division of the Institute for Solid State Physics at TUD.	since Dec. 2008
Tutored problems class for the course “Theoretical Physics III”.	Nov. 2008 - Mar. 2009
Tutored problems class for the course “Physics for Biologists”.	Oct. 2007 - Mar. 2008

Summer Schools

Summer University for Plasma Physics IPP at the Max Planck Institut for Plasma Physics in Greifswald, Germany.	Sep. 2007

Computer Skills

C++, Python, Perl, various other Programming Languages,
Linux (Administrator Level)

Languages

native:	German
fluent:	English, Spanish
moderate:	French
basic:	Finnish, Russian (UNIcert I), Italian, Portuguese