

Fakultät für Physik und Astronomie  
Ruprecht-Karls-Universität Heidelberg

Diplomarbeit  
im Studiengang Physik  
vorgelegt von  
Daniel Huebschmann  
aus Heidelberg  
2005



# Simulation früh- und spätreplizierender Chromatin-Domänen

Diese Diplomarbeit wurde von Daniel Huebschmann ausgeführt  
am Kirchhoff-Institut für Physik  
unter der Betreuung von  
Herrn Prof. Dr. Dr. Christoph Cremer



**Simulation früh- und spätreplizierender Chromatin-Domänen** Fluoreszenzmikroskopische Untersuchungen zeigten, daß das Chromatin eines Zellkerns in der Interphase mehreren Kompaktierungsstufen unterworfen ist. Auf diesen Erkenntnissen beruht das *Spherical 1-Mbp Chromatin Domain* (SCD) Modell, in dem jedes Chromosom durch eine lineare Kette von 1 Megabasenpaar großen Domänen angenähert wird. In dieser Arbeit werden der Bandentyp (spät replizierende C- und G-Banden und früh replizierende R-Banden) der Chromatindomänen und spezielle für das splicing zuständige Proteinkomplexe, die SC35-Domänen, in das SCD-Modell integriert. Wechselwirkungen zwischen den SC35-Domänen und den Chromatindomänen verschiedenen Bandentyps werden modelliert. Dadurch wird zum ersten Mal ein funktioneller Zusammenhang zwischen Proteinkomplexen und Chromatindomänen im SCD-Modell realisiert. Es werden das Kontaktverhalten von G- und R-Bandendomänen mit speckles und die Verteilung von G- und R-Bandendomänen auf der Oberfläche und im Inneren der Chromosomenterritorien untersucht. Ein Vergleich mit experimentellen Daten ermöglicht die Bestimmung der Stärke der Wechselwirkung zwischen R-Banden und speckles. Durch die Reduktion der Laufzeitordnung einer Funktion, welche in jedem Monte Carlo Schritt aufgerufen wird, wird die Monte Carlo Relaxation von Zellkernkonfigurationen im SCD-Modell beschleunigt.

**Simulation of early and late replicating chromatin domains** According to investigations by fluorescence microscopy on cell nuclei, chromatin is subject to different levels of compaction in the interphase. Based on these findings the *Spherical 1-Mbp Chromatin Domain* (SCD) model was developed. In this model every chromosome is approximated by a linear chain of domains of the size of 1 mega-basepair. In the present work the type of the chromatin domains (late replicating C- and G-Bands and early replicating R-Bands) and protein complexes responsible for splicing, the SC35-domains, are integrated into the SCD-model. The interactions between the chromatin domains of different band type and the SC35-domains are modelled. For the first time a functional relationship between protein complexes and chromatin domains is implemented in the SCD-model. The contacts between domains of different type and the speckles and the distribution of G-band domains and R-band domains on the surface and the interior of the chromosome territories are analyzed. A comparison with experimental data opens the possibility to determine the strength of the interaction between R-bands and speckles. The Monte Carlo relaxation of configurations of cell nuclei in the SCD-model is accelerated by a reduction of the order of a function called at every Monte Carlo step.



# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>5</b>
<b>2</b>	<b>Biologische Grundlagen</b>	<b>7</b>
2.1	Zellbiologie . . . . .	7
2.1.1	Desoxyribonucleinsäure - DNS . . . . .	7
2.1.2	Replikation und Transkription . . . . .	9
2.1.3	Chromatin . . . . .	9
2.1.4	Zellzyklus . . . . .	11
2.2	Fluoreszenzmarkierungsverfahren . . . . .	12
2.3	Höhere Chromatinorganisationsformen . . . . .	13
2.3.1	Metaphase . . . . .	13
2.3.2	Interphase . . . . .	13
2.4	Bandenmuster . . . . .	14
2.5	Speckles und SC35-Domänen . . . . .	15
2.6	Das Interphase Chromatin Domain (ICD) Modell . . . . .	16
<b>3</b>	<b>Mathematische und physikalische Grundlagen</b>	<b>19</b>
3.1	Polymerphysik . . . . .	19
3.1.1	Die ideale Kette . . . . .	19
3.1.2	Die Abstandsverteilungsfunktion . . . . .	19
3.1.3	Der Gyrationradius eines Polymers . . . . .	20
3.1.4	Nicht-ideale Ketten . . . . .	21
3.1.5	Entropische Federkraft . . . . .	21
3.1.6	Elastizitätstheorie . . . . .	22
3.2	Monte Carlo Methoden . . . . .	23
3.2.1	Simple Sampling . . . . .	23
3.2.2	Importance Sampling . . . . .	24
<b>4</b>	<b>Das SCD-Modell</b>	<b>26</b>
4.1	Modellierung und Simulation . . . . .	26
4.1.1	Der Chromatin-Linker . . . . .	26
4.1.2	Excluded Volume Wechselwirkung . . . . .	27
4.1.3	Tensegrity Force Wechselwirkung . . . . .	28
4.1.4	Monte-Carlo Relaxation . . . . .	29
4.1.5	Visualisierung . . . . .	29
4.2	Erweiterungen . . . . .	30
4.2.1	Generierung von Startkonfigurationen für Fibroblasten . . . . .	30
4.2.2	Speckles und SC35-Domänen . . . . .	31
4.2.3	Zuweisung des Bandentyps für Chromatindomänen . . . . .	32
4.2.4	Wechselwirkungen zwischen den Domänen im Zellkern . . . . .	35
4.2.5	Modifikation der Funktion updateposition . . . . .	36
4.2.6	Algorithmus für das Auffinden von clustern . . . . .	38

4.2.7	Auffinden der Oberfläche eines Territoriums . . . . .	39
<b>5</b>	<b>Ergebnisse</b>	<b>46</b>
5.1	Parametervariation . . . . .	46
5.2	Bestimmung einer unteren Schranke für die Größe der speckles . . . . .	50
5.3	Relaxationen mit den ausgewählten Parameterwerten . . . . .	52
5.4	Geometrische Merkmale und Größen . . . . .	55
5.4.1	Gyrationsradien spezieller Genregionen . . . . .	55
5.4.2	Oberfläche und Inneres der Territorien . . . . .	56
5.4.3	Auswertung von $GBandsSrf2Int$ und $RBandsSrf2Int$ für einzelne Chromosomen . . . . .	58
5.4.4	Auswertung von $G2RRatio$ für einzelne Chromosomen . . . . .	61
<b>6</b>	<b>Diskussion und Ausblick</b>	<b>63</b>
6.1	Diskussion . . . . .	63
6.1.1	Weiterentwicklung des SCD-Modells . . . . .	63
6.1.2	Variation der Stärke der Wechselwirkung zwischen speckles und R-Banden $\epsilon_1$ . . . . .	64
6.1.3	Kontaktverhalten von früh- und spätreplizierenden Chromatindomänen . . . . .	64
6.1.4	Geometrische Eigenschaften der Chromosomenterritorien . . . . .	65
6.2	Ausblick . . . . .	66

# Abbildungsverzeichnis

2.1	DNS–Doppelhelix . . . . .	8
2.2	Chromatin . . . . .	10
2.3	Zellzyklus einer eukaryotischen Zelle . . . . .	11
2.4	<i>Radial loop / Scaffold</i> Modell . . . . .	14
2.5	Ideogramm . . . . .	15
2.6	SC35-Domänen . . . . .	16
2.7	Schematische Darstellung des ICD Modells . . . . .	18
3.1	Bestimmung der Elastizität des Chromatin–Linkers . . . . .	22
4.1	Tensegrity Force Wechselwirkung . . . . .	28
4.2	Startkonfigurationen mit 30 speckles . . . . .	32
4.3	Startkonfiguration mit 500 speckles am Rand des Zellkerns . . . . .	32
4.4	teilrelaxierte Konfiguration mit 500 speckles am Rand des Zellkerns . . . . .	33
4.5	Relaxierte Kerne . . . . .	38
4.6	Oberfläche und Inneres der Territorien 1 und 2 . . . . .	45
5.1	Domänenkontakt . . . . .	46
5.2	Parametervariation . . . . .	48
5.3	Parametervariation bei kleineren speckles . . . . .	51
5.4	Schematische Darstellung der untersuchten Genregionen . . . . .	52
5.5	Virtuelle Mikroskopie: G-Banden der Chromosomen 7 und 17 . . . . .	54
5.6	Virtuelle Mikroskopie: R-Banden der Chromosomen 7 und 17 . . . . .	54
5.7	Verteilung von R- und G-Banden auf Oberfläche und Inneres der Territorien . . . . .	59
5.8	Differenz RSrf2Int-GSrf2Int . . . . .	59
5.9	Verhältnis von G-Banden-zählraten zu R-Banden-zählraten . . . . .	61
5.10	Differenz G2RInt-G2RSrf . . . . .	61

# Tabellenverzeichnis

5.1	Kriterien I und II . . . . .	52
5.2	Kriterien III und IV . . . . .	53
5.3	Kriterium V . . . . .	55
5.4	Gyrationsradien . . . . .	55
5.5	Verteilung von R- und G-Banden auf Oberfläche und Inneres aller Territorien im Zellkern . . . . .	57
5.6	Verteilung von R- und G-Banden auf Oberfläche und Inneres der Territorien, für jedes Chromosom einzeln aufgelistet . . . . .	60
5.7	Verhältnis von G-Bandenanzahlraten zu R-Bandenanzahlraten . . . . .	62

# Kapitel 1

## Einleitung

Lichtmikroskopische Untersuchungen an fluoreszent markierten Chromatinabschnitten haben gezeigt, daß das Chromatin eines Zellkerns nicht nur während der Mitose, sondern auch in der Interphase eine spezifische Organisation aufweist. Die Chromosomen nehmen Bereiche begrenzten Volumens ein, die Chromosomenterritorien. Darüber hinaus ist eine weitere Kompartimentalisierung des Chromatins in 300 bis 800 nm große Domänen, auch Foci genannt, beobachtet worden. Diese haben einen DNA-Gehalt von etwa  $10^6$  Basenpaaren (1 Mega-Basenpaar, Mpb).

Gregor Kreth entwickelte anhand von diesen Beobachtungen in [Kre00] ein Computermodell für Interphase-Zellkerne, das *Spherical 1 Mbp Chromatin Domain* (SCD) Modell. In diesem Modell werden die Foci bzw. Domänen durch Kugeln mit einem Durchmesser von 500 nm dargestellt. Jedes Chromosom besteht aus einer linearen Kette solcher Domänen. Die Wirkung des Chromatinlinkers zwischen zwei Domänen wird in der Simulation durch ein entropisches Federpotential modelliert. Jeder Domäne wird ein DNA-Gehalt von 1 Mbp zugeschrieben, wobei die Anzahl der Domänen eines Chromosoms proportional zu seinem DNA-Gehalt ist. Ausgehend von einer Startkonfiguration werden Zellkernkonfigurationen entweder mit dem Verfahren der Brown'schen Dynamik oder mit einem Monte Carlo Verfahren relaxiert.

In dieser Arbeit werden ausgehend von einer Ideogrammdarstellung eines menschlichen Chromosomensatzes die einzelnen Domänen im SCD-Modell verschiedenen Banden zugeordnet: den spätreplizierenden und in der Giemsa-Färbung dunklen G-Banden, den frühreplizierenden und in der Giemsa-Färbung hellen R-Banden und den Zentromerbanden (C-Banden).

Weiterhin werden Proteinkomplexe, sogenannte *SC35-Domänen* oder *speckles* in das SCD-Modell integriert. In diesen findet das mRNA-splicing statt. Sie sind 0.5 bis 3  $\mu\text{m}$  groß und kompakt. Im SCD-Modell werden sie durch Kugeln mit einem Durchmesser von 1.5  $\mu\text{m}$  dargestellt. Die SC35-Domänen wechselwirken je nach Bandentyp unterschiedlich mit den Chromatindomänen. Die Wechselwirkung zwischen SC35-Domänen und G- und C-Banden wird als rein repulsiv angenommen, die Wechselwirkung zwischen SC35-Domänen und R-Banden wird mit einem repulsiven und einem attraktiven Teil modelliert.

Damit wird zum ersten Mal ein funktioneller Zusammenhang zwischen Proteinkomplexen und Chromatindomänen im SCD-Modell realisiert. Es kann der Einfluß dieser Proteinkomplexe auf die Kernstruktur quantitativ untersucht werden.

Außerdem wird eine Funktion, welche bei der Monte Carlo Relaxation nach der Verrückung einer Domäne die neue Position in die Zellkernkonfiguration schreibt, überarbeitet. Es wird eine Reduktion der Laufzeitordnung von quadratisch auf linear erreicht, was zu einer Reduktion der für eine Relaxation benötigten Rechenzeit auf ein Elftel der ursprünglichen Rechenzeit führt. Dies macht numerische Studien in großem Umfang möglich.

Bei der Auswertung relaxierter Zellkernkonfigurationen werden verschiedene Zahlenwerte berechnet und mit experimentellen Daten von L. Shopland et al. (vergleiche [SJB<sup>+</sup>03]) verglichen. Es werden die Kontakte zwischen G-Banden und SC35-Domänen und zwischen R-Banden und SC35-Domänen, sowohl für den ganzen Zellkern, als auch für spezielle Genregionen untersucht.

Es wird die Coassoziaton zweier syntenischer<sup>1</sup> Genregionen analysiert (Coassoziaton zwischen zwei Genregionen A und B besteht dann, wenn mindestens eine SC35-Domäne zu mindestens einer Chromatindomäne der Genregion A und mindestens einer der Genregion B simultan Kontakt hat). Schließlich wird die Verteilung von früh- und spätreplizierenden Domänen auf der Oberfläche und im Inneren der Chromosomenterritorien untersucht.

---

<sup>1</sup>auf demselben Chromosom liegender

# Kapitel 2

## Biologische Grundlagen

Jedes Lebewesen besteht aus Zellen. Hierbei handelt es sich um membranbegrenzte Einheiten, die wässrige Lösungen verschiedener chemischer Substanzen sowie Suborganellen enthalten. Die einfachsten Lebewesen bestehen aus nur einer Zelle, höhere Organismen bestehen aus Zellverbänden, welche mit einander kommunizieren und kooperieren. In einem solchen Organismus haben verschiedene Zellen verschiedene Aufgaben und ergänzen sich.

Zellen sind in der Lage, durch Zweiteilung Kopien ihrer selbst herzustellen.

### 2.1 Zellbiologie

Typische Zellen haben einen Durchmesser von 5 bis 20  $\mu m$ . Sie sind erst mit einem Mikroskop zu beobachten. Wenige Zellen sind so groß, daß man sie mit bloßem Auge sehen kann, so hat z.B. die Eizelle eines Krallenfroschs einen Durchmesser von 1  $mm$ . Alle Zellen sind umgeben von einer Membran, der **Plasmamembran**. Innerhalb dieser befindet sich das **Zytosol**, ein wässriges Gel. Die verschiedenen Organellen in einer Zelle sind ihrerseits durch Membranen vom Zytosol abgegrenzt. Weitere Informationen in [ABJ<sup>+</sup>99].

Zellen unterscheiden sich bezüglich der Anzahl und Art der Organellen, die sich in ihrem Zytosol befinden. Man teilt alle Zellen in zwei Gruppen ein: **Eukaryonten** sind Zellen, welche ein Zellorganell Zellkern enthalten; **Prokaryonten** sind Zellen, welche keinen Zellkern enthalten. Die Begriffe Prokaryont und Bakterium werden synonym verwandt, hierbei handelt es sich ausschließlich um einzellige Organismen. Zu den Eukaryonten gehören einzellige Organismen wie Hefen und Amöben, aber auch alle vielzelligen Lebewesen wie Pflanzen und Tiere.

#### 2.1.1 Desoxyribonucleinsäure - DNS

Der Zellkern wird durch die **Kernmembran** vom Zytoplasma abgegrenzt. Sie enthält Kernporen, durch die bestimmte Moleküle selektiv passieren können. Im Zellkern befindet sich das genetische Material der Zelle in Form von DNS-Molekülen<sup>1</sup>. Ein solches Molekül besteht aus zwei langen Polynukleotidketten. Jede dieser Ketten wird als DNA-Kette oder DNA-Strang bezeichnet. Ein Nukleotid, das ist das Monomer der Polynukleotidkette, besteht aus einem Zucker mit fünf Kohlenstoffatomen, der Desoxyribose. An diesen Zucker sind eine Phosphatgruppe und eine der vier stickstoffhaltigen Nukleinbasen **Adenin (A)**, **Cytosin (C)**, **Guanin (G)** und **Thymin (T)** gebunden. Adenin und Guanin sind Purinbasen, welche sich vom heterozyklischen Molekül Purin ableiten, Cytosin und Thymin sind Pyrimidinbasen, welche sich vom ebenfalls heterozyklischen Pyrimidin ableiten. Jede Untereinheit ist vollständig durch ihre Base charakterisiert, die Nukleinbasen stellen das Alphabet der Erbinformation dar.

---

<sup>1</sup>Der englische Ausdruck ist *desoxyribonucleic acid* (DNA). DNA und DNS werden im folgenden synonym gebraucht

Über die Zucker und Phosphate sind die Nukleotide kovalent mit einander zu einer Kette verknüpft. Die beiden DNA-Stränge eines DNA-Moleküls werden von **Wasserstoffbrückenbindungen** zwischen den Basen der Nukleotide zusammengehalten. Hierbei nehmen die beiden Polynukleotidketten die räumliche Anordnung einer **Doppelhelix** ein (siehe Abbildung 2.1). Alle Basen befinden sich im Inneren der Helix, während die Zucker-Phosphatgerüste nach außen zeigen. Zwischen Adenin und Thymin können sich zwei Wasserstoffbrücken ausbilden, zwischen Guanin und Cytosin drei. Diese beiden **Basenpaare** sind energetisch sowie aufgrund der Konformation am günstigsten, daher kommen nur sie und keine anderen Kombinationen zwischen den Nukleinbasen in der DNA vor. Es ist also immer eine Purinbase mit einer Pyrimidinbase gepaart.

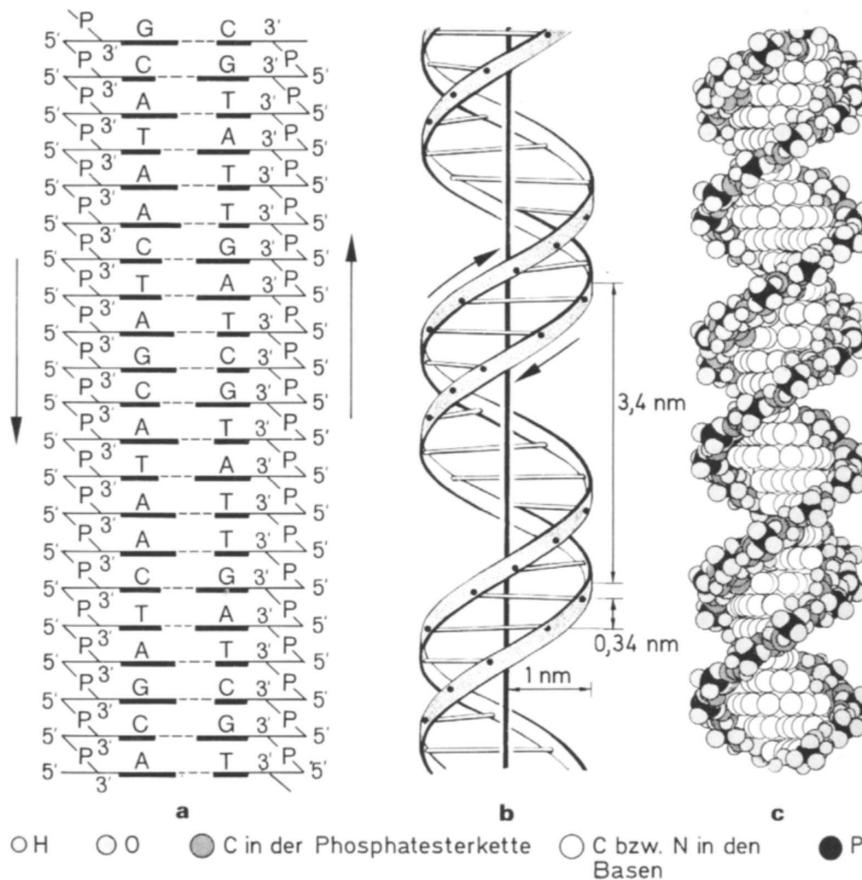


Abbildung 2.1: Schematische Darstellung der DNS-Doppelhelix. **a**: antiparallele Anordnung beider komplementären Polynukleotidstränge; **b**: Windung der Doppelhelix: eine Windung umfaßt 10 Basenpaare, die Höhe beträgt 3.4 nm **c**: Kalottenmodell der DNS (aus Knippers [KPSF90])

Jeder der beiden DNA-Stränge besitzt eine **Polarität**. Ein Ende eines Stranges wird als 3'-Ende bezeichnet, das andere als 5'-Ende. Diese Nomenklatur leitet sich von den Kohlenstoffatomen in der Desoxyribose ab, welche an der kovalenten Bindung zwischen den Nukleotiden beteiligt sind. In der Doppelhelix verlaufen die beiden Stränge antiparallel, d.h. sie haben entgegengesetzte Orientierung. Eine wichtige Konsequenz aus den oben beschriebenen Charakteristika ist, daß die beiden Stränge eines DNA-Moleküls **komplementär** sind. Dies bedeutet, daß ein Strang bereits die komplette Information des gesamten Moleküls trägt; ebenso kann aus einem Einzelstrang wieder ein komplettes Molekül aufgebaut werden, indem Schritt für Schritt die komplementären, d.h. die in das jeweilige Basenpaar passenden Basen, an einen wachsenden Strang hinzugefügt werden.

### 2.1.2 Replikation und Transkription

Die biologische Aufgabe der DNA ist die Speicherung der Erbinformation einer Zelle. Die Zelle muß nun zwei große Probleme lösen: einerseits muß die Erbinformation bei einer Teilung an Tochterzellen vollständig und korrekt weitergegeben werden, andererseits muß die Erbinformation ausgelesen und in biochemische Prozesse umgesetzt werden, damit die Zelle die zu ihrem Überleben nötigen und unter Umständen die in einem Zellverband erforderlichen Prozesse umsetzen kann. Die Lösungsstrategien aller Zellen heißen: Replikation und Translation.

**Replikation** bedeutet Verdopplung der Erbinformation. Die Replikation wird bewerkstelligt von der Replikationsmaschinerie im Zellkern. Hierbei handelt es sich um eine Zusammenarbeit mehrerer Proteinkomplexe, der wichtigste Bestandteil ist die DNA-Polymerase. Dabei wird die DNA-Doppelhelix an bestimmten Stellen entwunden, die Wasserstoffbrückenbindungen zwischen den Nucleinbasen werden gelöst und es werden komplementäre Stränge zu den bereits bestehenden angelegt.

Unter **Transkription** versteht man ein Umschreiben der in der DNA kodierten Information auf ein anderes Molekül, die *Ribonukleinsäure*, abgekürzt RNS, englisch *ribonucleic acid*, abgekürzt **RNA**. Die RNA ist ähnlich aufgebaut wie die DNA, es gibt jedoch folgende Unterschiede: der Zucker im Zucker-Phosphat-Gerüst ist Ribose und nicht Desoxyribose, anstatt der stickstoffhaltigen Base Thymin wird die ebenfalls stickstoffhaltige Base Uracil eingebaut, welche jedoch wie Thymin bevorzugt mit Adenin über Wasserstoffbrücken gepaart ist. Darüber hinaus besteht RNA aus nur einem Strang, welcher sich allerdings in eine räumliche Struktur falten kann und dabei Wasserstoffbrückenbindungen mit sich selbst eingeht. Das wichtigste Protein in der Transkriptionsmaschinerie ist die RNA-Polymerase. RNA-Moleküle, welche direkt aus der Transkription hervorgegangen sind, heißen Transkripte. Man unterscheidet nun zwischen verschiedenen Transkripten, je nach dem, wofür die RNA bestimmt ist:

- Ribosomale RNA - **rRNA** ist ein Baustein von Ribosomen. Ribosomen sind Komplexe, welche aus Proteinen und RNA-Molekülen bestehen und sind verantwortlich für die Synthese aller Proteine aus Aminosäuren. Die DNA, welche für rRNA codiert, ist zu Suborganellen des Zellkerns geclustert, den Nukleoli.
- Transfer RNA - **tRNA** stellt einen Adapter dar, der eine Aminosäure auswählt und im Ribosom an der richtigen Stelle zum Einbau in ein Protein bereithält.
- Messenger RNA - **mRNA** oder Boten-RNA. Diese Moleküle dienen den Ribosomen als Baupläne für Proteine. Hierbei wird die Nucleinbasensequenz mit Hilfe des genetischen Codes in eine Aminosäuresequenz übersetzt. Dieser Vorgang wird **Translation** genannt.

In Eukaryonten enthält jede mRNA normalerweise die Information von einem Gen, einem Abschnitt der DNA, welcher für genau ein Protein kodiert. In Bakterien werden häufig eine Reihe benachbarter Gene in eine mRNA transkribiert, die damit Information über mehrere verschiedene Proteine enthält. In beiden Fällen sind jedoch die mRNA-Moleküle wesentlich kürzer als die entsprechenden DNA-Moleküle, dasselbe gilt für rRNA- und tRNA-Moleküle. Innerhalb eines Gens auf der DNA befinden sich Abschnitte, welche keine Informationen für die Proteinsynthese enthalten. Auf der transkribierten mRNA finden sich diese Abschnitte wieder, sie werden Introns genannt (die codierenden Abschnitte werden Exons genannt). Bevor die mRNA-Transkripte zur nächsten Station, der Translation, gelangen, werden aus ihnen die nicht-codierenden Abschnitte herausgeschnitten. Dieser Vorgang wird **Spleißen**<sup>2</sup> genannt.

Anschließend werden die RNA Moleküle aus dem Zellkern durch Kernporen in das Zytoplasma exportiert, wo sich die Ribosomen befinden und wo deshalb die Translation stattfindet.

### 2.1.3 Chromatin

Ein DNA-Molekül erfährt verschiedene Packungs- und Kompaktierungsstufen (siehe auch 2.3). Diese Packung wird ermöglicht durch Proteine und Proteinkomplexe, an welche die DNA assoziiert

<sup>2</sup>englisch: splicing

ist. Die DNA zusammen mit den assoziierten Proteinen wird **Chromatin** (griechisch *chroma*, Farbe) genannt.

Die erste Packungsstufe ist das **Nukleosom**. Ein solches besteht aus acht **Histon**proteinen, welche im Zentrum des Nukleosoms ein Histonoktamer oder den Histonkern bilden, und einem Abschnitt DNA-Doppelhelix von ungefähr 200 Basenpaaren. Ein ungefähr 146 Basenpaare langer Abschnitt windet sich um den Histonkern und bildet mit diesem ein Nukleosomenkernpartikel. Der Rest ist die Linker-DNA, welche zwei Nukleosomenkernpartikel verbindet. Chromatin, welches in dieser Packungsstufe vorliegt, z.B. nachdem die höheren Organisations- und Packungsstrukturen zerstört worden sind, ist im Elektronenmikroskop als Perlenkette zu sehen (siehe Abbildung 2.2, linke Teilabbildung). Die Histonproteine sind in sehr großer Menge im Chromatin vorhanden, ihre Gesamtmasse entspricht ungefähr der der DNA selbst.

Die Nukleosomen sind ihrerseits dicht aneinander gepackt und bilden die **30-nm-Faser**, die Größenangabe bezieht sich auf ihren Durchmesser (siehe Abbildung 2.2, rechte Teilabbildung). Auch hier ist ein Histonprotein an der Bildung der Struktur beteiligt, allerdings nicht als Oktamer. Auch die 30-nm-Faser ist im Elektronenmikroskop sichtbar, nicht jedoch im klassischen Lichtmikroskop. Sie hat eine Packungsdichte von 6 Nukleosomen pro 11 nm; bei einem Durchmesser von 30 nm führt das, wenn man eine erneute helikale Wicklung annimmt, zu 6 Nucleosomen pro Windung (Solenoidmodell von Finch & Klug 1976). Neuere Untersuchungen legen jedoch eine unregelmäßigere Anordnung nahe.

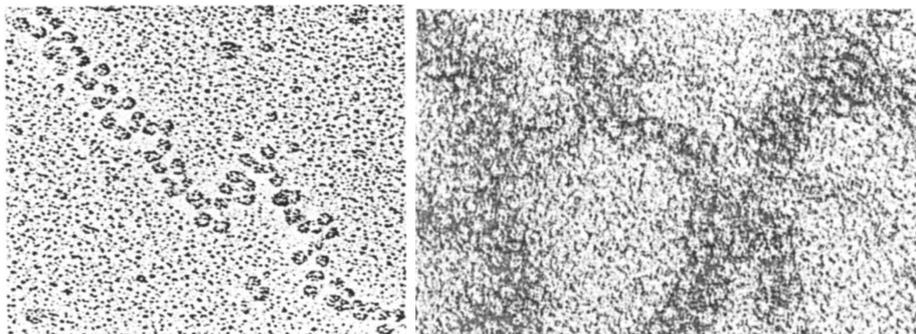


Abbildung 2.2: Elektronenmikroskopische Aufnahmen von Chromatin **a**: bei einer Salzkonzentration von 5 mM (*Beads on a String*) und **b**: bei physiologischen Bedingungen ( $\sim 145$  mM, 30 nm Chromatinfaser) (aus Voet & Voet [VV95])

Als Produkt mehrerer weiterer Packungsstufen eines DNA-Moleküls entsteht ein **Chromosom**, welches seinen höchsten Kompaktierungsgrad in der Mitose (siehe 2.1.4) hat. Alle eukaryontischen Zellen haben mindestens zwei Chromosomen. Prokaryonten hingegen haben nur ein einziges sogenanntes Chromosomenäquivalent, welches ringförmig ist und sich im Zytoplasma befindet.

Fast alle Zellen komplexer Organismen, welche sich geschlechtlich fortpflanzen, haben in ihrem Zellkern zwei Kopien jedes Chromosoms, eines vom Vater und eines von der Mutter. Solche Zellen nennt man **diploid**. Das maternale und das paternale Chromosom eines Paares werden **homologe Chromosomen** genannt. Die einzigen nichthomologen Chromosomen in Zellen von Spezies, welche sich geschlechtlich fortpflanzen, sind die Geschlechtschromosomen bei den männlichen Individuen, welche ein Y-Chromosom vom Vater und ein X-Chromosom von der Mutter geerbt haben.

Die Gameten, i.e. Spermien und Eizellen, die Zellen also, welche für die geschlechtliche Fortpflanzung verantwortlich sind, haben nur einen einfachen Chromosomensatz, sie werden **haploid** genannt. Prokaryontische Zellen sind ebenfalls haploid.

Diploide menschliche Zellkerne haben 46 Chromosomen, davon 22 homologe Chromosomenpaare<sup>3</sup> und zwei Geschlechtschromosomen.

<sup>3</sup>auch *Autosomen* genannt

### 2.1.4 Zellzyklus

Wenn eine Zelle sich teilt, durchläuft sie bei diesem Prozess mehrere aufeinander folgende Phasen (vergleiche Abbildung 2.3). Man unterscheidet die M-Phase,  $G_1$ -Phase, S-Phase und  $G_2$ -Phase. Diese werden zusammen als Zellzyklus bezeichnet. In den Zyklus eingebettet sind verschiedene Kontrollpunkte, an denen der Zyklus angehalten oder verzögert werden kann. So kann eine Zelle beispielsweise zu Anfang der  $G_1$ -Phase in eine Ruhephase, die  $G_0$ -Phase, treten und schreitet dann nicht weiter im Zellzyklus fort.

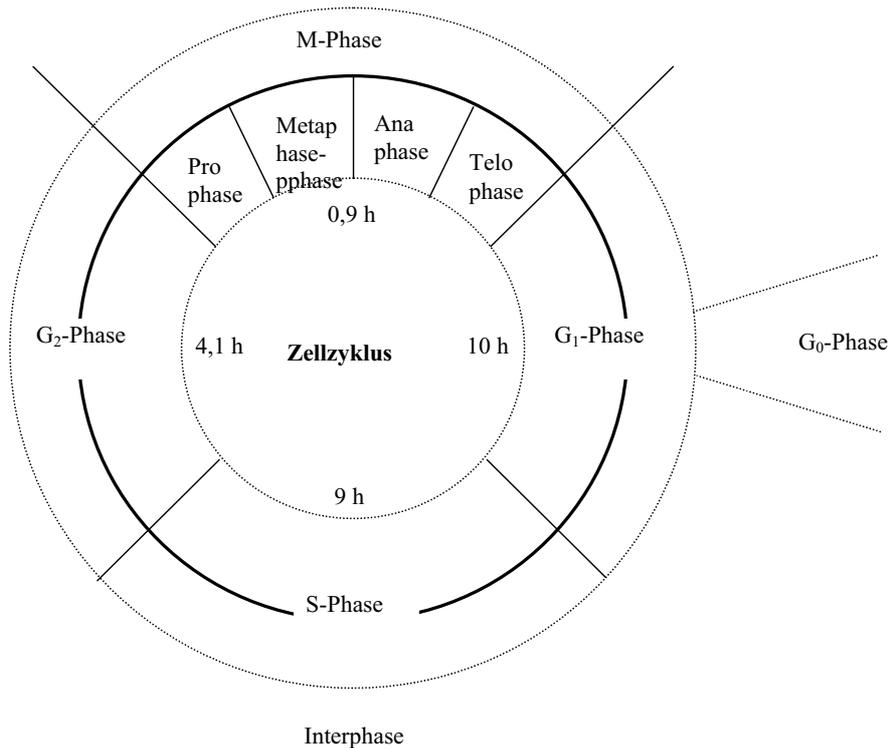


Abbildung 2.3: Der Zellzyklus einer eukaryotischen Zelle. Dargestellt ist ein Zyklus von 24 h, er kann jedoch zwischen 10 und 30 h variieren. Die  $G_1$ -Phase variiert dabei am meisten in der Zeit. Ruhende Zellen verharren in der  $G_1$ -Phase: diesen Zustand bezeichnet man auch als  $G_0$ -Phase.

- Die **M-Phase** besteht aus der Mitose und der Cytokinese. Die Mitose ist die Teilung des Zellkerns, die Cytokinese die Teilung der restlichen Zelle. Die **Mitose** ihrerseits wird in verschiedene Phasen eingeteilt:
  - Die in der S-Phase (s.u.) des vorhergehenden Zellzyklus replizierten Chromosomen, welche aus zwei eng miteinander assoziierten Schwesterchromatiden bestehen, kondensieren in der **Prophase** zu einer kompakten Struktur. Die aus Mikrotubuli bestehende mitotische Spindel wird außerhalb des Zellkerns zwischen ihren beiden Polen, den Centrosomen aufgebaut.
  - Die Auflösung der Kernmembran stellt den Beginn der **Prometaphase** dar. Die Chromosomen heften sich an die mitotische Spindel und erfahren eine aktive Bewegung.
  - Die Chromosomen sind in der Äquatorialebene der Spindel angekommen. In der **Meta-  
phase** orientieren sie sich so, daß je ein Schwesterchromatid zu einem der beiden Pole der mitotischen Spindel weist.

- In der **Anaphase** trennen sich die gepaarten Chromatiden synchron und bilden zwei Tochterchromosomen, welche langsam zu den entsprechenden Spindelpolen gezogen werden.
- Während der **Telophase** kommen die beiden Tochterchromosomensätze an den Spindelpolen an. Es bildet sich um jeden Satz wieder eine neue Kernhülle. Die Mitose ist damit abgeschlossen.

Während der **Cytokinese** wird das Cytoplasma in zwei Hälften geteilt. Hierzu bedient sich die Zelle des kontraktilen Rings, einer Struktur aus Aktin- und Myosinfilamenten, welche sich an der Plasmamembran zwischen den beiden neuen Kernen ausbildet. Der kontraktile Ring schnürt die Zelle ein und läßt auf diese Weise zwei Tochterzellen entstehen. Kleine Zellorganellen, welche in großer Zahl vorhanden sind, werden zufällig auf die Tochterzellen verteilt, große Organellen zerfallen in eine Vielzahl an Membranvesikeln, welche ihrerseits dann auf die Tochterzellen verteilt werden.

- Der Ausdruck  **$G_1$ -Phase** kommt vom englischen “gap” und bezeichnet den Teil des Zellzyklus, welcher sich zwischen M-Phase und S-Phase befindet. In der  $G_1$ -Phase befindet sich ein wichtiger Kontrollpunkt, an dem der Zellzyklus angehalten oder verzögert werden kann und der somit die Möglichkeit einer Koordination des zeitlichen Ablaufs des gesamten Zellzyklus ermöglicht.
- Während der **S-Phase** wird die DNS im Kern repliziert, d.h. verdoppelt. Dies ist nötig, damit in der nächsten M-Phase die Teilung ordentlich ablaufen kann.
- Auch die  **$G_2$ -Phase** leitet sich vom englischen “gap” ab und bezeichnet den Teil des Zellzyklus, der sich zwischen S-Phase und M-Phase befindet. Hier befindet sich ein weiterer Kontrollpunkt, der ebenfalls der Koordination des Zellzyklus dient.

$G_1$ -Phase, S-Phase und  $G_2$ -Phase werden zusammenfassend als Interphase bezeichnet.

## 2.2 Fluoreszenzmarkierungsverfahren

Gelingt es, eine bestimmte chemische Substanz in einer Zelle, im folgenden *Ziel* genannt, spezifisch zu markieren, so läßt sich mit dieser Markierung der Aufenthaltsort und die Verteilung der markierten Substanz ermitteln. Zur Markierung verwendet man sogenannte Sonden. Es handelt sich dabei um Moleküle, welche spezifisch an die zu markierende Zielsubstanz binden und welche als eigentliche Marker entweder radioaktive Isotope oder fluoreszierende Molekülgruppen, sogenannte Fluorochrome, tragen.

Mit dem Verfahren der *in situ Hybridisierung*, ISH, können bestimmte Abschnitte der DNA markiert werden. Die Sonde ist ihrerseits eine Polynukleotidkette. Trägt sie als Marker einen Fluoreszenzfarbstoff, so spricht man von der *Fluoreszenz in situ Hybridisierung*. Bei diesem Verfahren geht man folgendermassen vor:

1. Man mischt die Sonde der Ziel-DNA bei.
2. Ziel-DNA und Sonde werde gemeinsam und gleichzeitig *denaturiert*, das heißt in ihre Einzelstränge zerlegt. Dies geschieht entweder durch Erhitzen oder durch drastische Änderungen des pH-Werts der Lösung.
3. Bei der nun folgenden *Hybridisierung*, das ist die Renaturierung, lagern sich komplementäre Nukleotide, unter anderem der Ziel- und der Sonden-DNA, aneinander. Da auch Ziel-DNA mit komplementärer Ziel-DNA und Sonden-DNA mit komplementärer Sonden-DNA rekombinieren kann, erhält man ein Gemisch aus diesen drei Kombinationen.

Die ISH erlaubt Markierungen mit sehr hoher Spezifität. Man unterscheidet weiterhin zwischen einer direkten und einer indirekten Markierung:

- Bei der direkten Markierung werden Nukleotide mit Fluoreszenzfarbstoffen in die Sonden-DNA eingebaut. Mit der Hybridisierung ist die (Fluoreszenz-) Markierung der Ziel-DNA im Genom abgeschlossen.
- Bei der indirekten Markierung werden Nukleotide mit Reportermolekülen in die Sonden-DNA eingebaut. Die Fluoreszenzmarkierung erfolgt in einem oder mehreren Schritten nach der Hybridisierung. Es kann hier ein stärkeres Signal als bei der direkten Markierung erreicht werden.

Die mit Fluoreszenzfarbstoffen markierten Proben können mit Hilfe der Fluoreszenzmikroskopie untersucht werden. Dabei werden die Farbstoffe mit einem Laserstrahl zur Fluoreszenz angeregt, das Fluoreszenzlicht wird von der Mikroskopoptik aufgefangen und dem Betrachter zugeführt. Durch den Einsatz verschiedener Techniken, zum Beispiel verschiedener Farben, lassen sich Abstände und Größen messen, welche teilweise wesentlich kleiner als die optische Auflösung eines Lichtmikroskops sind.

Mit Hilfe von Fluoreszenzmarkierungsverfahren und Fluoreszenzmikroskopie wurden neue Untersuchungen von Struktur und Anordnung von Zellorganellen und sogar einzelnen Molekülen, vor allem von DNA-Abschnitten, möglich.

## 2.3 Höhere Chromatinorganisationsformen

Die Kompaktierung der DNA in Nukleosomen und deren Packung zu einer Faser sind Grundlage aller Modellvorstellungen über höhere Chromatinorganisationsformen, deren genaue Details und Struktur Gegenstand intensiver Forschung sind. Diese höheren Organisationsformen müssen so angelegt sein, daß sie die wesentlichen Veränderungen und die Hauptaufgaben des Chromatins ermöglichen und begünstigen: **Replikation** und **Transkription** (vergleiche 2.1.1).

Je nachdem, wo sich eine Zelle im Zellzyklus befindet, weist ihr Chromatin unterschiedliche Packungs- und Organisationsformen auf. Es werden im folgenden die beiden Extremfälle beschrieben, zwischen denen jeder Punkt im Zellzyklus liegt: die höchste Kompaktierung in der Metaphase, und eine quasi relaxierte Form der Chromosomen in der Interphase.

### 2.3.1 Metaphase

In der Metaphase besteht ein Chromosom aus zwei annähernd zylinderförmigen Schwesterchromatiden mit einem Durchmesser von ungefähr  $0.8\mu\text{m}$  und einer Länge zwischen 2 und  $10\mu\text{m}$ . Sie sind im Lichtmikroskop gut sichtbar. Die Verbindungsstelle beider Chromatiden wird Centromer genannt, wodurch das Chromosom in einen kürzeren p- und einen längeren q-Arm unterteilt wird. Jedes Schwesterchromatid besteht aus jeweils *einem* durchgehenden DNA-Faden. Für die Beschreibung des hohen Kondensationsgrades der DNA im Metaphase-Chromosom - der DNA-Faden ist ungefähr  $10^5$ -fach gepackt - schlugen Marsden & Laemmli 1979 das *radial loop / scaffold* - Modell vor (siehe Abbildung 2.4). Es wird angenommen, daß jedes Chromatid aus einem zentralen nicht histonischen Proteingerüst, dem sogenannten *Scaffold*, besteht, an welches DNA-Schleifen mit einer Länge von 30 bis 120 kBp gebunden sind.

### 2.3.2 Interphase

Das Chromatin von Interphasechromosomen liegt nicht in allen Bereichen im gleichen Verpackungsgrad vor. Einige Regionen, vor allem diejenigen, welche gerade in RNA transkribiert werden, sind stärker entfaltet oder lockerer als andere, vor allem transkriptionell inaktive Bereiche. Die Struktur eines Interphasechromosoms unterscheidet sich also von Zelle zu Zelle, je nachdem welche Gene exprimiert werden.

Die am stärksten kondensierte Form des Interphasechromatins wird als **Heterochromatin** bezeichnet (von griechisch *heteros*, verschieden; Heterochromatin läßt sich stark anfärben und ist deshalb im Lichtmikroskop deutlicher zu sehen als die umgebende Kernmaterie). Ungefähr 10 %

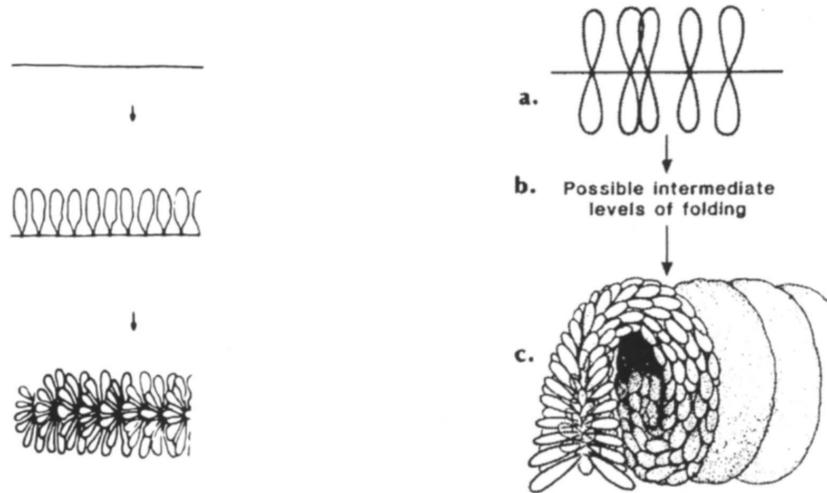


Abbildung 2.4: Schematische Darstellung des *Radial loop / Scaffold* Modells **Links:** 30 nm Chromatinfaser (dünne Linie) wird mit Hilfe nicht historischer Proteine zu 30–120 kbp großen Schleifen gebunden. Die Basispunkte der Schleifen sind entlang der Achse des Chromatids am Scaffold befestigt (aus Marsden & Laemmli [ML79]). **Rechts:** zusätzliche helikale Struktur des Scaffolds (aus Rattner & Lin [RL85])

eines Interphasechromosoms sind normalerweise Heterochromatin, hierzu zählen die Centromerregionen und die Enden der Chromosomen. Heterochromatin ist wie Mitosechromatin transkriptionell inaktiv.

Das übrige Interphasechromatin, welches verschiedene lockere Formen umfaßt, wird als **Euchromatin** (von griechisch *eu*, wahr, normal) bezeichnet. In einer typischen ausdifferenzierten eukaryontischen Zelle sind ungefähr 10 % dieses Chromatins in einem aktiv transkribierten oder leicht aktivierbaren Zustand. Dieser stellt die am wenigsten kondensierte Form des Chromatins dar.

In den 70er Jahren des 20. Jahrhunderts wurden Experimente entwickelt, um die Anordnung von Chromosomen in Interphasezellkernen zu untersuchen. Es wurden kleine Bereiche in Zellkernen mit einem fokussierten UV-Laserstrahl bestrahlt [ZCCZ79], [CCB+82]. In der nachfolgenden Metaphase zeigte sich, daß nur wenige Chromosomen bzw. chromosomale Abschnitte eine Schädigung durch die Bestrahlung erfahren hatten. Dies deutete darauf hin, daß ein einzelnes relaxiertes Chromosom nicht über den ganzen Zellkern verteilt ist, sondern daß eine territoriale Organisation besteht. Man spricht daher von **Chromosomenterritorien**. Diese wurden 1988 mit Hilfe der *Fluoreszenz in situ Hybridisierung* (vergleiche 2.2) in einem Fluoreszenzmikroskop direkt visualisiert [CLB+88], [LCB+88], [PLC+88] (ein Überblick findet sich in Cremer et al. [CKZ+93]). Mit weiterentwickelten Markierungsverfahren konnte dann eine weitere Kompartimentalisierung der Chromosomenterritorien in kleinere Aggregate, die **Chromatindomänen** oder **Foci**, gezeigt werden [NMS86], [NB89], [JP98].

## 2.4 Bandenmuster

Färbt man Chromatin mit einem Farbstoff an, so bilden sich im allgemeinen positive (gefärbte) und negative (nicht gefärbte) Bereiche. Benutzt man speziell die Giemsa-Färbung, so erhält man eine Struktur, bei der die Grenzen zwischen positiven und negativen Bereichen in der Mitose senkrecht zur Richtung des betreffenden Chromosoms verlaufen. Man nennt diese Banden. Es sind nun folgende Bandentypen zu unterscheiden:

- Giemsa-dunkle Banden, abgekürzt **G-Banden**; diese enthalten wenig Nukleotide mit den

Basen Guanin und Cytosin (sie sind GC-arm). Auf ihnen befinden sich hauptsächlich Gene, die nur in bestimmten Geweben oder in bestimmten Phasen der Entwicklung exprimiert werden. In einem ausdifferenzierten Zelltyp sind näherungsweise 97% der in G-Banden lokalisierten Gene transkriptionsinaktiv

- Giemsa-helle Banden<sup>4</sup>, abgekürzt **R-Banden**; enthalten viele Nukleotide mit den Basen Guanin und Cytosin (GC-reich). Sie sind genreicher als G-Banden und enthalten Gene zur Aufrechterhaltung der Zellstruktur und des Zellstoffwechsels (sogenannte Haushaltsgene), aber auch gewebespezifische Gene. R-Banden sind dem Euchromatin zuzuordnen.
- Wendet man die Färbetechnik nach einer Denaturierung mit anschließender Renaturierung an, so erhält man Centromer-Banden, abgekürzt **C-Banden**. Sie sind AT-reich, werden dem konstitutiven Heterochromatin zugeordnet und enthalten repetitive, transkriptionsinaktive DNA-Sequenzen.

In der S-Phase replizieren zuerst die R-Banden, dann die G-Banden und in der zweiten Hälfte der S-Phase die C-Banden. Das Bandenmuster eines Chromosoms (oder auch des gesamten Genoms) wird in einem sogenannten Ideogramm dargestellt (siehe Abbildung 2.5).

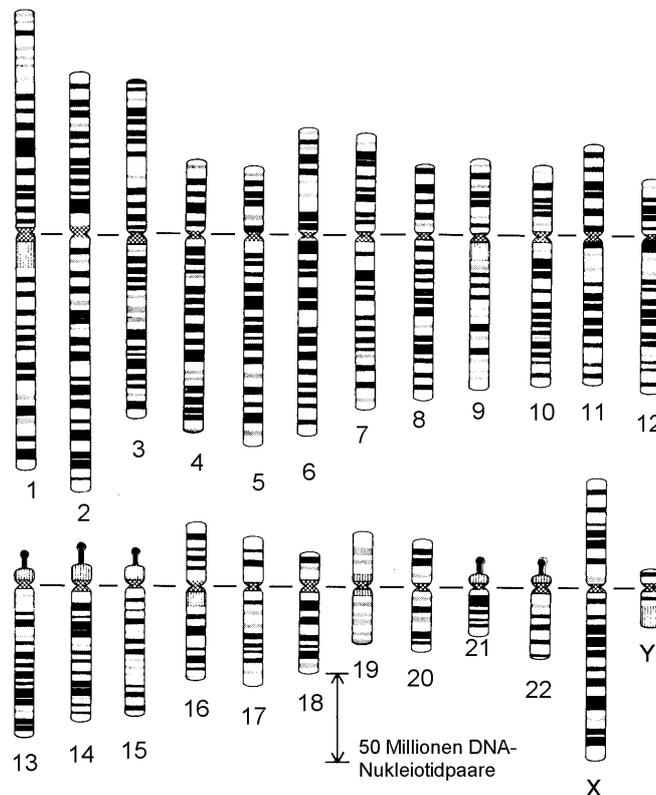


Abbildung 2.5: Ideogramme aller Chromosomen des Menschen

## 2.5 Speckles und SC35-Domänen

Im Zellkern einer eukaryontischen Zelle sind mehrere metabolische Faktoren (biochemisch aktive Substanzen), darunter preRNA-splicing Faktoren, nicht homogen über das ganze Kernvolumen verteilt, sondern treten an 20 bis 40 Stellen im Kern konzentriert auf. Im Rest des Nukleoplasmas

<sup>4</sup>englisch *reverse* Bands, daher die Abkürzung R

sind sie zwar vorhanden sind, aber in sehr kleinen Konzentrationen. Die Stellen hoher Konzentration werden als *Speckles*<sup>5</sup> bezeichnet.

Einer dieser splicing Faktoren ist der *spliceosome assembly factor* SC35. Speckles, die ihn enthalten, werden *SC35-domains*<sup>6</sup> oder *splicing factor compartments*<sup>7</sup> (SFCs) genannt. Diese sind diskret, irregulär geformt und können mit einem Antikörper gegen SC35 visualisiert werden (siehe Abbildung 2.6). Sie haben einen Durchmesser von 0.5 bis 3  $\mu\text{m}$  und enthalten außer SC35 noch weitere splicing Faktoren, SR Proteine, poly(A) RNA, poly(A) RNA bindendes Protein II, hyperphosphorylierte RNA-Polymerase II, Lamine und Faktoren, welche in den Export von RNA verwickelt sind. Alle diese Faktoren sind ebenfalls nicht ausschließlich in den SC35-Domänen zu finden, sondern auch diffus über das Nukleoplasma verteilt. Obwohl die SC35-Domänen relativ immobil sind, das heißt ihre Positionen verändern sich wenig, sind sie dynamische Strukturen in der Hinsicht, daß die in ihnen

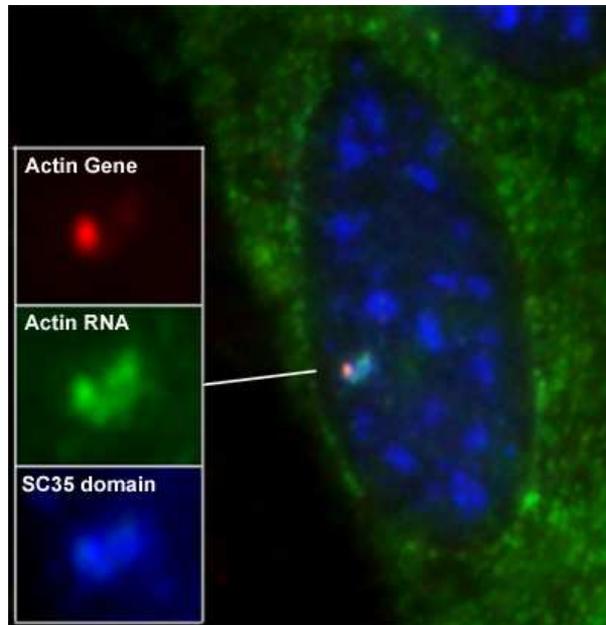


Abbildung 2.6: rot: Actin-Gen; grün: Actin-RNA; blau: SC35-Domänen; Aufnahme von <http://redbone.umassmed.edu/SC35.html>

enthaltenen Faktoren einer von ihrem Phosphorylierungszustand abhängigen Strömung unterworfen sind. Es wird geschätzt, daß bis zu 150 verschiedene Proteine zumindest vorübergehend mit diesen Domänen kolokalisieren (Informationen von <http://redbone.umassmed.edu/SC35.html>).

SC35-Domänen enthalten mit großer Wahrscheinlichkeit keine DNA. Sie befinden sich überwiegend in inneren, euchromatischen Regionen des Zellkerns. Es wird vermutet, daß sie Reservoirs für splicing-Komponenten sind, welche dann zu nahegelegenen aktiven Genen, in deren Nähe splicing stattfindet, geleitet werden. Experimentelle Daten zeigen, daß SC35-Domänen Strukturen sind, um die mehrere mRNA-codierende Gene clustern, und die die Transkripte dieser Gene enthalten. Es ist möglich, daß diese *euchromatic neighborhoods*<sup>8</sup> eine höhere Organisationsform der protein-codierenden Gene darstellen, analog zu der Clusterung von rRNA-codierender DNA zu Nukleoli.

## 2.6 Das Interphase Chromatin Domain (ICD) Modell

Die Sichtbarmachung der territorialen Chromosomenorganisation in Säugerzellen 1988 ([CLB<sup>+</sup>88], [LCB<sup>+</sup>88], [PLC<sup>+</sup>88]) mit Hilfe der Fluoreszenzmikroskopie führte 1993 zur Vorstellung eines Modells der Interphaseorganisation: das *Interphase Chromatin Domain (ICD) Compartment* Modell [CKZ<sup>+</sup>93].

Dieses Modell ordnet den Bereichen zwischen den Territorien (dem sogenannten ICD-Raum) eine entscheidende funktionelle Bedeutung bei der Abwicklung zellkernspezifischer Prozesse, wie mRNA Transport, DNA Replikation, Transkription oder Reparaturmechanismen zu. Die Oberflächen der Chromosomenterritorien wurden deshalb als bevorzugte Bereiche für die Lokalisation aktiver Chromatinabschnitte angesehen. Ausgehend von weiteren Untersuchungen wurde später

<sup>5</sup>deutsch: Flecken

<sup>6</sup>deutsch: SC35-Domänen

<sup>7</sup>deutsch: Spleiß-Faktor-Kompartimente

<sup>8</sup>deutsch: euchromatische Nachbarschaften

[CKK<sup>+</sup>00] eine weitere Kompartimentalisierung der Chromosomarmterritorien in 1 Mbp große Domänen angenommen, die wiederum aus etwa 10 – 100 kbp großen Domänen bestehen (siehe Abbildung 2.7). Diese 100-kbp Domänen könnten durch Faltung der Chromatinfaser zu Schleifen entstehen, vorstellbar wären aber auch Konglomerate aus unregelmäßig gefalteten Chromatin-, bzw. Nukleosomenkettenabschnitten, die durch *Tensegrity Force*<sup>9</sup> Wechselwirkungen zusammengehalten werden [CKK<sup>+</sup>00], siehe auch Abschnitt 4.1.3. Durch Änderung der Anzahl bzw. der Stärke dieser Wechselwirkungen bei den 100-kbp Domänen wird beim erweiterten ICD Modell ein Konzept der *offenen* bzw. *geschlossenen* Domänen postuliert [CKK<sup>+</sup>00]. Danach sind aufgrund starker *Tensegrity Force* Wechselwirkungen abgeschlossene Domänen weitgehend genetisch inaktiv, da die Transkriptionsfaktor-Bindungsstellen räumlich nicht zugänglich sind. Bei offenen Domänen sind aktive Chromatinabschnitte untereinander zugänglich und damit genetisch aktiv.

Ausgehend von Kernporen erstreckt sich der **ICD Raum** zwischen allen Kompartimentalisierungen der Chromosomenterritorien. Damit wird angenommen, daß die chromatinfreien Bereiche zwischen den 1-Mbp bzw. 100-kbp Domänen kleiner als die Bereiche zwischen den Chromosomenterritorien sind. Nach diesem Modell sind aktive Chromatinabschnitte demnach auch innerhalb von Chromosomenterritorien in den 100-kbp Domänen lokalisiert, entsprechend dem Konzept *offener* und *geschlossener* Domänen. Zur Aufrechterhaltung eines minimalen ICD Raumes von einigen nm wurde beim ICD Modell außerdem eine abstoßende elektrostatische Wechselwirkung zwischen den Oberflächen der Chromosomenterritorien postuliert, die durch die nur teilweise kompensierten negativen Ladungen des Phosphatrückgrats der DNS hervorgerufen werden [CKZ<sup>+</sup>93]. Solche Potentialverläufe wurden bei Messungen an menschlichen Tumor T-Zell Lymphozyten (JURKAT), bei denen die Spannung im Zellkern in Abhängigkeit von der beweglichen Ladungsträgerkonzentration bestimmt wurde, gefunden [CKK<sup>+</sup>00].

1974 wurden von Berezney & Coffey [BC74] eine aus Proteinen bestehende **Kernmatrix** im Zellkern sichtbar gemacht. Sie wurde als der unlösliche Rest betrachtet, der nach einer Reihe von biochemischen Extraktionsschritten im Zellkern zurückbleibt. Für Abläufe zellkernspezifischer Prozesse (z.B. Transkription, Reperaturprozesse und DNS Replikation) wird dieser Matrix eine ähnlich bedeutende Funktion wie dem ICD Raum beim ICD Modell zugeordnet. Unklar ist jedoch wie eine derart komplexe, funktionell bedeutsame 3D-Struktur nach jedem Zellzyklus wieder neu gebildet wird.

---

<sup>9</sup>Kunstwort im Englischen, welches sich zusammen setzt aus *tension* und *integrity*; bezeichnet ein System, in dem sich Strukturen durch Druck und Spannung selbst stabilisieren

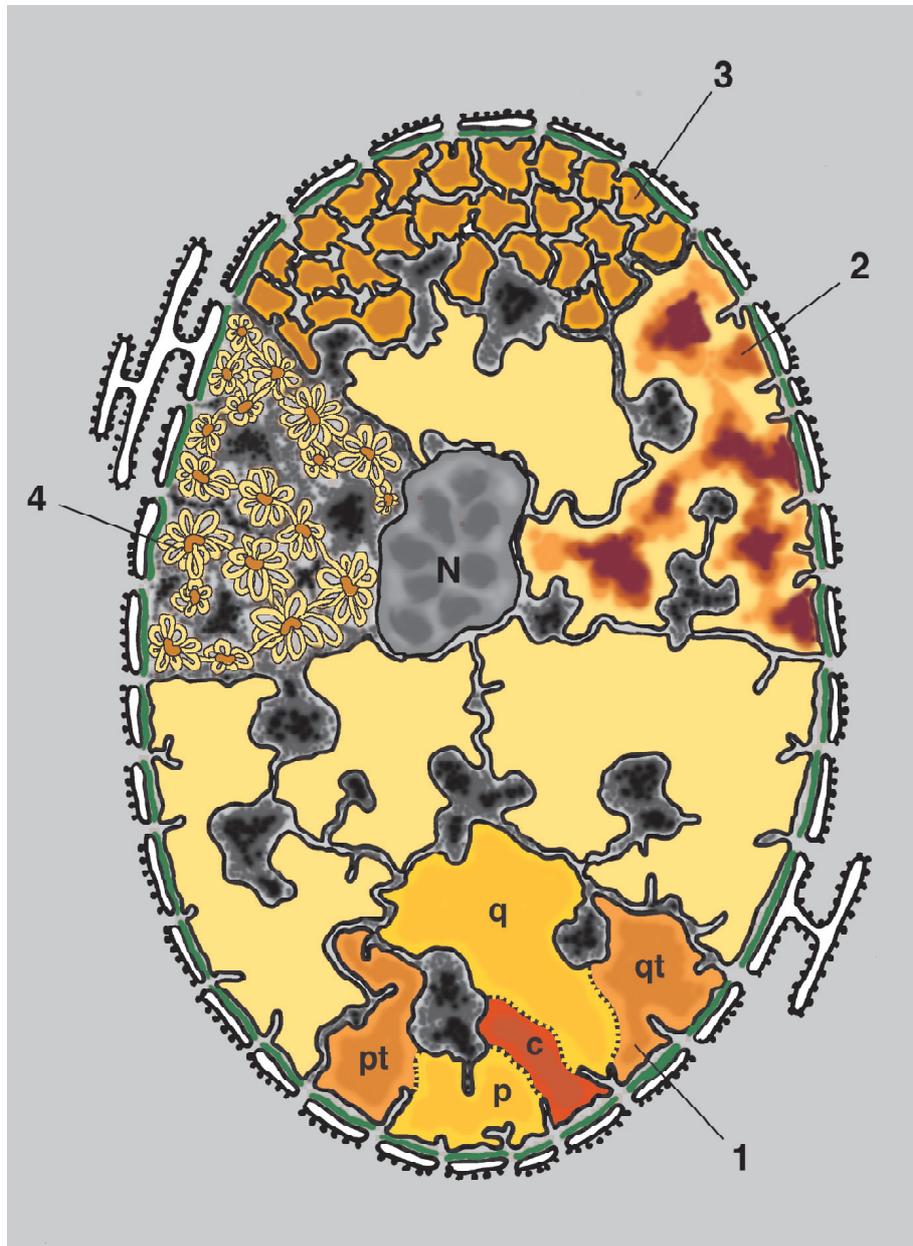


Abbildung 2.7: Schematische Darstellung des erweiterten ICD Modells [CKK<sup>+</sup>00]. Jedes Chromosomenterritorium (2) ist danach in weitere Untereinheiten kompartmentalisiert, die wiederum eigene voneinander abgegrenzte Bereiche oder Domänen einnehmen. Die Chromosomenterritorien (1) bestehen demnach aus 1-Mbp Domänen (3), die im Durchmesser 300 bis 800 nm groß und durch Chromatin-Linker miteinander verbunden sind. Diese sind wiederum aus den bis jetzt noch hypothetischen ~ 100 kbp Domänen aufgebaut, welche beim MLS Modell [ML98] z.B. den Schleifen entsprechen (4). Der ICD Raum erstreckt sich zwischen all diesen Untereinheiten, wobei seine Größe durch den Kompartimentalisierungsgrad bestimmt wird (von Dr. I. Solovei, München).

# Kapitel 3

## Mathematische und physikalische Grundlagen

### 3.1 Polymerphysik

Ein Polymer ist ein großes Molekül, welches aus vielen kleinen, chemisch einfachen Einheiten besteht, welche kovalent mit einander verbunden sind. Die Anzahl dieser Einheiten wird Polymerisationsgrad  $N$  genannt. Viele Polymere sind repetitiv, das heißt die kleinen Grundeinheiten, aus denen sie aufgebaut sind, sind alle gleich. Eine solche Grundeinheit wird Monomer oder Monomeinheit genannt. Für nähere Informationen siehe [Doi97] oder [Str97].

Ein Molekül wird üblicherweise Polymer genannt, wenn sein Polymerisationsgrad größer als 100 ist. Synthetisch hergestellte Polymere bestehen aus bis zu  $10^5$  Einheiten. Die DNA, welche auch ein Polymer ist, hat einen Polymerisationsgrad von  $10^7$ , es gibt sogar natürlich vorkommende Moleküle mit einem Polymerisationsgrad von über  $10^9$ .

#### 3.1.1 Die ideale Kette

Im folgenden wird ein Polymer betrachtet, welches aus  $N$  geraden Segmenten der Länge  $b$  besteht. Um die Verbindungen zwischen den Segmenten seien Drehungen erlaubt. Das  $n$ -te Segment sei durch den Vektor  $\mathbf{r}_n$  gekennzeichnet. Der End-zu-End-Vektor dieses Polymers ist  $\mathbf{R} = \sum_{n=1}^N \mathbf{r}_n$ . Allerdings ist der Mittelwert  $\langle \mathbf{R} \rangle$  gleich Null, da sich Beiträge von  $\mathbf{R}$  und  $-\mathbf{R}$  gegenseitig aufheben. Andererseits gilt für den quadratischen End-zu-End-Abstand:

$$\langle \mathbf{R}^2 \rangle = \sum_{n=1}^N \sum_{m=1}^N \langle \mathbf{r}_n \cdot \mathbf{r}_m \rangle \quad (3.1)$$

In der Modellvorstellung einer idealen Kette bestehen keine Wechselwirkungen zwischen verschiedenen Segmenten, daher gibt es auch keine Korrelation zwischen verschiedenen Segmentvektoren und es gilt für  $n \neq m$ :  $\langle \mathbf{r}_n \cdot \mathbf{r}_m \rangle = \langle \mathbf{r}_n \rangle \cdot \langle \mathbf{r}_m \rangle = 0$ . Daher ist:

$$\langle \mathbf{R}^2 \rangle = \sum_{n=1}^N \langle \mathbf{r}_n^2 \rangle = Nb^2 \quad (3.2)$$

#### 3.1.2 Die Abstandsverteilungsfunktion

Mit  $P(\mathbf{R}, N)$  werde im folgenden die Wahrscheinlichkeit bezeichnet, daß das Ende des Polymers mit  $N$  Segmenten an der Stelle  $\mathbf{R}$  liegt. Die Wahrscheinlichkeitsverteilung kann unter der Annahme, daß das Polymer auf einem Gitter der Koordinationszahl  $z$  liegt, folgendermaßen hergeleitet werden: seien  $\{\mathbf{b}_i, i = 1 \dots z\}$  die möglichen Segmentvektoren, die das Polymer annehmen kann.

In einem Polymer aus  $N$  Segmenten, das zum Ort  $\mathbf{R}$  reicht, nimmt das  $(N - 1)$ -te Segment eine der Positionen  $\mathbf{R} - \mathbf{b}_i$  ein, wovon jede die Wahrscheinlichkeit  $\frac{1}{z}$  hat. Daher gilt für die Wahrscheinlichkeit:

$$P(\mathbf{R}, N) = \frac{1}{z} \sum_{i=1}^z P(\mathbf{R} - \mathbf{b}_i, N - 1). \quad (3.3)$$

Eine Taylorentwicklung unter den Voraussetzungen daß  $N \gg 1$  und  $|\mathbf{R}| \gg |\mathbf{b}_i|$  ergibt:

$$P(\mathbf{R} - \mathbf{b}_i, N - 1) = P(\mathbf{R}, N) - \frac{\partial P}{\partial N} - \frac{\partial P}{\partial R_\alpha} b_{i\alpha} + \frac{1}{2} \frac{\partial^2 P}{\partial R_\alpha \partial R_\beta} b_{i\alpha} b_{i\beta}. \quad (3.4)$$

Hierbei sind  $b_{i\alpha}$  und  $R_\alpha$  die Komponenten von  $\mathbf{b}_i$  und  $\mathbf{R}$ , außerdem wurde die Einsteinsche Summationskonvention verwendet. Es gilt nun:

$$\frac{1}{z} \sum_{i=1}^z b_{i\alpha} = 0 \text{ und } \frac{1}{z} \sum_{i=1}^z b_{i\alpha} b_{i\beta} = \frac{\delta_{\alpha\beta} b^2}{3}. \quad (3.5)$$

Einsetzen von 3.4 in 3.3 unter Berücksichtigung von 3.5 ergibt die Differentialgleichung

$$\frac{\partial P}{\partial N} = \frac{b^2}{6} \frac{\partial^2 P}{\partial \mathbf{R}^2}. \quad (3.6)$$

Unter der Bedingung, daß  $\mathbf{R} = 0$  falls  $N = 0$  ist die Lösung dieser Differentialgleichung:

$$P(\mathbf{R}, N) = \left( \frac{3}{2\pi N b^2} \right)^{\frac{3}{2}} \exp\left( -\frac{3\mathbf{R}^2}{2N b^2} \right). \quad (3.7)$$

Dies ist das typische Verhalten eines *random walk* (RW). Läßt man die Annahme, daß das Polymer auf einem Gitter liegt, wieder fallen und geht statt dessen direkt von einer Wahrscheinlichkeitsverteilung der Art 3.7 aus, so gelangt man zum Modell der Gauß'schen Kette, welche aufgrund der oben gemachten Annahmen  $N \gg 1$  und  $|\mathbf{R}| \gg |\mathbf{b}_i|$  ein Spezialfall der idealen Kette ist.

### 3.1.3 Der Gyrationradius eines Polymers

Der Gyrationradius eines Polymers ist definiert durch:

$$R_g^2 = \frac{1}{2N^2} \sum_{n=1}^N \sum_{m=1}^N \langle (\mathbf{R}_n - \mathbf{R}_m)^2 \rangle. \quad (3.8)$$

Mit dem Massenschwerpunkt des Polymers  $\mathbf{R}_G = \frac{1}{N} \sum_{n=1}^N \mathbf{R}_n$  kann man umformen:

$$R_g^2 = \frac{1}{N} \sum_{n=1}^N \langle (\mathbf{R}_n - \mathbf{R}_G)^2 \rangle \quad (3.9)$$

das heißt der Gyrationradius ist gleich dem mittleren Abstand zwischen einem Segment und dem Massenschwerpunkt des Polymers. Er ist Teil des ersten Koeffizienten einer Taylorentwicklung der Fouriertransformierten der Paarkorrelationsfunktion  $g(\mathbf{r}) = \frac{1}{N} \sum_{n=1}^N \sum_{m=1}^N \langle \delta(\mathbf{r} - (\mathbf{R}_m - \mathbf{R}_n)) \rangle$  des Polymers. Daher ist er in Streuexperimenten direkt meßbar und stellt eine praktische Möglichkeit dar, die Größe eines Polymers zu quantifizieren (siehe [Doi97], Seite 8).

Der Gyrationradius einer Gauß'schen Kette läßt sich wie folgt berechnen:

$$\begin{aligned} \langle (\mathbf{R}_n - \mathbf{R}_m)^2 \rangle &= |n - m| b^2 \\ R_g^2 &= \frac{1}{2N^2} \sum_{n=1}^N \sum_{m=1}^N |n - m| b^2 \\ &\doteq \frac{b^2}{2N^2} \int_0^N dn \int_0^N dm |n - m| = \frac{1}{6} N b^2 \end{aligned} \quad (3.10)$$

### 3.1.4 Nicht-ideale Ketten

Zwischen den Segmenten eines Polymers können verschiedene Wechselwirkungen bestehen.

Bei *kurzreichweitigen Wechselwirkungen* gilt für den quadratischen End-zu-End-Abstand (für eine Herleitung vergleiche [Doi97], Seite 4):  $\langle \mathbf{R}^2 \rangle = Nb_{eff}^2$ , ebenso gilt für den Gyrationradius:

$$R_g = N^{\frac{1}{2}} b_{eff} \quad (3.11)$$

Die Länge  $b$  eines Segments muß ersetzt werden durch eine effektive Länge  $b_{eff}$ . An den statistischen Eigenschaften ändert sich jedoch nichts,  $R_g$  bleibt proportional zu  $N^{\frac{1}{2}}$  und die Abstandsverteilung bleibt gaußförmig. 'Kurzreichweitig' bezieht sich hier auf die Entfernung zweier Segmente entlang des Polymers. Kurzreichweitige Wechselwirkungen können in das Modell der idealen Kette integriert werden.

Wechselwirkungen, welche nur von der Entfernung zweier Segmente im Raum, nicht aber entlang des Polymers abhängen, werden *langreichweitige Wechselwirkungen* genannt. Unter ihrer Berücksichtigung erhält man andere statistische Eigenschaften. So führt der *excluded volume effect*, der zwei Segmenten verbietet, sich gegenseitig zu durchdringen, zu folgendem Verhalten des Gyrationradius:

$$R_g \propto N^\nu b_{eff} \quad (3.12)$$

Der Parameter  $\nu$ , welcher nicht analytisch berechnet werden kann, hat im Falle der *excluded volume*-Wechselwirkung den Wert 0.588, wie numerische Simulationen ergaben. Darüber hinaus ist die Abstandsverteilung nicht mehr gaußförmig.

Eine wichtige Größe für die Charakterisierung der Steifheit eines Polymers ist die Kuhn'sche Segmentlänge  $b_K$ . Sie ist definiert durch:

$$b_K = \frac{\langle \mathbf{R}^2 \rangle}{L_{max}} \quad (3.13)$$

Hierbei ist  $L_{max}$  die Konturlänge des Polymers, das ist die Länge der voll ausgestreckten Kette:  $L_{max} = N \cdot b$ . Für ideale Ketten gilt nun:  $b_K = \frac{b_{eff}^2}{b}$  bzw.  $b_K = b$ , falls es keine kurzreichweitigen Wechselwirkungen gibt.

### 3.1.5 Entropische Federkraft

Im folgenden soll die Streckelastizität eines Polymers bestimmt werden unter der Annahme, daß sie allein auf die thermischen Bewegungen der Segmente des Polymers zurückzuführen ist. Dazu wird angenommen, daß von einer Kette aus  $N$  Segmenten der Länge  $b$  ein Ende im Ursprung festgehalten wird, während am anderen Ende gezogen wird.

Die Anzahl an idealen Ketten  $W(\mathbf{R})dR$ , welche mit  $N$  Segmenten einen End-zu-End-Vektor zwischen  $\mathbf{R}$  und  $\frac{\mathbf{R}}{|\mathbf{R}|} \cdot (R + dR)$  erreichen, ist  $P(\mathbf{R}, N)dR$ . Daher ist:

$$W(\mathbf{R}) = P(\mathbf{R}, N) = \left( \frac{3}{2\pi N b^2} \right)^{\frac{3}{2}} \exp \left( -\frac{3\mathbf{R}^2}{2N b^2} \right) \quad (3.14)$$

Die freie Energie der idealen Kette ist dann:

$$A = -k_B T \ln(W(\mathbf{R})) + A_0 = \frac{3k_B T}{2N b^2} \mathbf{R}^2 + const + A_0 \quad (3.15)$$

$A_0$  ist eine Konstante, die den Nullpunkt der Energieskala festlegt. Auf das Ende der Kette wirkt daher die Spannkraft:

$$\mathbf{F} = -\nabla A = -\frac{3k_B T}{N b^2} \mathbf{R} \quad (3.16)$$

Die Ursache dieser Spannkraft ist nicht in einer Änderung der Energie, sondern allein in einer Änderung der Entropie zu sehen. Wenn die Enden einer Polymerkette gestreckt werden, wird die

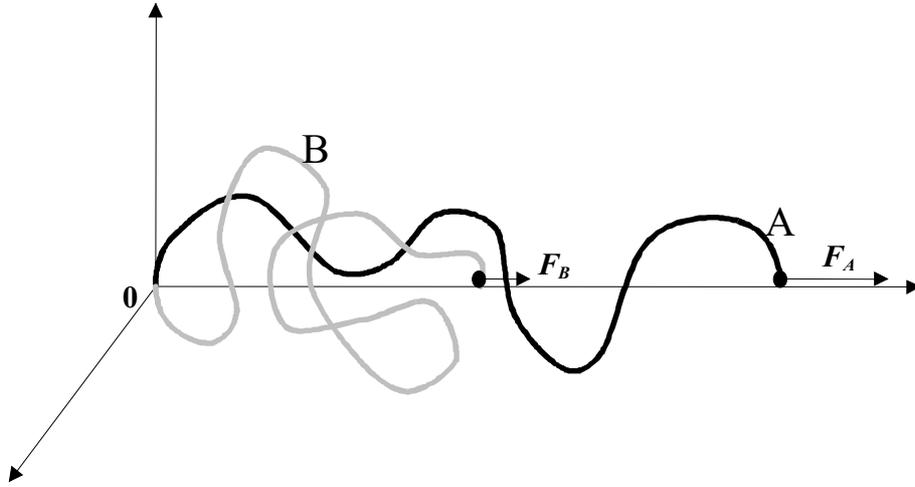


Abbildung 3.1: Schematische Darstellung zur entropischen Federkraft. Eine Kette, die im Zustand A festgehalten wird, kann weniger Konformationen annehmen als die Kette, die in B festgehalten wird, weshalb die „rückstellende“ entropische Kraft in A größer ist (aus Kreth [Kre00]).

Anzahl der möglichen Konfigurationen, die das Polymer einnehmen kann, reduziert. Die Spannung bewirkt eine rückstellende Federkraft, welche die Kette wieder in einen Zustand mit einer größeren Anzahl an Konfigurationen, das heißt größerer Entropie, zurückbringt, vergleiche Abbildung 3.1.

### 3.1.6 Elastizitätstheorie

Die in den obigen Abschnitten hergeleiteten Größen können zu den elastischen Eigenschaften eines Polymers in Beziehung gesetzt werden. Dazu soll zunächst eine homogene elastische Stange betrachtet werden.  $s$  sei eine kontinuierliche krummlinige Koordinate, deren Achse dem Verlauf der Stange folgt; sie nimmt Werte zwischen 0 und  $L_{max}$ , der Konturlänge, an.  $\mathbf{u}(s)$  bezeichne den Einheitstangentenvektor zur Stange an der Stelle  $s$ . Werden Verformungen und lokale Krümmungen der Stange durch thermische Fluktuationen hervorgerufen, so fällt die Winkelkorrelation zwischen zwei Einheitstangentenvektoren an den Stellen  $s$  und  $s'$  exponentiell ab:

$$\langle \cos(\Theta) \rangle = \langle \mathbf{u}(s) \cdot \mathbf{u}(s') \rangle = e^{-k|s-s'|} \quad (3.17)$$

Die Persistenzlänge  $L_P$  ist definiert als der Abstand, bei dem die Winkelkorrelation auf  $\frac{1}{e}$  abgefallen ist:  $L_P = \frac{1}{k}$ . Sie hängt mit dem elastischen Biegemodul der Stange wie folgt zusammen (siehe [LL86]):

$$L_P = \frac{B}{k_B T} \quad (3.18)$$

Mit Gleichung 3.17 läßt sich der mittlere quadratische End-zu-End-Abstand der Stange berechnen:

$$\begin{aligned} \langle \mathbf{R}^2 \rangle &= \int_0^{L_{max}} ds \int_0^{L_{max}} ds' \langle \mathbf{u}(s) \cdot \mathbf{u}(s') \rangle = \int_0^{L_{max}} ds \int_0^{L_{max}} ds' e^{-\frac{|s-s'|}{L_P}} \\ &= 2L_P L_{max} \left( 1 - \frac{L_P}{L_{max}} \left( 1 - e^{-\frac{L_{max}}{L_P}} \right) \right) \end{aligned} \quad (3.19)$$

Im Grenzfall  $L_{max} \gg L_P$  geht Gleichung 3.19 über in:

$$\langle \mathbf{R}^2 \rangle = 2L_{max} L_P \quad (3.20)$$

Modellvorstellungen über die Elastizität eines Polymers bewegen sich nun zwischen folgenden Grenzfällen:

- Im *freely jointed chain*<sup>1</sup> Modell gilt: die Länge eines Segmentes ist größer als die Persistenzlänge:  $b \gg L_P$ . In einem solchen *coarse-grained*<sup>2</sup> Modell spielen die elastischen Eigenschaften keine Rolle.
- Im *wormlike chain*<sup>3</sup> Modell ist  $b \sim L_P$ . Hier ist eine Berücksichtigung der elastischen Eigenschaften erforderlich.

Aus der Definition für die Kuhnsche Segmentlänge eines Polymers,  $b_K = \frac{\langle \mathbf{R}^2 \rangle}{L_{max}}$ , folgt:  $\langle \mathbf{R}^2 \rangle = L_{max} b_K$ . In der obigen Näherung ist also  $b_K = 2L_P$ .

## 3.2 Monte Carlo Methoden

Konfigurationen von Systemen mit vielen Freiheitsgraden werden durch einen Zustandsvektor  $\mathbf{x}$  angegeben. Dieser ist ein Element des Phasenraumes des betrachteten Systems. Er kann z.B. die Koordinaten und Geschwindigkeiten aller Teilchen eines Gases in einem festgelegten Volumen enthalten, oder die Koordinaten aller Segmente eines Polymers. In der statistischen Physik ist man interessiert am Erwartungswert  $\langle A(\mathbf{x}) \rangle$  einer Observablen  $A$ , welcher Aufschluß über die physikalischen Eigenschaften des Systems gibt und welcher unter Umständen mit experimentellen Daten vergleichbar ist. Klassische Beispiele sind die Magnetisierung eines Kristalls oder die End-zu-end-Länge eines Polymers.

Untersucht man ein System, welches durch ein kanonisches Ensemble beschrieben wird, so ist der Erwartungswert der Observablen  $A$  gegeben durch:

$$\langle A(\mathbf{x}) \rangle = \frac{1}{Z} \int e^{-\frac{H(\mathbf{x})}{k_B T}} A(\mathbf{x}) d\mathbf{x} = \int p(\mathbf{x}) A(\mathbf{x}) d\mathbf{x} \quad (3.21)$$

Dabei ist  $Z = \int e^{-\frac{H(\mathbf{x})}{k_B T}} d\mathbf{x}$  die Zustandssumme und der Boltzmann-Faktor

$$p(\mathbf{x}) = \frac{1}{Z} e^{-\frac{H(\mathbf{x})}{k_B T}} \quad (3.22)$$

gibt an, wie stark die Konfiguration  $\mathbf{x}$  im Mittelwert der Gleichung 3.21 gewichtet wird. Das hochdimensionale Integral in 3.21 kann jedoch in den meisten Fällen nicht analytisch ausgeführt werden, außerdem ist eine genaue Kenntnis der mikroskopischen Konfigurationen weder möglich, noch erwünscht.

Die Idee von Monte Carlo Verfahren ist, das Integral in 3.21 durch eine Summe über eine endliche Teilmenge des Phasenraumes  $\{\mathbf{x}_1, \mathbf{x}_2 \dots \mathbf{x}_M\}$  zu ersetzen:

$$\langle A(\mathbf{x}) \rangle \approx \frac{\sum_{i=0}^M \exp\left(\frac{H(\mathbf{x}_i)}{k_B T}\right) A(\mathbf{x}_i)}{\sum_{i=0}^M \exp\left(\frac{H(\mathbf{x}_i)}{k_B T}\right)} \quad (3.23)$$

Man unterscheidet nun verschiedene Monte Carlo Verfahren nach der Art und Weise, wie die Zustände  $\mathbf{x}_1, \mathbf{x}_2 \dots \mathbf{x}_M$  ausgewählt werden. Für detailliertere Informationen siehe [BH97].

### 3.2.1 Simple Sampling

Beim *simple sampling* werden die Zustände  $\mathbf{x}_1, \mathbf{x}_2 \dots \mathbf{x}_M$  zufällig, von einander unabhängig und mit gleicher Wahrscheinlichkeit ausgewählt. Hierzu bedient man sich eines Pseudozufallszahlengenerators. Die Qualität der Pseudo-Zufallszahlen ist entscheidend für die Güte der berechneten Erwartungswerte. Es ist:

<sup>1</sup>englisch für "frei verbundene Kette"

<sup>2</sup>englisch für "grobkörnig"

<sup>3</sup>englisch für "wurmartige Kette"

$$\langle A(\mathbf{x}) \rangle \approx \sum_{i=0}^M \tilde{p}(\mathbf{x}_i) A(\mathbf{x}_i) \quad (3.24)$$

$$\text{wobei } \tilde{p}(\mathbf{x}_i) = \frac{1}{\sum_{i=0}^M \exp\left(\frac{H(\mathbf{x}_i)}{k_B T}\right)} \cdot \exp\left(\frac{H(\mathbf{x}_i)}{k_B T}\right)$$

Anwendungen des *simple sampling* sind z.B. *random walk*- oder *self avoiding walk*-Probleme aus der Polymerphysik.

Die Vorteile von *simple sampling* Algorithmen sind einerseits die einfache Implementierung und andererseits die Tatsache, daß die ausgewählten Konfigurationen statistisch unabhängig sind und daher standardisierte Methoden zur Fehlerabschätzung verwendet werden können. Ein großer Nachteil ist jedoch, daß alle Konfigurationen mit der gleichen Wahrscheinlichkeit ausgewählt werden. Daher kommen auch viele Konfigurationen  $\mathbf{x}_l$  vor, die aufgrund ihres Gewichts  $p(\mathbf{x}_l)$  nicht oder nur wenig zu  $\langle A(\mathbf{x}) \rangle$  beitragen. Dieser Nachteil wird durch das im folgenden beschriebene *importance sampling* überwunden.

### 3.2.2 Importance Sampling

Wählt man die Zustände  $\{\mathbf{x}_l, l = 1 \dots M\}$  nicht mit gleicher Wahrscheinlichkeit, sondern gemäß einer Wahrscheinlichkeitsverteilung  $P(\mathbf{x}_l)$  aus dem Phasenraum aus, so muß Gleichung 3.23 modifiziert werden zu:

$$\langle A(\mathbf{x}) \rangle \approx \frac{\sum_{i=0}^M \frac{\exp\left(\frac{H(\mathbf{x}_i)}{k_B T}\right)}{P(\mathbf{x}_i)} A(\mathbf{x}_i)}{\sum_{i=0}^M \frac{\exp\left(\frac{H(\mathbf{x}_i)}{k_B T}\right)}{P(\mathbf{x}_i)}} = \sum_{i=0}^M \frac{\tilde{p}(\mathbf{x}_i)}{P(\mathbf{x}_i)} A(\mathbf{x}_i) \quad (3.25)$$

$$\text{wobei } \tilde{p}(\mathbf{x}_i) = \frac{1}{\sum_{i=0}^M \frac{\exp\left(\frac{H(\mathbf{x}_i)}{k_B T}\right)}{P(\mathbf{x}_i)}} \cdot \exp\left(\frac{H(\mathbf{x}_i)}{k_B T}\right)$$

Die einfachste Wahl für  $P(\mathbf{x}_l)$  ist  $P(\mathbf{x}_l) \propto e^{-\frac{H(\mathbf{x}_l)}{k_B T}}$ . Dann vereinfacht sich Gl. 3.25 zu

$$\langle A(\mathbf{x}) \rangle \approx \frac{1}{M} \sum_{i=0}^M A(\mathbf{x}_i) \quad (3.26)$$

Das Problem allerdings ist, eine Vorschrift zu finden welche die Zustände  $\mathbf{x}_1, \mathbf{x}_2 \dots \mathbf{x}_M$  gemäß dieser Wahrscheinlichkeitsverteilung auswählt. Metropolis et al. schlugen 1953 vor, die Zustände nicht unabhängig von einander zu wählen, sondern sie durch einen Markov Prozess zu generieren, bei dem ein Zustand  $\mathbf{x}_{l+1}$  aus dem vorhergehenden und bereits existierenden Zustand  $\mathbf{x}_l$  mit Hilfe einer Übergangswahrscheinlichkeit  $W(\mathbf{x}_l \rightarrow \mathbf{x}_{l+1})$  gebildet wird. Sie zeigten, daß es Verteilungen  $P(\mathbf{x}_l)$  gibt, welche im Grenzwert  $M \rightarrow \infty$  gegen die Gleichgewichtsverteilung  $P_{Gl}(\mathbf{x}) = \frac{1}{Z} e^{-\frac{H(\mathbf{x})}{k_B T}}$  konvergieren. Dafür ist es hinreichend, an die Übergangswahrscheinlichkeit die Bedingung der detaillierten Balance (mikroskopische Reversibilität) zu stellen:

$$P_{Gl}(\mathbf{x}_i) W(\mathbf{x}_i \rightarrow \mathbf{x}_j) = P_{Gl}(\mathbf{x}_j) W(\mathbf{x}_j \rightarrow \mathbf{x}_i) \quad (3.27)$$

Dann hängt das Verhältnis der Wahrscheinlichkeiten für die Übergänge  $\mathbf{x}_i \rightarrow \mathbf{x}_j$  und  $\mathbf{x}_j \rightarrow \mathbf{x}_i$  nur von der Energiedifferenz  $\Delta H = H(\mathbf{x}_j) - H(\mathbf{x}_i)$  ab:

$$\frac{W(\mathbf{x}_i \rightarrow \mathbf{x}_j)}{W(\mathbf{x}_j \rightarrow \mathbf{x}_i)} = e^{-\frac{\Delta H}{k_B T}} \quad (3.28)$$

Dies legt  $W(\mathbf{x}_i \rightarrow \mathbf{x}_j)$  nicht eindeutig fest, eine mögliche Wahl ist:

$$W(\mathbf{x}_i \rightarrow \mathbf{x}_j) = \begin{cases} \frac{1}{\tau_s} e^{-\frac{\Delta H}{k_B T}} & \text{falls } \Delta H > 0 \\ \frac{1}{\tau_s} & \text{falls } \Delta H \leq 0 \end{cases} \quad (3.29)$$

Hierbei kann  $\tau_s$  als Einheit der Monte-Carlo-Zeit interpretiert werden, sofern eine dynamische Interpretation des Monte Carlo Verfahrens möglich ist. Im folgenden soll nun gezeigt werden, daß die durch die Übergangswahrscheinlichkeit 3.29 generierten Konfigurationen  $\mathbf{x}_1, \mathbf{x}_2, \dots$ , auch Markov-Kette genannt, tatsächlich die Bedingung erfüllen, daß ihre Wahrscheinlichkeitsverteilung  $P(\mathbf{x}_l)$  gegen die Gleichgewichtsverteilung  $P_{Gl}(\mathbf{x}_l)$  konvergiert.

Man betrachte mehrere Markov-Ketten gleichzeitig und nehme an, daß an einem vorgegebenen Schritt im Prozess  $N_r$  Systeme sich im Zustand  $r$  befinden und  $N_s$  Systeme im Zustand  $s$ . Man nehme ferner an, daß  $H(\mathbf{x}_r) < H(\mathbf{x}_s)$ . Die Anzahl an Übergängen von  $\mathbf{x}_r$  nach  $\mathbf{x}_s$ ,  $N_{r \rightarrow s}$ , ist dann:

$$N_{r \rightarrow s} = N_r W(\mathbf{x}_r \rightarrow \mathbf{x}_s) = N_r \frac{1}{\tau_s} \exp\left(-\frac{H(\mathbf{x}_s) - H(\mathbf{x}_r)}{k_B T}\right) \quad (3.30)$$

und die Anzahl der Übergänge von  $\mathbf{x}_s$  nach  $\mathbf{x}_r$  ist:

$$N_{s \rightarrow r} = N_s W(\mathbf{x}_s \rightarrow \mathbf{x}_r) = N_s \frac{1}{\tau_s} \quad (3.31)$$

Die effektive Anzahl an Übergängen ist:

$$\begin{aligned} \Delta N_{r \rightarrow s} &= N_{r \rightarrow s} - N_{s \rightarrow r} \\ &= N_r \frac{1}{\tau_s} \left( \frac{\exp\left(-\frac{H(\mathbf{x}_s)}{k_B T}\right)}{\exp\left(-\frac{H(\mathbf{x}_r)}{k_B T}\right)} - \frac{N_s}{N_r} \right) \end{aligned} \quad (3.32)$$

Nun gilt: ist das Verhältnis  $\frac{N_s}{N_r}$  kleiner als es der Gleichgewichtsverteilung entspricht, so ist  $\Delta N_{r \rightarrow s} > 0$  und  $\frac{N_s}{N_r}$  wächst solange an, bis die Gleichgewichtsverteilung erreicht ist, ist andererseits  $\frac{N_s}{N_r}$  größer als es der Gleichgewichtsverteilung entspricht, so ist  $\Delta N_{r \rightarrow s} < 0$  und  $\frac{N_s}{N_r}$  fällt solange ab, bis die Gleichgewichtsverteilung erreicht ist.

Der folgende Algorithmus (Metropolisalgorithmus) stellt eine Realisierung des *importance sampling* dar:

1. Beginne mit einem Startzustand bzw. einer Startkonfiguration  $\mathbf{x}_0$  im Phasenraum.
2. Wähle im  $l$ -ten Schritt zufällig irgendeinen anderen Zustand  $\mathbf{x}_l$  im Phasenraum aus.
3. Berechne  $\Delta H = H(\mathbf{x}_l) - H(\mathbf{x}_{l-1})$ .
  - Falls  $\Delta H < 0$ , so akzeptiere den Zustand  $\mathbf{x}_l$ , inkrementiere  $l$  und fahre mit Schritt 2 fort.
  - Falls  $\Delta H \geq 0$ , so akzeptiere  $\mathbf{x}_l$  mit der Wahrscheinlichkeit  $\exp\left(-\frac{\Delta H}{k_B T}\right)$ . Dies wird wie folgt implementiert: Generiere eine Zufallszahl  $\epsilon$  mit  $0 \leq \epsilon < 1$ ; falls  $\epsilon < \exp\left(-\frac{\Delta H}{k_B T}\right)$ , so akzeptiere  $\mathbf{x}_l$ , inkrementiere  $l$  und fahre mit Schritt 2 fort; ansonsten verwerfe  $\mathbf{x}_l$  und beginne von neuem mit Schritt 2.

# Kapitel 4

## Das SCD-Modell

### 4.1 Modellierung und Simulation

In dem von Gregor Kreth entwickelten *Spherical 1-Mpb Chromatin Domain* (SCD) Modell (vergleiche [Kre00]) werden die chromosomalen Foci durch kugelförmige, 1 Mbp große Domänen angenähert, welche durch ein abstoßendes Potential mit einander wechselwirken. Sie sind verbunden durch einen Chromatin-Linker. Dessen Detailstruktur wird nicht explizit ausgerechnet, statt dessen wird mit einem entropischen Federpotential seine Wirkung auf zwei benachbarte Domänen simuliert. Es handelt sich also um ein *beads on a spring*<sup>1</sup> - Modell. Das Modell erlaubt keine Aussagen über die Substruktur der Foci, ermöglicht aber z.B. die Simulation eines Zellkerns mit dem gesamten menschlichen Genom in annehmbarer Rechenzeit auf einem PC. Es sollen nun die Modellannahmen genauer beschrieben werden.

Jedes Chromosom wird durch eine lineare Kette kugelförmiger Domänen beschrieben, dies entspricht der Tatsache, daß pro Schwesterchromatid nur ein durchgehender DNA-Faden existiert. Für jede Domäne wird eine Größe von 1 Mbp und ein Durchmesser von 500 nm angenommen. Dies entspricht der mittleren gemessenen Größe von Replikationfoci (vergleiche [Kre00]). Die Anzahl der 1-Mbp Domänen eines Chromosoms wird proportional zu seinem DNA-Gehalt gewählt: das Chromosom #1 mit einem DNA-Gehalt von 245 Mbp wird durch eine Kette aus 245 Domänen modelliert.

#### 4.1.1 Der Chromatin-Linker

In Anlehnung an experimentelle Daten (vergleiche [Kre00]) wird eine Persistenzlänge  $L_P$  (siehe 3.1.6) von 150 nm zugrunde gelegt, was einer Kuhn'schen Segmentlänge  $b_K$  (siehe 3.1.4 und 3.1.6) von 300 nm entspricht falls die Konturlänge  $L_{max}$  viel größer als die Persistenzlänge ist. Dieser Fall ist beim MLS-Modell erfüllt, denn es gilt: bei dem Basengehalt der DNA, bei dem einer Kuhn'schen Segmentlänge 30 kbp entsprechen, hat ein Linker, welcher 120 kbp groß ist, eine Konturlänge von 1200 nm. Daher gilt:

$$L_{max} = 1200nm \gg 150nm = L_P \quad (4.1)$$

Der Linker kann folglich als ideale Kette angenommen werden. Seine Wirkung auf die Domänen, die er verbindet, ist die einer entropischen Federkraft, es gilt Gleichung 3.16 unter Berücksichtigung von Gleichung 3.20:

$$\mathbf{F} = -\nabla A = -\frac{3k_B T}{N b^2} \mathbf{R} = -\frac{3k_B T}{2L_P L_{max}} \mathbf{R} = -\frac{3k_B T}{l_0^2} \mathbf{R} = -k \mathbf{R} \quad (4.2)$$

---

<sup>1</sup>Perlen auf einer Feder

mit  $l_0 = \sqrt{2L_P L_{max}}$  und  $k = \frac{3k_B T}{Nb^2}$ . Das zu dieser Kraft gehörende Potential ist:

$$U = \frac{1}{2} k \mathbf{R}^2 \quad (4.3)$$

Es stellt sich im thermodynamischen Gleichgewicht jedoch kein Gleichgewichtsabstand von 0 ein - was bedeutet, daß die Kette, welche das DNA-Polymer modelliert, nicht kollabiert, sondern entsprechend der Boltzmannverteilung ist

$$\langle x^2 \rangle = \frac{\int_0^\infty dx x^2 \exp\left(-\frac{U(x)}{k_B T}\right)}{\int_0^\infty dx \exp\left(-\frac{U(x)}{k_B T}\right)} = \frac{\int_0^\infty dx x^2 \exp\left(-\frac{kx^2}{2k_B T}\right)}{\int_0^\infty dx \exp\left(-\frac{kx^2}{2k_B T}\right)} = \frac{k_B T}{k} = \frac{Nb^2}{3} \quad (4.4)$$

Dasselbe gilt für die anderen Koordinatenrichtungen und es ergibt sich:

$$\langle x^2 \rangle + \langle y^2 \rangle + \langle z^2 \rangle = Nb^2 = l_0^2 \quad (4.5)$$

$l_0$  ist also die mittlere Länge des Chromatin-Linkers. Eine Domäne  $i$  im Polymer, mit Ausnahme derjenigen an den beiden Enden, hat zwei Nachbarn, daher wirkt auf sie die Kraft

$$\mathbf{F}_i = -\frac{3k_B T}{l_0^2} (l_{i-1} \mathbf{u}_{i-1} - l_i \mathbf{u}_i) \quad (4.6)$$

wobei  $\mathbf{u}_i$  ein Einheitsvektor ist, welcher von Domäne  $i$  zu Domäne  $i + 1$  zeigt, und  $l_i$  der euklidische Abstand zwischen den Domänen  $i$  und  $i + 1$  ist. Das Gesamtpotential, welches auf durch den Chromatinlinker auf die Domäne  $i$  wirkt, ist daher:

$$U_i = \frac{3k_B T}{2l_0^2} (l_{i-1}^2 + l_i^2) \quad (4.7)$$

Für die tatsächliche Berechnung im Modell werden nur die Potentiale gebraucht, da Energien verglichen werden.

### 4.1.2 Excluded Volume Wechselwirkung

Da die Chromatinfaser eine kompakte Struktur besitzt, kann sie sich nicht beliebig durchdringen. Das muß auch bei den 1-Mbp Domänen berücksichtigt werden. Deshalb wird in das Modell eine Wechselwirkung eingeführt, die verhindert, daß sich zwei Kugeln überlappen (*excluded volume*). Die Chromatindichte in einer Domäne steigt von außen nach innen an, siehe dazu die Arbeit zum *Multi Loop Subcompartment* Modell (MLS) für Inter- und Metaphasechromosomen von Münkel und Langowski [ML98]. Daraus ergibt sich ein kurzreichweitiges, ansteigendes Potential für die 1-Mbp Domänen. Dieses Potential entspricht auf einer größeren Längenskala dem Potential der Wechselwirkung der Beads beim MLS Modell und ist rein empirisch.

$$U_E(r) = U_0 k_B T \left( 1 + \frac{r^4 - 2D^2 r^2}{D^4} \right) \quad (4.8)$$

$$\mathbf{F}_E(\mathbf{r}) = -U_0 \frac{4k_B T}{D^4} (r^2 - D^2) \mathbf{r} \quad D = 500 \text{ nm} \quad (4.9)$$

Bei einem Abstand  $D$  größer oder gleich 500 nm gibt es keine Wechselwirkung zwischen den Domänen; wenn der Abstand kleiner wird, steigt die Stärke der Wechselwirkung langsam an.  $U_0$  bestimmt die Höhe der Potentialbarriere und damit die Wahrscheinlichkeit des Durchdringens zweier Domänen. Untersuchungen mit verschiedenen Werten von  $U_0$  hat Gregor Kreth in [Kre00] durchgeführt.

### 4.1.3 Tensegrity Force Wechselwirkung

Experimentellen Beobachtungen (siehe Kapitel 2.3.2) zufolge bilden Chromosomen in der Interphase eine Territorienstruktur aus. Mit den bisher eingeführten Wechselwirkungen alleine würde sich bei der Beschreibung eines Chromosoms durch die *beads on a spring* noch keine entsprechende Verpackung ergeben. Um den experimentellen Beobachtungen gerecht zu werden, wird bei der Simulation der Chromosomenketten nach dem SCD Modell eine schwache kugelförmige Potentialbarriere um jede Chromosomenkette gelegt (analog zu [HHB<sup>+</sup>93]). Das Volumen der Potentialbarriere wird in Abhängigkeit des DNS Gehaltes des jeweiligen Chromosoms am Gesamtvolumen des Zellkernes festgelegt [HHL<sup>+</sup>89] (Abbildung 4.1):

$$U_T(r) = \begin{cases} 0 & \text{für } r < R_{\text{Terr}} - a \\ 0.1 \cdot U_0 \frac{k_B T}{a} (r - R_{\text{Terr}} - a) & \text{für } r \geq R_{\text{Terr}} - a \end{cases} \quad (4.10)$$

$$R_{\text{Terr}} = R_{\text{Kern}} \left( \frac{c_{\text{Chromosom}}}{c_{\text{Genom}}} \right)^{\frac{1}{3}} \quad r = |\mathbf{r} - \mathbf{r}_c| \quad \mathbf{r}_c = \frac{1}{N} \sum_{i=1}^N \mathbf{r}_i \quad (4.11)$$

Hierbei ist  $a$  der Beadradius,  $R_{\text{Terr}}$  der Radius der Potentialbarriere und  $R_{\text{Kern}}$  der Radius der *Zellkernkugel*.  $c_{\text{Chromosom}}$  und  $c_{\text{Genom}}$  bezeichnen jeweils den DNA-Gehalt des Chromosoms beziehungsweise des gesamten Genoms.

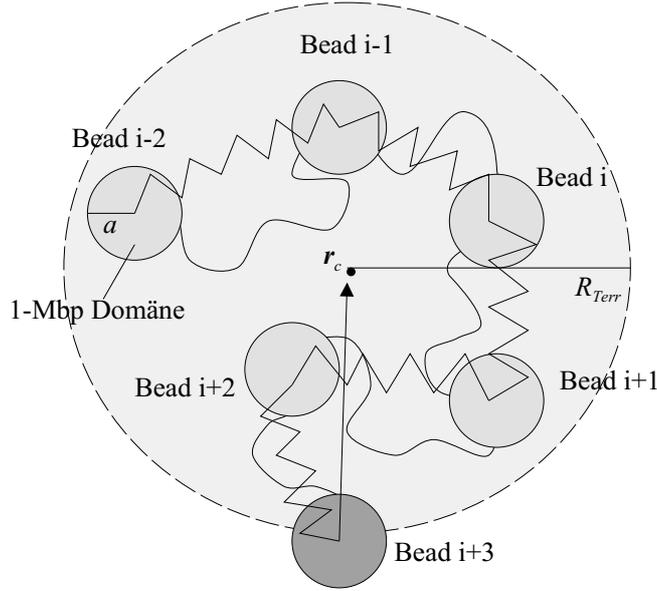


Abbildung 4.1: Schematische Darstellung der Tensegrity Force Wechselwirkung: um jede Chromosomenkette wird eine schwache, kugelförmige Potentialbarriere (mit Radius  $R_{\text{Terr}}$ ) gelegt. Eine 1-Mbp Domäne, die sich außerhalb dieser Barriere bewegt (Bead  $i+3$ ), erfährt eine konstante Kraft, die zum Schwerpunkt  $r_c$  der Kette gerichtet ist (aus Kreth [Kre00]).

Alle Beads, die weiter als  $R_{\text{Terr}} - a$  vom Schwerpunkt des Chromosomenterritoriums entfernt sind, erfahren eine konstante Kraft mit dem Betrag  $0.1 \cdot U_0 \frac{k_B T}{a}$  in Richtung des Schwerpunktes  $r_c$  des Territoriums (zum Beispiel Bead  $i+3$  in Abbildung 4.1).

Das MLS Modell verwendet eine solche Barriere nur für die Simulation einzelner Chromosomen, um die Anwesenheit benachbarter Chromosomen zu simulieren [ML98]. Für die Aufrechterhaltung der Territorienstruktur ist sie allerdings nicht notwendig und deshalb wird sie bei der Simulation ganzer Zellkerne auch nicht angewendet. Beim MLS Modell kommt es vermutlich durch Verhakungen der Chromatin-Linker während der Relaxation zu einer Kompaktierung, so daß sich ein Abstandsverhalten von  $\sim N^{\frac{1}{3}}$  ergibt.

#### 4.1.4 Monte-Carlo Relaxation

Mit Hilfe des unter 3.2.2 vorgestellten Metropolis-Algorithmus kann man thermodynamische Gleichgewichtskonfigurationen des *Spherical 1-MBp Chromatin Domain* (SCD)-Modells berechnen. Man beginnt hierbei mit einer Startkonfiguration und relaxiert die darin angeordneten Chromosomenketten. Das Erreichen des Gleichgewichtszustands kann durch geometrische Moden wie z.B. den Gyrationradius (vergleiche 3.1) kontrolliert werden.

Die Generierung von **Mitose**- oder Pseudo-Mitose-Startkonfigurationen erfolgt mit dem *Simple Sampling Monte Carlo* Verfahren (vergleiche 3.2.1). Im ersten Schritt werden in zufälliger Reihenfolge Territorien ausgewählt. Jedem Territorium wird eine *Territorienkugel* zugeordnet, deren Größe dem DNA-Gehalt des Territoriums angepasst wird:

$$R_{Terr} = \left( \frac{3}{4\pi} V_{Kern} \cdot \nu_s \cdot \frac{c_{Chr}}{c_{Ge}} \right)^{\frac{1}{3}} \quad (4.12)$$

Hierbei ist  $R_{Terr}$  der Radius der Territorienkugel,  $V_{Kern}$  das Volumen des gesamten Zellkerns,  $\nu_s$  ein gemeinsamer Faktor, um den die Volumina sämtlicher Territorienkugeln verkleinert werden, um die hohe Kompaktierung der Territorien gegen Ende der Mitose darzustellen ( $\nu_s = 0.22$  ist eine gute Wahl),  $c_{Chr}$  ist der DNA-Gehalt des dem Territorium entsprechenden Chromosoms und  $c_{Ge}$  ist der DNA-Gehalt des gesamten Genoms. Im Falle eines kugelförmigen Zellkerns ergibt sich:

$$R_{Terr} = R_{Kern} \cdot \left( \nu_s \cdot \frac{c_{Chr}}{c_{Ge}} \right)^{\frac{1}{3}} \quad (4.13)$$

Für ein Territorium werden nun solange verschiedene, zufällig ausgewählte Plätze getestet, bis keine Überlappung mit anderen Territorienkugeln oder der Zellkernhülle besteht.

Die 1-Mbp-Domänen einer Chromosomenkette werden nun hintereinander zu Zylindern angeordnet, ein solcher Zylinder entspricht einer gestreckten Chromatide in der Ana- und Telophase (vergleiche 2.1.4). Der Mittelpunkt eines Zylinders wird in den Mittelpunkt der zugehörigen Territorienkugel gelegt. In dieser zylindrischen Anordnung ist der Abstand zwischen zwei aufeinander folgenden 1-Mbp-Domänen so klein, daß sie sich stark durchdringen.

Als Startkonfiguration kann auch eine bereits vorhandene Kernkonfiguration dienen, die aus einem früheren Simulationslauf hervorgegangen ist.

Die Generierung weiterer Startkonfigurationen und deren biologische Interpretation wird in Abschnitt 4.2.1 und 4.2.2 behandelt.

Im Monte-Carlo Schritt wird nun in jedem Territorium eine Domäne zufällig ausgewählt und innerhalb einer Kugel mit Radius  $\delta r_{max}$  verrückt; dann wird die Energie der neuen Konfiguration ausgerechnet und mit der Energie der alten verglichen und die Bedingungen des Metropolis Algorithmus überprüft.  $\delta r_{max}$  wird dabei vom Monte Carlo Programm so eingestellt, daß mindestens die Hälfte aller Schritte akzeptiert wird. Eine Abfrage zur Justierung von  $\delta r_{max}$  erfolgt alle 500 Monte Carlo Schritte.

#### 4.1.5 Visualisierung

Sowohl Startkonfigurationen, als auch relaxierte Kerne können mit Hilfe der ray-tracing Software POV-Ray ("persistence of vision ray-tracing") visualisiert werden. Dabei werden die Domänen durch Kugeln dargestellt, der Chromatin-Linker zwischen zwei Domänen durch einen Zylinder zwischen den Kugeln. Verschiedene Chromosomenterritorien können verschieden eingefärbt werden. Technisch gesehen wird zunächst aus der Zellkernkonfiguration eine POV-Ray-Quellcodedatei generiert, aus der dann die POV-Ray-Software ein Bild erzeugt.

## 4.2 Erweiterungen

In diesem Abschnitt werden Erweiterungen des SCD-Modells vorgestellt, welche im Rahmen dieser Arbeit entwickelt wurden.

### 4.2.1 Generierung von Startkonfigurationen für Fibroblasten

Fibroblasten haben ellipsoidförmige Zellkerne. Außerdem haben sie im Zellkern eine Chromosomenverteilung, die sich nach der Größe der Chromosomen richtet. Um Startkonfigurationen für Fibroblasten zu erhalten, wurden diese Eigenschaften in das Programm zur Generierung von Startkonfigurationen integriert. Dieses Programm arbeitet ebenfalls nach dem Monte-Carlo Verfahren und geht gemäß folgenden Schritten vor:

1. Für ein Territorium  $i$  wird mit dem Pseudo-Zufallszahlengenerator ein Richtungsvektor  $\vec{u} = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}$  erzeugt, wobei die Zahlen  $u_1$ ,  $u_2$  und  $u_3$  alle in  $(0,1)$  sind.
2. Es wird nun überprüft, ob ein Territorium, welches seinen Mittelpunkt in  $\vec{u}$  hat, vollständig im Zellkern liegt. Das geschieht mit Hilfe der Ellipsengleichung:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1 \quad (4.14)$$

Falls dies nicht erfüllt ist, so beginne von neuem mit Schritt (1).

3. Überprüfe, ob das Territorium, wenn es seinen Mittelpunkt in  $\vec{u}$  hat, sich mit anderen bereits eingebauten Territorien durchdringt. Falls ja, so beginne erneut mit Schritt (1).
4. Falls ein Ortsvektor gefunden ist, welcher die obigen geometrischen Zwangsbedingungen erfüllt, so überprüfe, ob dieser Ortsvektor der Größenverteilung der Chromosomen bzw. Territorien im Kern entspricht: Die Gerade, welche durch den Ursprung (den Mittelpunkt der zu erstellenden Kernkonfiguration) geht und  $\vec{u}$  als Richtungsvektor hat, hat folgende Gleichung:

$$g : \begin{pmatrix} x \\ y \\ z \end{pmatrix} = t \cdot \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} \quad (4.15)$$

Berechne nun den Schnittpunkt zwischen dieser Geraden und dem Ellipsoid, welches die Kernhülle darstellt:

$$\begin{aligned} t_s^2 \cdot \left( \frac{u_1^2}{a^2} + \frac{u_2^2}{b^2} + \frac{u_3^2}{c^2} \right) &= 1 \\ \Rightarrow t_{s1,2} &= \pm \frac{1}{\sqrt{\frac{u_1^2}{a^2} + \frac{u_2^2}{b^2} + \frac{u_3^2}{c^2}}} \end{aligned} \quad (4.16)$$

Einsetzen in die Geradengleichung ergibt:

$$\vec{x}_s = \begin{pmatrix} x_s \\ y_s \\ z_s \end{pmatrix} = t_s \cdot \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}$$

Der quadrierte Betrag dieses Vektors liefert das Quadrat der Länge des Strahls mit dem Richtungsvektor  $\vec{u}$  im Ellipsoid:

$$|\vec{x}_s|^2 = t_s^2 \cdot (u_1^2 + u_2^2 + u_3^2) = \frac{u_1^2 + u_2^2 + u_3^2}{\frac{u_1^2}{a^2} + \frac{u_2^2}{b^2} + \frac{u_3^2}{c^2}}$$

Bilde nun die Verhältnisse

$$frac2 = \frac{|\vec{u}|^2}{|\vec{x}_s|^2} = \frac{u_1^2 + u_2^2 + u_3^2}{\frac{u_1^2 + u_2^2 + u_3^2}{\frac{a^2}{u_1^2} + \frac{b^2}{u_2^2} + \frac{c^2}{u_3^2}}} = \frac{u_1^2}{a^2} + \frac{u_2^2}{b^2} + \frac{u_3^2}{c^2}$$

und

$$frac = \sqrt{frac2} = \frac{|\vec{u}|}{|\vec{x}_s|}$$

und berechne den Wert

$$p = A \cdot e^{frac \cdot \frac{Size[46]}{Size[i]}} \quad (4.17)$$

wobei  $Size[46]$  die Größe von Territorium 46 (das kleinste von allen) und  $Size[i]$  die Größe von Territorium  $i$  ist.  $A$  ist ein Parameter, der so angepasst wird, daß eine sinnvolle Akzeptanzrate entsteht (siehe untenstehende Fallunterscheidung). Vergleiche den Wert  $p$  mit einer Zufallszahl  $z$ :

- Falls  $p > z$  akzeptiere die Position  $\vec{u}$  für Territorium  $i$ , erhöhe  $i$  um 1 und suche eine Position für das nächste Territorium, fahre mit Schritt (1) fort
- Falls  $p \leq z$  verwerfe die Position  $\vec{u}$  für Territorium  $i$ , wiederhole die Prozedur ab Schritt (1), jedoch ohne  $i$  zu inkrementieren.

Bei großen Territorien ist der Quotient  $\frac{Size[46]}{Size[i]}$  klein. Damit ein großes Territorium einen ähnlich großen Wert  $p$  zugewiesen bekommen kann wie ein kleines, muß sein  $frac$ -Wert größer sein, denn dann ist der Exponent im Ausdruck für  $p$  ähnlich groß. Bei gleicher Akzeptanzrate werden größere Territorien also bevorzugt am Rand des Zellkerns positioniert, für die kleineren bleiben eher Positionen im Inneren des Zellkerns übrig. Man erhält eine größenabhängige Radialverteilung der Chromosomen im Zellkern, wie bei Fibroblasten erforderlich.

## 4.2.2 Speckles und SC35-Domänen

Das SCD-Modell wird im Rahmen dieser Arbeit um *SC35-Domänen* oder *speckles* erweitert (cf. Kapitel 2.5). Diese müssen sowohl in die Programme zur Generierung von Startkonfigurationen, als auch in die Relaxationsprogramme integriert werden. In beiden Fällen wurden verschiedene Möglichkeiten implementiert. Um die Flexibilität möglichst groß zu halten, wurden den Kernkonfigurationen zusätzliche Territorien angefügt, in denen die Positionen der SC35-Domänen oder *speckles* abgespeichert werden. Ein Territorium bietet Platz für bis zu 300 Domänen, sodaß z.B. für die Bearbeitung von 700 speckles 3 zusätzliche Territorien erforderlich sind. Die Größe der speckles wurde so gewählt, daß ihr Gesamtvolumen so groß ist, wie das Gesamtvolumen der typischerweise 30 unter dem Mikroskop sichtbaren SC35-Domänen einer normalen Zelle.

$$\begin{aligned} speckleNum \cdot V_{speckle} &= 30 \cdot V_{Aggregat} \\ \Rightarrow speckleNum \cdot \frac{4\pi}{3} r_{speckle}^3 &= 30 \cdot \frac{4\pi}{3} r_{Aggregat}^3 \\ \Rightarrow r_{speckle} &= \left( \frac{30}{speckleNum} \right)^{\frac{1}{3}} \cdot r_{Aggregat} \end{aligned} \quad (4.18)$$

Die Größe  $r_{Aggregat}$  wird als dreimal so groß wie der Radius einer normalen Chromatindomäne gesetzt. Die Programme zur Generierung von Startkonfigurationen ermöglichen nun folgende Alternativen:

1. Die *SC35-Domänen* oder *speckles* werden zufällig über das gesamte Zellkernvolumen einer "normalen" Mitose-Startkonfiguration verteilt (vergleiche Abbildung 4.2).
2. Die *SC35-Domänen* oder *speckles* werden zufällig, aber ausschließlich am Rand des Zellkernvolumens einer "normalen" Mitose-Startkonfiguration verteilt. Dies ist sinnvoll unter der Annahme, daß die metabolischen Faktoren, welche die *SC35-Domänen* oder *speckles* ausmachen, aus dem Zytoplasma stammen (vergleiche Abbildung 4.3).

3. Die *SC35-Domänen* oder *speckles* werden zufällig, aber ausschließlich am Rand des Zellkernvolumens einer relaxierten Zellkernkonfiguration, die bis dahin ohne *SC35-Domänen* oder *speckles* relaxiert wurde, verteilt (vergleiche Abbildung 4.4). Dies wird motiviert durch die Annahme, daß der Import der metabolischen Faktoren, aus denen die *SC35-Domänen* oder *speckles* bestehen, vor allem in der Interphase erfolgt und die Chromosomen daher zum Zeitpunkt der Wechselwirkung mit diesen metabolischen Faktoren nicht mehr den Packungsgrad der Mitose besitzen.

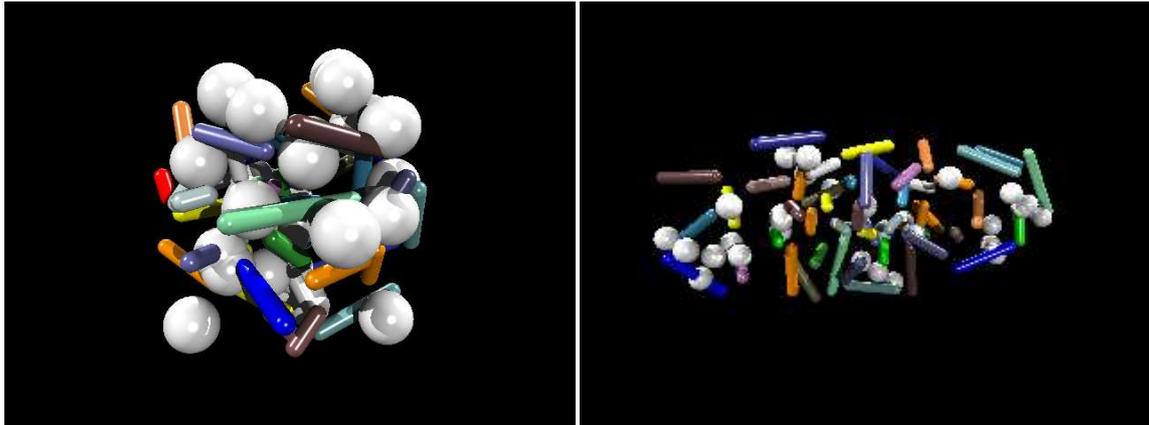


Abbildung 4.2: POVRAY-Visualisierung von Startkonfigurationen mit 30 über den ganzen Kern verteilten SC35-Domänen; links für Lymphozyten mit sphärischem Zellkern, rechts für Fibroblasten mit ellipsoidalem Zellkern. Die relaxierten Konfigurationen dieser Kerne sind in Abb. 4.5 visualisiert.

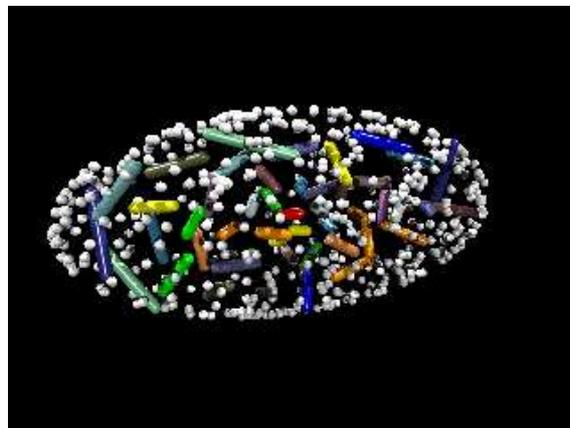


Abbildung 4.3: POVRAY-Visualisierung einer Startkonfiguration mit 500 über die Oberfläche des Zellkerns eines Fibroblasten verteilten speckles. Diese Abbildung hat im Vergleich zu Abb. 4.2 wesentlich mehr speckles, da das Gesamtvolumen der speckles jedoch gleich ist, sind sie hier kleiner.

### 4.2.3 Zuweisung des Bandentyps für Chromatindomänen

Die in Zellkernkonfigurationen des SCD-Modells integrierten *SC35-Domänen* oder *speckles* sollen mit Domänen, die verschiedenen Banden zugeordnet sind, also C-, G- und R-Banden, verschieden wechselwirken. Jede sphärische 1-MBp-Chromatindomäne erhält daher ein Label, in dem spezifiziert wird, von welchem Bandentyp diese Domäne ist. Diese Zuweisung erfolgt durch eine Funkti-

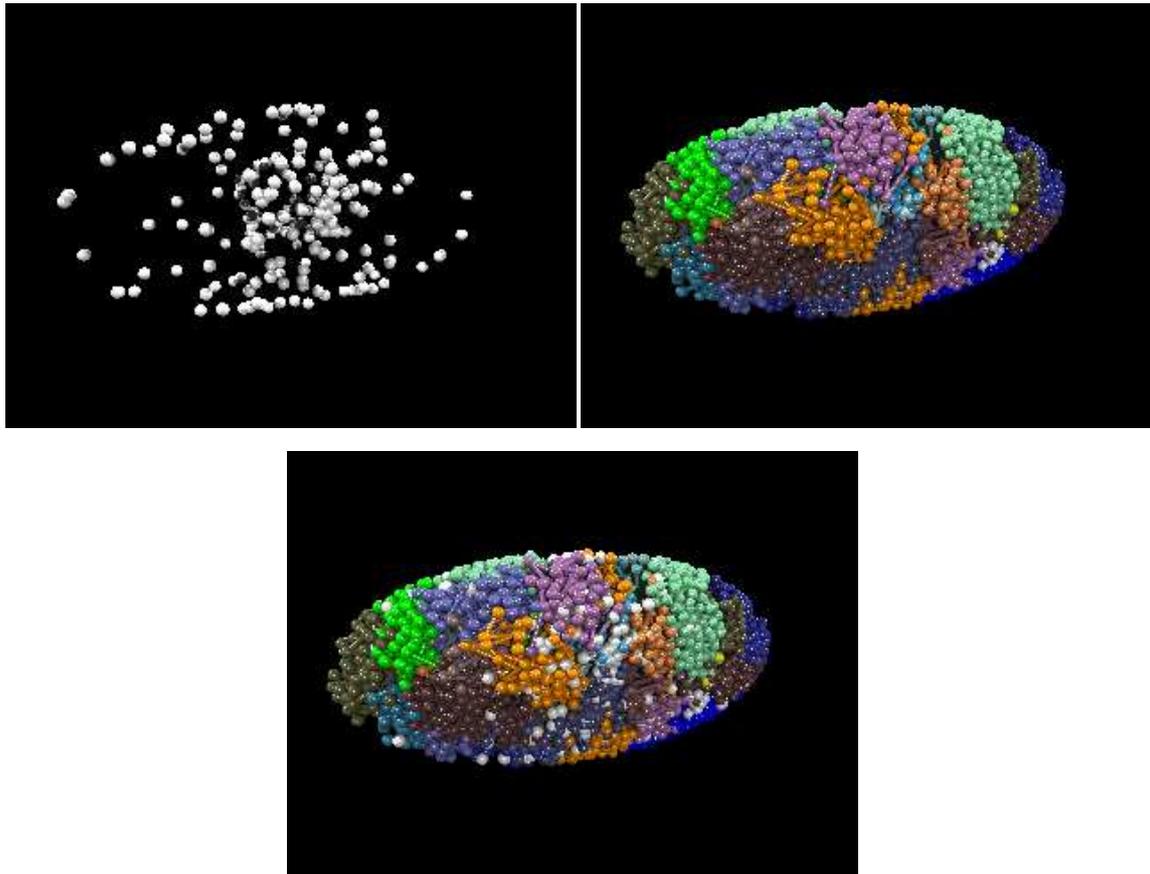


Abbildung 4.4: POVRAY-Visualisierung einer ohne speckles vorrelaxierten Konfiguration (rechts oben), bei der 500 speckles (links oben) nachträglich auf der Oberfläche des Zellkerns verteilt wurden (unteres Bild).

on attributeBands(const Nucleus &nuc), welche in alle Programmteile des SCD-Modells eingefügt wurde. Hierbei wird folgende Konvention verwendet:

- Domänen, die Teil einer C-Bande sind, die also zu Zentromeren gehören, erhalten das Label 1.
- Domänen, die Teil einer G-Bande sind, erhalten das Label 2.
- Domänen, die Teil einer R-Bande sind, erhalten das Label 3.
- SC35-Domänen, speckles oder nuclear bodies erhalten das Label 4.
- Domänen, welche den Nucleoli zugeschrieben werden, erhalten das Label 0.

Die Funktion attributeBands geht dabei wie folgt vor: In die header-Datei global.h wurden Arrays eingefügt, in denen die Anfangsstellen der verschiedenen Banden verzeichnet sind. Dabei sind alle Chromosomenterritorien auf eine Einheitsgröße normiert; um die Information dieser Positionen auf die 1 Mbp-Domänen umzurechnen, wird daher mit der Chromosomengröße multipliziert und anschließend eine Rundungsoperation durchgeführt. Zuerst werden die G-Banden abgearbeitet:

```
for(i=1; i<=46 ; i++)
{
```

```

for(k=1; k<= NUMBERDARKBANDS[t2c[i]];k++)
{
  start=(int)(ChromosomeSize[t2c[i]]*darkbandstart[t2c[i]-1][k]+0.5);
  length=(int)(ChromosomeSize[t2c[i]]*darkbandlength[t2c[i]-1][k]+0.5);
  if(length != 0)
  {
    for(j=start; j<=start+length-1; j++)
    {
      Bands[i][j] = 2;
    }
  }
}
}

```

Hierbei ist  $i$  ein Index, der alle Territorien durchläuft und  $k$  ein Index, der die Anzahl an G-Banden auf diesem Territorium durchläuft. Dieselben Operationen werden für Centromer-Domänen durchgeführt. Jedes Territorium hat nur ein Centromer, daher braucht man in diesem Fall keine Doppelschleife.

```

for(i=1; i<=46 ; i++)
{
  start=(int)(ChromosomeSize[t2c[i]]*centro_band_start[t2c[i]-1][k]+0.5);
  length=(int)(ChromosomeSize[t2c[i]]*centro_band_length[t2c[i]-1][k]+0.5);
  if(length != 0)
  {
    for(j=start; j<=start+length-1; j++)
    {
      Bands[i][j] = 2;
    }
  }
}

```

Alle Domänen, die jetzt noch kein Label haben, die aber aufgrund ihrer Position zu einem Chromosomenterritorium gehören, sind R-Banden. Daher wird ihnen das Label für R-Banden zugewiesen:

```

for(k = 0; k<=ChromosomeSize[t2c[i]]-1; k++)
{
  if(Bands[i][k]==0) Bands[i][k]=3;
}

```

Es müssen nun noch die *SC35-Domänen* oder *speckles* ihr Label erhalten. Hierfür sind jedoch keinerlei Abfragen nötig, da sie sich in zusätzlichen Territorien befinden, in denen es nur Domänen dieses einen Typs gibt:

```

multiple = speckleNum/300;
rest      = speckleNum%300;
for(i=0; i<multiple; i++)
{
  for(k=0; k<300; k++) Bands[49+i][k]=4;
}
for(k=0; k<rest; k++) Bands[49+multiple][k]=4;

```

Die hierbei ausgeführten Modulo-Operationen verteilen die *SC35-Domänen* oder *speckles* auf mehrere Territorien, falls ihre Gesamtanzahl (`speckleNum`) größer ist als die Größe eines Territoriums (300 Domänen).

#### 4.2.4 Wechselwirkungen zwischen den Domänen im Zellkern

In Zellkernkonfigurationen des SCD-Modells, welche um *SC35-Domänen* oder *speckles* (im folgenden werden die Ausdrücke SC35-Domänen und speckles synonym verwendet) erweitert wurden, liegen verschiedene Wechselwirkungen vor. Es können zwei Chromatindomänen miteinander, zwei SC35-Domänen miteinander oder eine Chromatindomäne mit einer SC35-Domäne wechselwirken, wobei unterschieden wird, ob die Chromatindomäne Teil einer C-, G- oder R-Bande ist. Die dabei verwendeten Modellannahmen werden im folgenden Absatz beschrieben.

Für die Wechselwirkung zwischen zwei Chromatindomänen untereinander, also ohne Beteiligung von SC35-Domänen, wird weiterhin das MLS-Potential verwandt, gegenüber bisherigen Relaxationen von Zellkernkonfigurationen mit dem SCD-Modell wurde an dieser Stelle nichts verändert.

$$V(r) = \epsilon \cdot \left(1 + \frac{r^4 - 2 \cdot \sigma^2 \cdot r^2}{\sigma^4}\right) \quad (4.19)$$

Die SC35-Domänen gehen mit den verschiedenen Bandentypen verschiedene Wechselwirkungen ein: mit den spät replizierenden C- und G-Banden wechselwirken sie rein repulsiv, mit den R-Banden wechselwirken sie sowohl attraktiv als auch repulsiv.

Für die Wechselwirkung zwischen SC35-Domänen und R-Banden wird daher ein Lennard-Jones-Potential angesetzt:

$$V(r) = 4 \cdot \epsilon \cdot \left(\frac{\sigma^{12}}{r^{12}} - \frac{\sigma^6}{r^6}\right) \quad (4.20)$$

Ein solches Potential hat einen attraktiven Teil ( $-4 \cdot \epsilon \cdot \frac{\sigma^6}{r^6}$ ) und einen repulsiven Teil ( $+4 \cdot \epsilon \cdot \frac{\sigma^{12}}{r^{12}}$ ), vergleiche mit dem Buch von Allen und Tildesley, [AT89].

Das Minimum eines solchen Potentials liegt bei:

$$\begin{aligned} 0 = V'(r_{min}) &= 4 \cdot \epsilon \cdot \left(-\frac{12 \cdot \sigma^{12}}{r_{min}^{13}} + \frac{6 \cdot \sigma^6}{r_{min}^7}\right) \\ \Rightarrow \frac{6 \cdot \sigma^6}{r_{min}^7} &= \frac{12 \cdot \sigma^{12}}{r_{min}^{13}} \Rightarrow r_{min}^6 = 2 \cdot \sigma^6 \\ &\Rightarrow r_{min} = 2^{\frac{1}{6}} \cdot \sigma = 1,1225 \cdot \sigma \end{aligned} \quad (4.21)$$

$r_{min}$  wird im folgenden Minimumsabstand genannt.

Für die Wechselwirkung zwischen zwei SC35-Domänen untereinander und zwischen SC35-Domänen und G-Banden wurden Potentiale gewählt, die nur dem repulsiven Teil eines Lennard-Jones-Potentials entsprechen:

$$V(r) = \begin{cases} 4 \cdot \epsilon \cdot \left(\frac{\sigma^{12}}{r^{12}} - \frac{\sigma^6}{r^6}\right) + \epsilon & \text{falls } r \leq r_{min} \\ 0 & \text{falls } r > r_{min} \end{cases} \quad (4.22)$$

Der repulsive Anteil des Lennard-Jones-Potentials geht dabei an der Stelle  $r_{min}$  stetig differenzierbar in die Gerade  $y = 0$  über.

Für jede Wechselwirkung zwischen zwei sphärischen Domänen wird der Minimumsabstand gleich der Summe der Radien der beteiligten Sphären gewählt. Wenn nun beispielsweise SC35-Domänen den dreifachen Radius einer Chromatindomäne haben, so ist der Minimumsabstand für die Wechselwirkung zwischen einer SC35- und einer Chromatindomäne  $3 + 1 = 4$  mal der Radius der Chromatindomäne, der Minimumsabstand für die Wechselwirkung zweier SC35-Domänen ist  $3 + 3 = 6$  mal der Radius einer Chromatindomäne.

Ausser den Minimumsabständen geht noch der Parameter  $\epsilon$  in die Potentiale ein. Der Wert eines Lennard-Jones-Potentials in seinem Minimum ist:

$$V(r_{min}) = 4 \cdot \epsilon \cdot \left( \frac{\sigma^{12}}{(2^{\frac{1}{6}}\sigma)^{12}} - \frac{\sigma^6}{(2^{\frac{1}{6}}\sigma)^6} \right) = 4 \cdot \epsilon \cdot \left( \frac{1}{2^2} - \frac{1}{2^1} \right) = 4 \cdot \epsilon \cdot \left( -\frac{1}{4} \right) = -\epsilon \quad (4.23)$$

Der Parameter  $\epsilon$  gibt also die Tiefe der Potentialmulde an. Für jede Wechselwirkung kann er einen anderen Wert annehmen und wird in Kapitel 5.1 auch Gegenstand einer Variation sein. Man erhält also folgende Wechselwirkungen:

- Chromatindomäne mit Chromatindomäne (implementiert als `mlspot(float distance)`):

$$V_0(r) = \epsilon_0 \cdot \left( 1 + \frac{r^4 - 2 \cdot \sigma_0^2 \cdot r^2}{\sigma_0^4} \right) \quad (4.24)$$

- SC35-Domäne mit einer R-Banden-Domäne (implementiert als `SC35_RBand_Pot(double distance)`):

$$V_1(r) = 4 \cdot \epsilon_1 \cdot \left( \frac{\sigma_1^{12}}{r^{12}} - \frac{\sigma_1^6}{r^6} \right) \quad (4.25)$$

- SC35-Domäne mit SC35-Domäne (implementiert als `SC35_SC35_Pot(double distance)`):

$$V_2(r) = \begin{cases} 4 \cdot \epsilon_2 \cdot \left( \frac{\sigma_2^{12}}{r^{12}} - \frac{\sigma_2^6}{r^6} \right) + \epsilon_2 & \text{falls } r \leq r_{min_2} \\ 0 & \text{falls } r > r_{min_2} \end{cases} \quad (4.26)$$

- SC35-Domäne mit einer Centromer- oder G-Banden-Domäne (implementiert als `SC35_GCBand_Pot(double distance)`):

$$V_3(r) = \begin{cases} 4 \cdot \epsilon_3 \cdot \left( \frac{\sigma_3^{12}}{r^{12}} - \frac{\sigma_3^6}{r^6} \right) + \epsilon_3 & \text{falls } r \leq r_{min_3} \\ 0 & \text{falls } r > r_{min_3} \end{cases} \quad (4.27)$$

Jeder Wechselwirkung wird ein cutoff zugeschrieben, dieser wurde gewählt als das 1.5-fache des Minimumsabstands. Bei diesem Abstand ist der Betrag des Potentials auf 16.79 % seines lokalen Maximums gesunken.

Welche Wechselwirkung für einen speziellen Rechenschritt in einer Relaxation auszuwählen ist, ermittelt das Programm in einer geschachtelten `switch`-Abfrage.

#### 4.2.5 Modifikation der Funktion `updateposition`

Die Funktion `updateposition` wird immer dann aufgerufen, wenn in einem Monte Carlo Schritt eine Verrückung akzeptiert worden ist. Sie aktualisiert die Koordinaten der verrückten Domäne, sowohl im zugehörigen Objekt der Klasse `Territory`, als auch auf dem Gitter, welches über das Volumen des simulierten Zellkerns gelegt wird. Die bisherige Funktion `updateposition` löschte zu diesem Zweck alle Einträge im Gitter durch Aufruf der Funktion `resetttables`, aktualisierte die Koordinaten der verrückten Domäne, und machte alle Einträge im Gitter neu durch Aufruf der Funktion `addtogrid`. Der Nachteil bei dieser Implementierung war, daß bei jedem Monte-Carlo Schritt auch alle diejenigen Einträge im Gitter neu berechnet und geschrieben wurden, welche gar keine Veränderung erfahren hatten. Dies wird durch die neue Implementierung der Funktion `updateposition` vermieden.

Dazu wird zunächst berechnet, in welchem Voxel des Gitters sich die verrückte Domäne vor (`oldVoxel`) und nach (`newVoxel`) der Verrückung befindet. Sind `oldVoxel` und `newVoxel` gleich, so wird die Domäne innerhalb eines Voxels verrückt und es müssen nur die Koordinaten im Gitter aktualisiert werden. Sind sie jedoch ungleich, so muß die verrückte Domäne aus der Liste der im Voxel `oldVoxel` befindlichen Domänen herausgenommen werden und alle ihre Nachfolger in der Liste

einen Platz nach vorne verschoben werden, damit keine Leerstelle in der Liste entsteht. Ausserdem wird die Anzahl der in oldVoxel befindlichen Domänen, elementnumber[oldVoxel], dekrementiert.

Anschließend werden die Anzahl der in newVoxel befindlichen Domänen, elementnumber[newVoxel], inkrementiert, die Koordinaten der verrückten Domäne aktualisiert und die verrückte Domäne in die Liste der in newVoxel befindlichen Domänen am Ende eingefügt.

Dadurch, daß keine Doppelschleife mehr für das Durchlaufen des gesamten Gitters benötigt wird, sondern nur noch eine einfache Schleife für das Aktualisieren des hinteren Teils der Liste der in oldVoxel befindlichen Domänen, erhält man gegenüber der alten Funktion eine Reduktion der Laufzeitordnung von quadratisch auf linear.

```
void Ray::updatePosition(const Nucleus &nuc,const Vector &upd,int terr,int segm,
int oldX, int oldY, int oldZ, int newX, int newY, int newZ)
{
int oldVoxel = oldZ*VOXELS*VOXELS+oldY*VOXELS+oldX;
int newVoxel = newZ*VOXELS*VOXELS+newY*VOXELS+newX;
int i = 0;

if(oldVoxel!=newVoxel)
{
//changes in old voxel
for(i=element[terr][segm]; i<elementnumber[oldVoxel]; i++)
{
voxelement1[oldVoxel][i] = voxelement1[oldVoxel][i+1];
voxelement2[oldVoxel][i] = voxelement2[oldVoxel][i+1];
voxelement3[oldVoxel][i] = voxelement3[oldVoxel][i+1];
voxelTerr[oldVoxel][i] = voxelTerr[oldVoxel][i+1];
voxelSegm[oldVoxel][i] = voxelSegm[oldVoxel][i+1];
//indicate elementnumber of element terr,segm in new voxel
element[voxelTerr[oldVoxel][i]][voxelSegm[oldVoxel][i]] = i;
typeband[oldVoxel][i]=typeband[oldVoxel][i+1];
}
elementnumber[oldVoxel]--; //decrease number of elements in old voxel

//changes in new voxel
elementnumber[newVoxel]++; //increase number of elements in new voxel
voxelement1[newVoxel][elementnumber[newVoxel]] = upd[0];
voxelement2[newVoxel][elementnumber[newVoxel]] = upd[1];
voxelement3[newVoxel][elementnumber[newVoxel]] = upd[2];
voxelTerr[newVoxel][elementnumber[newVoxel]] = terr;
voxelSegm[newVoxel][elementnumber[newVoxel]] = segm;
//indicate elementnumber of element terr,segm in new voxel
element[terr][segm] = elementnumber[newVoxel];
typeband[newVoxel][elementnumber[newVoxel]]=Bands[terr][segm];

//indicate in wich voxel element i,j is
voxelnumberx[terr][segm] = newX; //indicate in wich voxel element i,j is
voxelnumbery[terr][segm] = newY;
voxelnumberz[terr][segm] = newZ;
}

else
{
voxelement1[newVoxel][element[terr][segm]] = upd[0];
voxelement2[newVoxel][element[terr][segm]] = upd[1];
voxelement3[newVoxel][element[terr][segm]] = upd[2];
}

nuc.getTer(terr).replaceElement(segm, upd);
}
```

Zum Vergleich der Algorithmen wurde eine Liste von 1,5 Mio Zufallszahlen erstellt. Der Zufallszahlengenerator in beiden Algorithmen wurde ersetzt durch eine Funktion, welche eine Zahl aus dieser Liste einliest. Dadurch hatten beide Algorithmen dieselben Zufallszahlen in derselben Reihenfolge zur Verfügung.

Es wurde mit dem Linux-Befehl time ermittelt, welche Rechenzeit jeder der beiden Algorithmen für einen Simulationslauf von 5000 Monte Carlo Schritten und mit derselben Startkonfiguration benötigt,  $\tilde{t}_a$  für den Algorithmus mit der "alten" Funktion updateposition,  $\tilde{t}_n$  für den Algorithmus mit der "neuen" Funktion. Davon wurde diejenige Rechenzeit abgezogen, welche die Algorithmen für das Einlesen der Startkonfiguration benötigen:  $t_0$ . Es folgende Werte wurden ermittelt:

$$t_0 = 0.797 \text{ sec}$$

$$\tilde{t}_a = 12 \text{ min } 52.625 \text{ sec} \Rightarrow t_a = \tilde{t}_a - t_0 = 12 \text{ min } 51.828 \text{ sec}$$

$$\tilde{t}_n = 1 \text{ min } 8.845 \text{ sec} \Rightarrow t_n = \tilde{t}_n - t_0 = 1 \text{ min } 8.048 \text{ sec}$$

Dies ergibt eine Verkürzung der Rechenzeit:  $\frac{t_a}{t_n} = 11.34$ ; der Algorithmus mit der “neuen” Funktion ist 11.34 mal so schnell wie derjenige mit der “alten”. Relaxierte Zellkernkonfigurationen eines Lymphozyten und eines Fibroblasten sind in Abbildung 4.5 dargestellt.

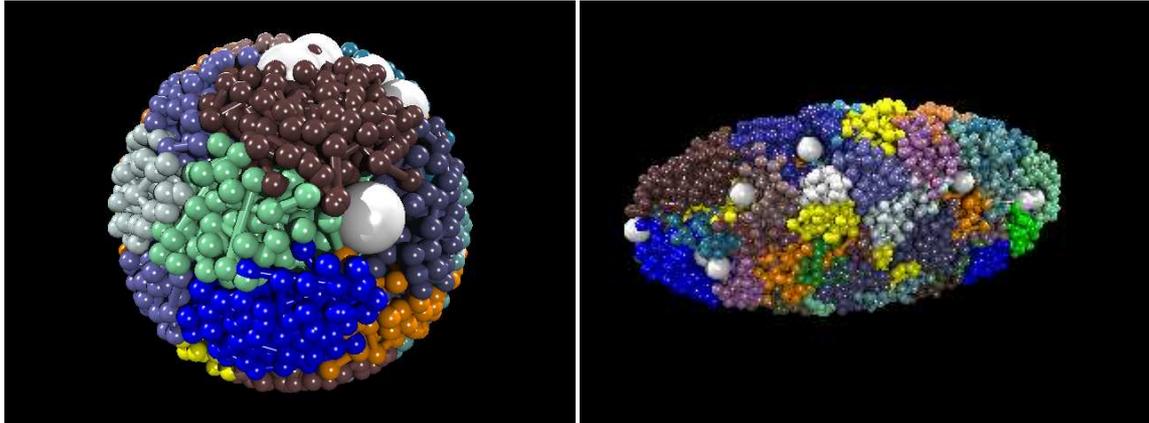


Abbildung 4.5: POVRAY-Visualisierung von relaxierten Kernen mit 30 über den ganzen Kern verteilten SC35-Domänen; links für einen Lymphozyten mit sphärischem Zellkern, rechts für einen Fibroblasten mit ellipsoidem Zellkern. Die Startkonfigurationen für diese beiden Kerne sind in Abb. 4.2 visualisiert.

#### 4.2.6 Algorithmus für das Auffinden von clustern

In diesem Absatz wird ein Algorithmus beschrieben, mit dem man cluster von speckles finden kann. Er lässt sich mit leichten Modifikationen auch auf Chromatindomänen anwenden.

Der Algorithmus clusteridentification weist jedem speckle, ein Label zu. Zu Beginn alle speckles das Label 0. Später werden die speckles im Label die Nummer des Clusters tragen, dem sie angehören. Nun werden alle speckles in einer for-Schleife durchlaufen. Stößt man dabei auf einen speckle  $i$ , der noch das Label 0 trägt, so hat man ein neues cluster gefunden. Ist dies der Fall, so wird der Zähler clusterCounter inkrementiert und dem speckle  $i$  als Label der Wert von clusterCounter zugewiesen. Dann werden alle speckles, welche nach dem gerade bearbeiteten im Array stehen in einer weiteren for-Schleife durchlaufen. Bei jedem dieser speckles  $j$  wird überprüft, ob er Kontakt zu dem speckle  $i$  hat. Wenn ja, so gibt es zwei mögliche Fälle: entweder speckle  $j$  gehört noch nicht zu einem cluster, er hat bisher das Label 0 und bekommt nun als Label den Wert von clusterCounter zugewiesen. Die andere Möglichkeit ist, daß speckle  $j$  bereits zu einem cluster gehört und ein Label ungleich Null trägt. Dann wird in einer dritten for-Schleife das ganze Array durchlaufen und jeder speckle, welcher dasselbe Label trägt wie speckle  $j$ , bekommt das Label von speckle  $i$ . Zuletzt wird noch die tatsächliche Anzahl an clustern bestimmt, diese ist nicht immer gleich clusteCounter, da letzterer auch dann inkrementiert wird, wenn im Zuge des Algorithmus zwei bereits bestehende Teilcluster zu einem großen verschmelzen.

```
int Ray::clusteridentification(const Nucleus &nucc)
{
    speckleNum = nucc.getSpeckleNum();
    speckleRadius = nucc.getSpeckleRadius();
    int clusterCounter = 0, i = 0, j = 0, k = 0;
    int totalCounter = 0, flag = 0;
    int terri = 0, terrj = 0, segmi = 0, segmj = 0;
    int label[1000];
```

```

int clusterSize[1000];
int minSpeckle = 0;
int maxSpeckle = speckleNum-1;
float dist = 0.0;
float contactLength = 2*speckleRadius;

for(i=0; i<1000; i++)
{ label[i]=0;
  clusterSize[i]=0;
}

for(i=minSpeckle; i<=maxSpeckle; i++)
{ terr_i = 49+i/300;
  segm_i = i%300;
  if(label[i]==0)
  { clusterCounter++;
    label[i]=clusterCounter;
  }
  for(j=i+1; j<= maxSpeckle; j++)
  { terr_j = 49+j/300;
    segm_j = j%300;
    //pos(i)-pos(j)
    dist=(Vector(nucc.getTer(terr_i).getElementPos(segmi)
      -nucc.getTer(terr_j).getElementPos(segmj))).length();
    if(dist<=contactLength)
    { if (label[j]==0)
      { label[j]=label[i];
        }
      else if (label[j]!=label[i])
      { flag = label[j];
        for(k=minSpeckle; k<=maxSpeckle; k++)
        { if(label[k]==flag) label[k]=label[i];
          }
        }
      }
    }
  }
}
}
}
for(i=minSpeckle; i<=maxSpeckle; i++)
{ clusterSize[label[i]]++;
}
for(i=1; i<=clusterCounter; i++)
{ if(clusterSize[i]!=0)
  { totalCounter++;
  }
}
return totalCounter;
}

```

### 4.2.7 Auffinden der Oberfläche eines Territoriums

Dieser Algorithmus, implementiert in der Funktion `surfaceFinder`, arbeitet in acht Schritten:

1. Am Anfang werden die Variablen initialisiert.

```

void Ray::surfaceFinder(const Nucleus &nuc, int inTer, surfData &dataSet)
{
  speckleNum = nuc.getSpeckleNum();
  speckleRadius = nuc.getSpeckleRadius();

  Vector centre = terrcentre[inTer];

  int i=0, j=0, k=0, l=0;
  int clusterCounter=0;
  int myLabel=0;
  int state=1, Domain1=0, Domain2=0;
  int domainNumber=nuc.getTer(inTer).getnElements();
  int shellFlag[domainNumber];
  int neighbours[domainNumber][30];
  int neighbourCounter[domainNumber];
  int shells[51][domainNumber];
  int shellSize[51];
  int surface[domainNumber];
  int clusterLabel[domainNumber];

  float contactLength=2*radEl;
  float dist=0.0;

  for(i=0; i<51; i++)

```

```

{
  shellSize[i]=0;
  for(j=0; j<domainNumber; j++)
  {
    shells[i][j]=0;
  }
}
for(i=0; i<domainNumber; i++)
{
  clusterLabel[i]=0;
  neighbourCounter[i]=0;
  shellFlag[i]=0;
  for(j=0; j<30; j++) neighbours[i][j]=0;
}

```

2. Dann werden alle Abstände zwischen allen Chromatindomänen eines Chromosomenterritoriums bestimmt. Berühren sich zwei von ihnen, so werden sie als Nachbarn in Listen abgespeichert.

```

for(i=0; i<domainNumber; i++)
{
  for(j=0; j<domainNumber; j++)
  {
    dist=(Vector(nuc.getTer(inTer).getElementPos(i) - nuc.getTer(inTer).getElementPos(j))).length();
    if(dist<=contactLength && j!=i)
    {
      neighbours[i][neighbourCounter[i]]=j;
      neighbourCounter[i]++;
    }
  }
}

```

3. Anschließend wird ermittelt, ob das Chromosomenterritorium in disjunkte Cluster zerfällt und wenn ja in wieviele. Dazu wird jeder Domäne in einem Array clusterLabel die Nummer des Clusters zugeordnet, in dem sie sich befindet. Isolierte Domänen, das sind solche, die keine Nachbarn noch obigem Kriterium haben, bekommen den Wert 1000 zugeordnet.

```

for(l=0; l<domainNumber; l++)
{
  if(neighbourCounter[l]==0) clusterLabel[l]=1000; //isolated domains
  else if(clusterLabel[l]==0)
  {
    clusterCounter++;
    clusterLabel[l]=clusterCounter;
    shells[l][0]=1;
    shellSize[l]=1;
    i=1;
    do
    {
      for(j=0; j<shellSize[l]; j++)
      {
        Domain1=shells[l][j];
        for(k=0; k<neighbourCounter[Domain1]; k++)
        {
          Domain2=neighbours[Domain1][k];
          if(clusterLabel[Domain2]==0)
          {
            clusterLabel[Domain2]=clusterCounter;
            shells[l+1][shellSize[l+1]]=Domain2;
            shellSize[l+1]++;
          }
        }
      }
      i++;
    }while(shellSize[l]!=0);
    for(i=0; i<51; i++)
    {
      shellSize[i]=0;
      for(j=0; j<domainNumber; j++)
      {
        shells[i][j]=0;
      }
    }
  }
}

```

4. Nun werden die Schwerpunkte bzw. Barycenters der einzelnen Cluster bestimmt.

```

Vector clusterCentres[clusterCounter+1];
Vector null(0.0,0.0,0.0);
int clusterSizes[clusterCounter+1];
int startDom[clusterCounter+1];
float minDist[clusterCounter+1];
for(i=0; i<=clusterCounter; i++)
{
    clusterSizes[i]=0;
    clusterCentres[i]=null;
    minDist[i]=1000.0;
}
for(i=0; i<domainNumber; i++)
{
    myLabel=clusterLabel[i];
    if(myLabel!=1000)
    {
        Vector dummy=clusterCentres[myLabel];
        clusterSizes[myLabel]++;
        clusterCentres[myLabel]=dummy+(Vector)nuc.getTer(inTer).getElementPos(i);
    }
}
for(i=1; i<=clusterCounter; i++)
{
    clusterCentres[i]=clusterCentres[i]/(1.0*clusterSizes[i]);
}

```

5. Es wird anschließend für in jedem Cluster die Domäne ermittelt, deren Position dem Schwerpunkt am nächsten ist. Sie wird dann als Zentrum des Clusters deklariert und sie wird der Schicht 1 zugeordnet (siehe unten).

```

for(i=0; i<domainNumber; i++)
{
    myLabel=clusterLabel[i];
    dist=(Vector(nuc.getTer(inTer).getElementPos(i) - clusterCentres[myLabel])).length();
    if(dist<minDist[myLabel])
    {
        minDist[myLabel]=dist;
        startDom[myLabel]=i;
    }
}
for(i=1; i<=clusterCounter; i++)
{
    while(neighbourCounter[startDom[i]]==0)
    {
        startDom[i]++;
    }
    shellFlag[startDom[i]]=1;
}

```

6. Nun wird das Cluster schichtweise durchlaufen. Das funktioniert folgendermaßen: alle Nachbarn der Domäne im Zentrum werden der Schicht 2 zugeordnet. Dann wird in einer for-Schleife die Schicht 2 durchlaufen, und alle Nachbarn von Domänen der Schicht 2, welche noch keiner Schicht zugeordnet sind (welche also nicht in Schicht 1 sind), werden der Schicht 3 zugeordnet. Dann wird Schicht 3 durchlaufen und so fort. Im n-ten Schritt wird in einer for-Schleife Schicht n durchlaufen, und alle Nachbarn von Domänen der Schicht n, welche noch keiner Schicht zugeordnet sind, werden der Schicht n+1 zugeordnet.

```

for(l=1; l<=clusterCounter; l++)
{
    shells[l][0]=startDom[l];
    shellSize[l]=1;
    i=1;
    do
    {
        for(j=0; j<shellSize[i]; j++)
        {
            Domain1=shells[i][j];
            for(k=0; k<neighbourCounter[Domain1]; k++)
            {
                Domain2=neighbours[Domain1][k];
            }
        }
    }
}

```

```

        if (shellFlag[Domain2]==0)
        {
            shellFlag[Domain2]=i+1;
            shells[i+1][shellSize[i+1]]=Domain2;
            shellSize[i+1]++;
        }
    }
    }
    i++;
}while(shellSize[i]!=0);
for(i=0; i<51; i++)
{
    shellSize[i]=0;
    for(j=0; j<domainNumber; j++)
    {
        shells[i][j]=0;
    }
}
}
}

```

7. Wenn in alle Cluster ganz durchlaufen sind, werden die Oberflächen ermittelt: jede nicht-isolierte Domäne, die keine Nachbarn einer höheren Schicht hat, gehört zur Oberfläche. Domänen im Inneren des Territoriums werden gekennzeichnet durch  $surface[i]=0$ ; Domänen auf der Oberfläche durch  $surface[i]=1$  und isolierte Domänen durch  $surface[i]=2$ .

```

for(i=0; i<domainNumber; i++)
{
    state=0;
    Domain1=i;
    if(neighbourCounter[Domain1]==0) surface[Domain1]=2;
    else if(shellFlag[Domain1]==0) surface[Domain1]=3;
    else if(clusterSizes[clusterLabel[i]]<4) surface[Domain1]=1;
    else
    {
        for(j=0; j<neighbourCounter[Domain1]; j++)
        {
            Domain2=neighbours[Domain1][j];
            if(shellFlag[Domain2]>shellFlag[Domain1]) state=1;
        }
        if(state==1) surface[Domain1]=0;
        else surface[Domain1]=1;
    }
}
}

```

8. Ist die Oberfläche des analysierten Territoriums an bestimmten Stellen konkav (dies passiert oft an Kontaktstellen mit den speckles), so gibt es Fälle, in denen der Algorithmus bis zu diesem Punkt nicht alle Oberflächendomänen gefunden hat. Daher werden alle Domänen, welche sich im inneren befinden, einem weiteren Kriterium unterworfen. Dafür werden aus der Liste der Nachbarn einer Domäne  $i$  alle möglichen Viererkombinationen von Domänen  $j$ ,  $k$ ,  $l$  und  $m$  bestimmt. Für jede dieser Viererkombinationen wird ermittelt, ob sich die untersuchte Domäne  $i$  in dem Tetraeder befindet, welcher durch die Domänen  $j$ ,  $k$ ,  $l$  und  $m$  aufgespannt wird. Wenn diese Bedingung für mindestens einen dieser Tetraeder erfüllt ist, so gehört die untersuchte Domäne  $i$  zum Inneren ihres Territoriums. Gibt es jedoch keinen von Nachbarn der Domäne  $i$  aufgespannten Tetraeder, welcher die Domäne  $i$  enthält, so gehört die Domäne  $i$  zur Oberfläche ihres Territoriums. Durch diese Kombination von Abfragen wird überprüft, ob sich die Domäne  $i$  im Inneren des von ihren Nachbarn aufgespannten Polyeders befindet. Der Polyeder wurde dazu in Tetraeder zerlegt.

```

Vector corner[4];
Vector diff[4];
Vector normal, point;
int m=0, n=0, testCounter=0, Domain0;
int State[4];
int oldSign=0, mySign=0, sign1=0, sign2=0;
float State1=0.0, State2=0.0;
float dists[4];

for (i=0; i<domainNumber; i++)
{

```



Quotienten aus diesen Zählraten gebildet werden. Diese Quotienten werden in 5.4.2, 5.4.3 und 5.4.4 ausgewertet.

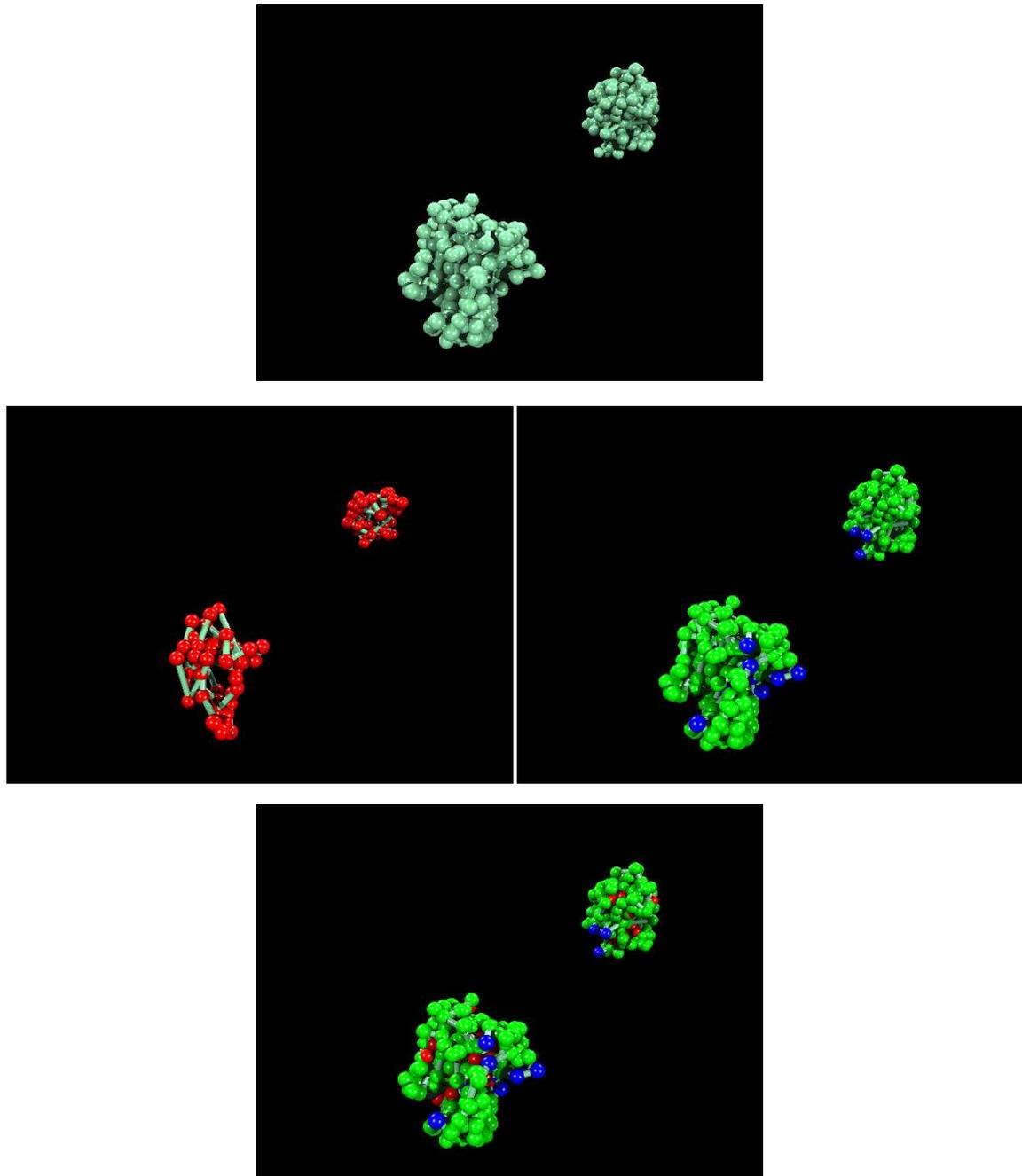


Abbildung 4.6: POV-Ray-Visualisierung des Ergebnisses des Algorithmus `surfaceFinder` anhand der Territorien 1 und 2 (Chromosom 1) eines relaxierten Fibroblastenzellkerns. In der ersten Teilabbildung sind die Territorien vollständig in der Farbe dargestellt, die sie auch bei einer Visualisierung des ganzen Zellkerns tragen. Die Teilabbildungen 2 und 3 darunter zeigen die Zerlegung in Inneres (rot) auf der einen, Oberfläche (grün) und insulierte Domänen (blau) auf der anderen Seite. Teilabbildung 4 zeigt die Superposition der Strukturen aus 2 und 3.

# Kapitel 5

## Ergebnisse

### 5.1 Parametervariation

Das im vorhergehenden Kapitel vorgestellte SCD-Modell und die in dieser Arbeit entwickelten Erweiterungen beinhalten mehrere Parameter, die sinnvoll gewählt werden müssen. Dies sind unter anderem die Stärken der Wechselwirkungen zwischen den verschiedenen Domänen,  $\epsilon_0$ ,  $\epsilon_1$ ,  $\epsilon_2$  und  $\epsilon_3$  sowie die Größe und die Anzahl der speckles bzw. SC35-Domänen. Die Werte für die Potentialstärke der Wechselwirkung zwischen 2 speckles,  $\epsilon_2$ , und für die Potentialstärke der Wechselwirkung zwischen G-Banden und speckles,  $\epsilon_3$ , wurden auf den Wert  $1.5 k_B T$  festgesetzt. Diese Wechselwirkungen haben damit dieselbe Stärke wie die Wechselwirkung zweier Chromatindomänen untereinander,  $\epsilon_0 = 1.5 k_B T$ , welche von Gregor Kreth in [Kre00] gewählt wurde und sich in Relaxationen bewährt hat. Bei kleineren Werten von  $\epsilon_0$ , insbesondere im Fall  $\epsilon_0 < 1.0$  können sich zwei Chromatindomänen gegenseitig vollständig durchdringen. Dies ist unerwünscht. Andererseits möchte man  $\epsilon_0$  so klein halten wie möglich, um der Natur weitestgehend zu entsprechen.

Es wird mit Fibroblasten-Zellkernkonfigurationen gearbeitet, die 30 speckles enthalten. Der Radius eines speckles ist dreimal so groß wie der Radius einer Chromatindomäne, der seinerseits 250 nm beträgt.

$$r_{\text{speckle}} = 3 \cdot r_{\text{Chr}} = 0.75 \mu\text{m} \quad (5.1)$$

Der Wert des Parameters  $\epsilon_1$ , der Potentialstärke der Wechselwirkung zwischen R-Banden und speckles, wurde variiert zwischen  $\epsilon_1 = 0.0 k_B T$  und  $\epsilon_1 = 12.0 k_B T$ , wobei zwischen  $\epsilon_1 = 0.0 k_B T$  und  $\epsilon_1 = 2.0 k_B T$  eine kleine Schrittweite gewählt wurde und zwischen  $\epsilon_1 = 2.0 k_B T$  und  $\epsilon_1 = 12.0 k_B T$  eine größere. Für jeden dieser Werte von  $\epsilon_1$  wurden vier bis zehn Relaxationen durchgeführt.

Die relaxierten Konfigurationen wurden anhand zweier Kriterien untersucht. Diese Kriterien liefern mit experimentellen Daten und biologischen Modellvorstellungen vergleichbare Zahlen und Messgrößen (vergleiche Abb. 5.2). Bevor diese Kriterien genauer beschrieben werden, soll an dieser Stelle kurz darauf eingegangen werden, was mit einem Kontakt zwischen zwei Domänen gemeint ist: zwei Domänen haben Kontakt mit einander, wenn der Abstand zwischen ihren Mittelpunkten kleiner oder gleich der Summe ihrer Radien ist. Dabei können die Domänen gleichen Typs, also auch gleicher Größe sein, oder auch verschiedener Größe. Kontakt beinhaltet auch den Fall, daß sich zwei Domänen durchdringen (vergleiche Abbildung 5.1).

Nun zu den Kriterien für die Auswertung:

- **Kriterium I.** In einer Variablen *GBandsCount* werden alle Kontakte zwischen G-Banden

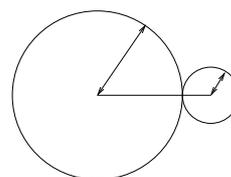


Abbildung 5.1: Skizze zur Illustration des Kontakts zwischen einem speckle und einer Chromatindomäne. Der Radius des speckles ist dreimal so groß wie der Radius der Chromatindomäne

und speckles, kurz G-Banden-Kontakte, aufsummiert:

$$GBandsCount = \sum_{i=1}^{46} \sum_{j=0}^{N_i-1} \sum_{k=0}^{N_{sp}-1} \delta_{ijk}^G$$

wobei 46 die Anzahl der Territorien,  $N_i$  die Größe des Territorium  $i$ ,  $N_{sp}$  die Anzahl der speckles im Kern ist und für  $\delta_{ijk}^G$  gilt:

$$\delta_{ijk}^G = \begin{cases} 1 & \text{falls Domäne } j \text{ in Territorium } i \text{ zu einer G-Bande gehört} \\ & \text{und Kontakt zu speckle } k \text{ hat} \\ 0 & \text{sonst} \end{cases}$$

Ebenso werden in einer Variablen  $RBandsCount$  alle Kontakte zwischen R-Banden und speckles, kurz R-Banden-Kontakte, aufsummiert:

$$RBandsCount = \sum_{i=1}^{46} \sum_{j=0}^{N_i-1} \sum_{k=0}^{N_{sp}-1} \delta_{ijk}^R$$

Bezeichnungen analog zu den G-Banden. Anschließend wird der Quotient von G-Banden-Kontakten durch R-Banden-Kontakte, **G2RContacts** gebildet:

$$G2RContacts = \frac{GBandsCount}{RBandsCount} \quad (5.2)$$

Die Auswertung von Kriterium I ist implementiert in der Funktion `clusterRatio`. Experimentellen Beobachtungen von L. Shopland et al. (siehe [SJB<sup>+</sup>03]) ergeben für **G2RContacts** einen Wert von  $\frac{1}{8}$ .

- **Kriterium II.** In der Variablen  $contacts_i$  wird für das Territorium  $i$  die Anzahl an Speckles abgelegt, zu denen irgend eine Domäne des Territoriums  $i$  Kontakt hat:

$$contacts_i = \sum_{l=0}^{N_{sp}-1} \delta_{il}$$

wobei  $N_{sp}$  hier wieder die Anzahl der speckles im Kern ist und für  $\delta_{il}$  gilt:

$$\delta_{il} = \begin{cases} 1 & \text{falls irgend eine Domäne des Territoriums Kontakt zu speckle } l \text{ hat} \\ 0 & \text{sonst} \end{cases}$$

Anschließend wird die mittlere Anzahl an kontaktierten Speckles pro Territorium, **contPerTer** gebildet:

$$contPerTer = \frac{1}{46} \sum_{i=1}^{46} contacts_i \quad (5.3)$$

Die Auswertung von Kriterium II ist implementiert in der Funktion `specklesPerTerritory`. Biologische Modellvorstellungen von L. Shopland et al. (siehe [SJB<sup>+</sup>03]) fordern für **contPerTer** einen Wert zwischen 3 und 4.

Bei der Variation von  $\epsilon_1$  wurden für jeden Potentialwert mehrere Relaxationen durchgeführt und Mittelwert sowie Standardabweichung der beiden Größen **G2RContacts** und **contPerTer** gebildet und in Abbildung 5.2 gegen  $\epsilon_1$  aufgetragen (rote Kurven).

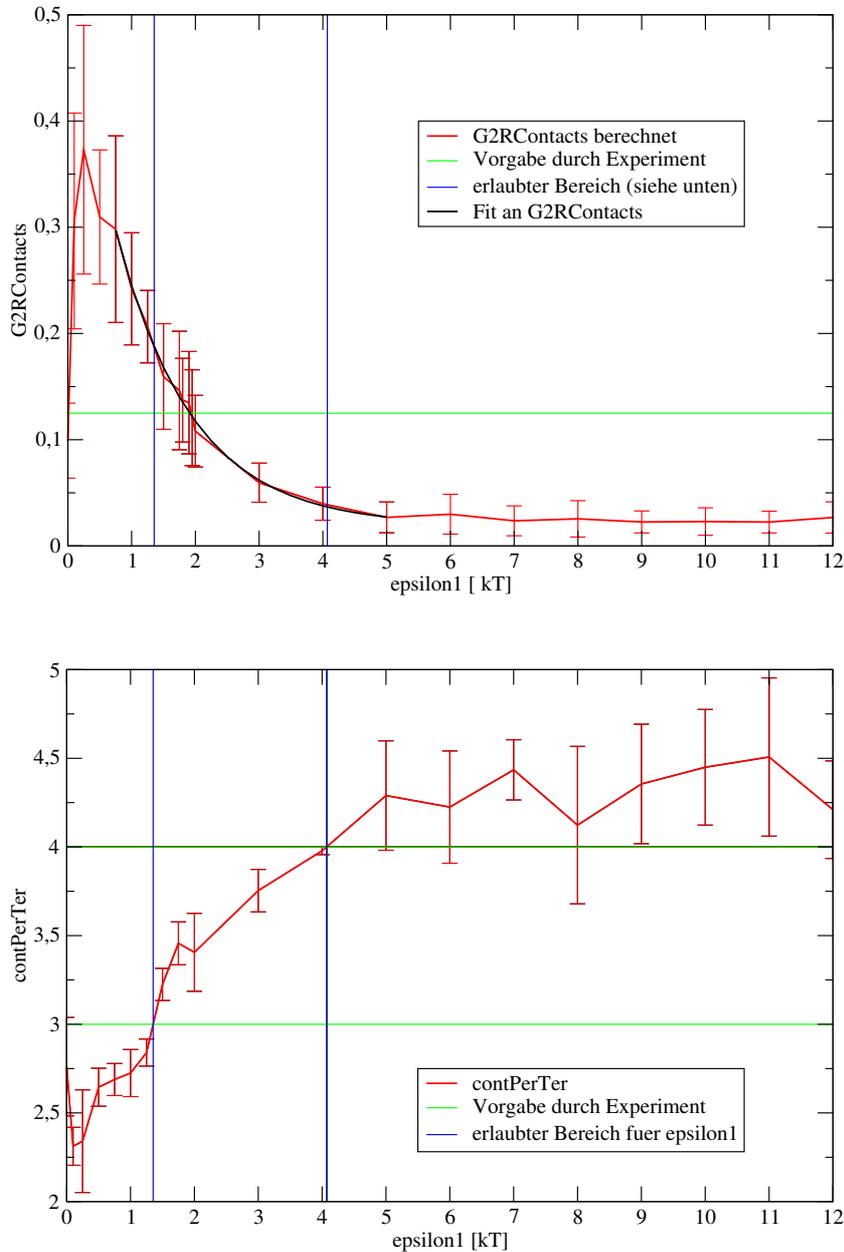


Abbildung 5.2: **Variation des Parameters  $\epsilon_1 = \epsilon_1$ , der R-Banden-Potentialstärke**, aufgetragen in Einheiten von  $k_B T$ . Die Werte für die Potentialstärke der Wechselwirkung zwischen 2 speckles,  $\epsilon_2$ , und für die Potentialstärke der Wechselwirkung zwischen G-Banden und speckles,  $\epsilon_3$ , wurden auf den Wert  $1.5 k_B T$  festgesetzt. Der Radius der speckles ist **dreimal** so groß wie der der Chromatindomänen. In der oberen Teilabbildung (Kriterium I) ist der Quotient von G-Bandenkontakten durch R-Bandenkontakte, **G2RContacts**, dargestellt. Die grüne Gerade stellt den Wert dar, der von L. Shopland et al. experimentell ermittelt wurde (vergleiche [SJB<sup>+</sup>03]). Die blauen Linien stellen den nach Kriterium II (untere Teilabbildung) erlaubten Bereich für den Parameter  $\epsilon_1$  dar. Der Schnittpunkt der grünen Linie mit der roten Kurve befindet sich im nach Kriterium II erlaubten Bereich, die Kriterien I und II sind mit einander vereinbar. In der unteren Teilabbildung dargestellt ist die mittlere Anzahl an Speckles, welche zu einem Territorium Kontakt haben, **contPerTer**. Die grünen Geraden stellen die Werte dar, die einem biologischen Modell entsprechen (vergleiche [SJB<sup>+</sup>03]). Die blauen Linien ergeben sich aus den Schnittpunkten der grünen Geraden mit der Kurve **contPerTer**.

In die zweite Teilabbildung von Abbildung 5.2 wurden zwei waagrechte Geraden bei  $y = 3$  und  $y = 4$  eingetragen (grün). Sie stellen den nach Kriterium II erlaubten Bereich für `contPerTer` dar. Die Projektionen auf die  $x$ -Achse der Schnittpunkte zwischen diesen Geraden und der Kurve `contPerTer` liefern den nach Kriterium II erlaubten Bereich für  $\epsilon_1$  (dieser wird durch die senkrechten blauen Geraden markiert). Es ergibt sich

$$\begin{aligned}\epsilon_{1,min} &= 1.35377 \pm 0.03799 k_B T \\ \epsilon_{1,max} &= 4.06977 \pm 0.09477 k_B T\end{aligned}\tag{5.4}$$

Ebenso wurde in die erste Teilabbildung von Abbildung 5.2 eine waagrechte Gerade bei  $y = \frac{1}{8}$  eingefügt. Diese stellt den nach Kriterium I optimalen Wert von `G2RContacts` dar. An den mittleren Teil der Kurve `G2RContacts` wurde ein Fit gelegt und anschließend der Schnittpunkt dieses Fits mit der Geraden  $y = \frac{1}{8}$  ermittelt. Die Projektion dieses Schnittpunkts auf die  $x$ -Achse liefert den nach Kriterium I optimalen Wert für  $\epsilon_1$ ,  $\epsilon_{1,opt}$ . Die Parametrisierung des Fits wurde wie folgt gewählt:

$$\begin{aligned}y &= A0 \cdot e^{-A1x} + A2 \text{ mit} \\ A0 &= 0.519037 \\ A1 &= 0.833047 \\ A2 &= 0.0192392\end{aligned}\tag{5.5}$$

Die oberen und unteren Enden der Fehlerbalken wurden ebenso gefittet und die Schnittpunkte dieser Fits mit der Geraden  $y = \frac{1}{8}$  ermittelt; ihre Projektionen auf die  $x$ -Achse dienen der Ermittlung des Fehlers, mit dem  $\epsilon_{1,opt}$  behaftet ist. Es ergibt sich:

$$\epsilon_{1,opt} = 1.90961^{+0.4248}_{-0.4616} k_B T\tag{5.6}$$

Es gilt also:  $\epsilon_{1,min} < \epsilon_{1,opt} < \epsilon_{1,max}$ . Der Wert  $\epsilon_{1,opt}$  liegt folglich im erlaubten Bereich (auch innerhalb der Fehlergrenzen), die Kriterien I und II sind mit einander vereinbar.

Die Kurve `G2RContacts` in der ersten Teilabbildung von Abb. 5.2 steigt für kleine  $\epsilon_1$  zunächst stark an, fällt dann für größere Werte von  $\epsilon_1$  jedoch wieder ab. Dies läßt sich physikalisch erklären: mit  $\epsilon_1$  wächst die Stärke der Wechselwirkung zwischen R-Banden und speckles. Diese Wechselwirkung hat einen akktraktiven und einen repulsiven Teil. Der attraktive Teil ist für den fallenden hinteren Teil der Kurve `G2RContacts` verantwortlich: R-Banden werden stärker zu den speckles hingezogen, es gibt mehr Kontakte zwischen R-Banden und speckles und der Nenner des Quotienten `G2RContacts` wächst. Zusätzlich ist es möglich, daß R-Bandendomänen G-Bandendomänen von der Oberfläche der speckles verdrängen, was den Zähler von `G2RContacts` verkleinert und dadurch die Verringerung dieses Wertes verstärkt. Der repulsive Teil der Wechselwirkung ist für den Anstieg der Kurve `G2RContacts` verantwortlich: R-Banden, die bei  $\epsilon_1 = 0$  die speckles beliebig durchdringen können, erfahren mit steigendem  $\epsilon_1$  eine steigende Abstoßung von den speckles, die Anzahl der Kontakte zwischen speckles und R-Banden nimmt ab, der Nenner von `G2RContacts` verkleinert sich.

Diese beiden gegenläufigen Effekte führen zu einem Maximum und einem Wendepunkt in der Auftragung von `G2RContacts`. Bei großen Werten von  $\epsilon_1$  fällt die Kurve unter ihren Wert bei  $\epsilon_1 = 0$ , es überwiegt hier also der attraktive Teil der Wechselwirkung.

Die Zunahme der Größe `contPerTer` mit steigendem  $\epsilon_1$  kann folgendermaßen erklärt werden: hier wird ermittelt, zu wievielen speckles ein bestimmtes Territorium Kontakt hat. Es ist jedoch egal, ob die Kontakte durch G-, C- oder R-Banden hervorgerufen werden. Da die Attraktivität zwischen speckles und R-Banden zunimmt (s.o.), kommt es zu insgesamt mehr Kontakten. Die Kurve scheint jedoch einen Sättigungsbereich zu erreichen, dies könnte daran liegen, daß die Anziehungskraft der speckles auf die R-Banden zu Spannungen in den Territorien führt, die Chromatinlinker zwischen den Chromatinomänen sowie die Tensegrity-Wechselwirkung der Territorien verhindern

eine allzu große Verzerrung der Chromatindomänenkette und begrenzen dadurch die mögliche Anzahl an Kontakten.

Rechnet man den Wert  $\epsilon_{1,opt}$  für  $T = 310K = 36.85^\circ C$  auf andere Energieeinheiten um, so erhält man:

$$\begin{aligned}\epsilon_{1,opt} &= 1.90961_{-0.4616}^{+0.4248} k_B T = 8.17_{-1.97}^{+1.82} \cdot 10^{-21} J = 0.051_{-0.012}^{+0.011} eV \\ &\doteq 4.91997_{-1.189}^{+1.094} \frac{kJ}{mol} = 1.17587_{-0.284}^{+0.261} \frac{kcal}{mol}\end{aligned}$$

Wasserstoffbrückenbindungen in wässriger Umgebung haben die Bindungsenergie:

$$\epsilon_H \sim 1 \frac{kcal}{mol} = 4.184 \frac{kJ}{mol}$$

Die Wechselwirkung zwischen speckles und R-Bandendomänen entspricht in ihrer Stärke folglich ungefähr 1.18 Wasserstoffbrückenbindungen. Dies ist erstaunlich, da die Kontaktfläche zwischen speckles und R-Bandendomänen wesentlich mehr Wasserstoffbrücken bereitstellen könnte. Eine mögliche Ursache für die viel geringere Anzahl der effektiv wirksamen Wasserstoffbrückenbindungen ist, daß sich die DNA schnell an den SC35-Proteinkomplexen vorbeibewegt.

## 5.2 Bestimmung einer unteren Schranke für die Größe der speckles

Die im vorigen Abschnitt 5.1 durchgeführte Variation der Stärke der Wechselwirkung zwischen R-Banden und speckles und die damit verbundenen Relaxationen und Rechnungen wurden erneut ausgeführt, wobei allerdings die Größe der speckles herabgesetzt wurde. Ihr Radius betrug nur noch doppelt so viel wie derjenige einer Chromatindomäne im Gegensatz zu dreimal so viel in der vorangegangenen Rechnung.

$$r_{speckle} = 2 \cdot r_{Chr} = 0.5 \mu m \quad (5.7)$$

Es wurden wieder die Kriterien I und II ausgewertet (siehe Abbildung 5.3). Mit Hilfe der Vorgaben des biologischen Modells in Kriterium II wurde wieder ein erlaubter Bereich für  $\epsilon_1$  ermittelt. In diesem Fall ist der erlaubte Bereich nur nach unten beschränkt, da in dem dargestellten Bereich kein Schnittpunkt zwischen der Geraden  $y = 4$  und der Kurve contPerTer vorliegt. Die Kurve contPerTer erreicht ihren Sättigungsbereich früher als in dem in Abbildung 5.2 dargestellten Fall mit den etwas größeren speckles, dies ist jedoch nicht verwunderlich, denn kleinere speckles können mit ihrer kleineren Oberfläche auch insgesamt weniger Kontakte ausbilden.

$$\epsilon_{1,min} = 2.03069 \pm 0.04801 k_B T \quad (5.8)$$

Man findet bei der Kurve G2RContacts wie im vorherigen Fall (siehe 5.1) einen ansteigenden und einen fallenden Teil. Die Ursachen dafür sind natürlich dieselben wie in 5.1 beschrieben, jedoch wurde  $\epsilon_1$  hier in größeren Schritten variiert, sodaß die Kurve größer ist. Mit den experimentellen Daten zu Kriterium I wurde der optimale Wert für  $\epsilon_1$  ermittelt. Mit Hilfe des Fits

$$\begin{aligned}y &= A0 \cdot e^{-A1x} + A2 \text{ mit} \\ A0 &= 0.434685 \\ A1 &= 0.814875 \\ A2 &= 0.00320794\end{aligned} \quad (5.9)$$

ergab sich für  $\epsilon_{1,opt}$ :

$$\epsilon_{1,opt} = 1.56139 k_B T \quad (5.10)$$

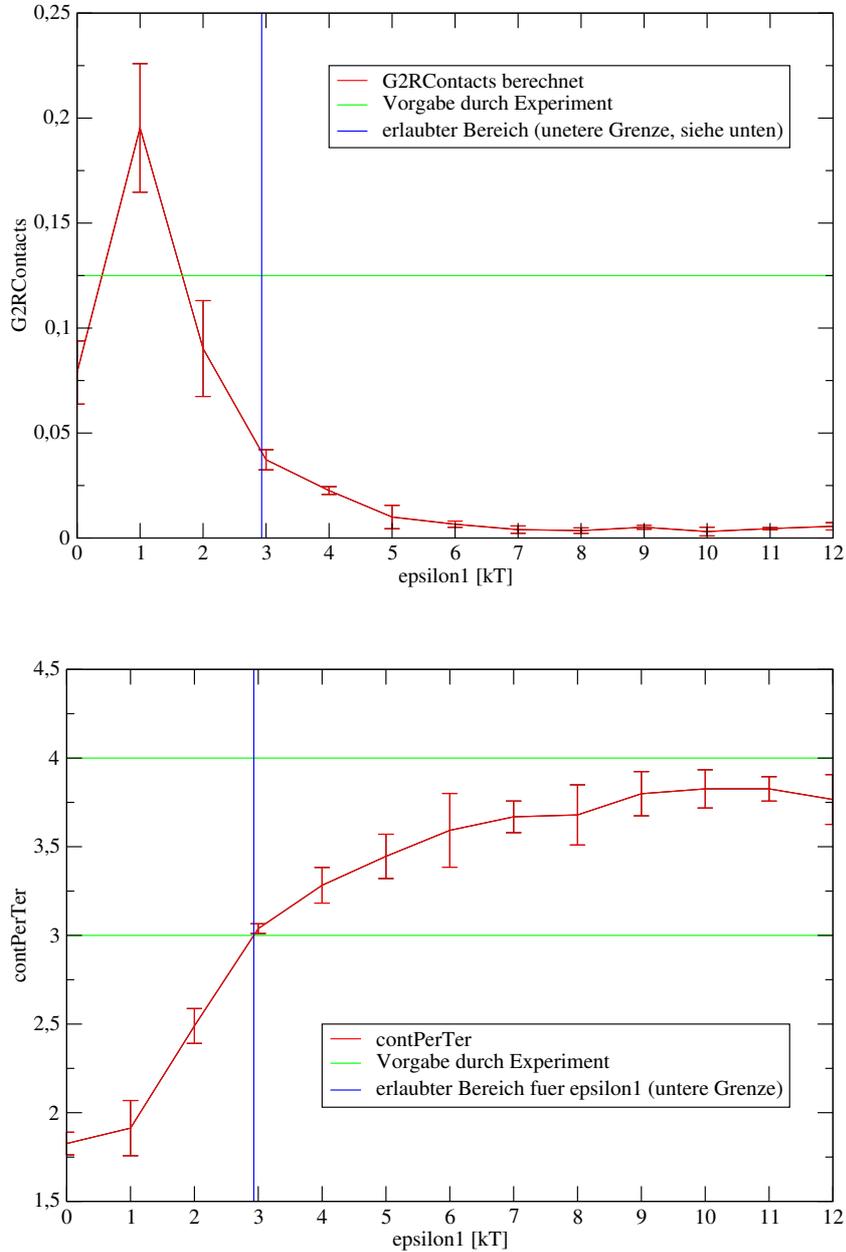


Abbildung 5.3: Variation des Parameters R-Banden-Potentialstärke bei kleineren speckles. Alle Parameter bis auf den Radius der speckles sind gewählt wie in Abbildung 5.2:  $\epsilon_2 = \epsilon_3 = 1.5 k_B T$ . Der Radius der speckles ist hier **doppelt** so groß wie der Radius einer Chromatindomäne. In der oberen Teilabbildung ist wieder der Quotient von G-Bandenkontakten durch R-Bandenkontakte, **G2RContacts**, dargestellt. In der unteren Teilabbildung ist die mittlere Anzahl an Speckles, welche zu einem Territorium Kontakt haben, **contPerTer**. Die grünen und blaue Geraden haben dieselbe Bedeutung wie in Abbildung 5.2. Der Schnittpunkt zwischen der roten Kurve und der grünen Geraden in Teilabbildung 1 liegt hier außerhalb des nach Kriterium II erlaubten Bereiches, die beiden Kriterien I und II sind nicht mit einander vereinbar.

Es gilt also:  $\epsilon_{1,opt} < \epsilon_{1,min}$ . Im Falle der kleineren speckles liegt der Wert  $\epsilon_{1,opt}$  folglich nicht im laut Kriterium II erlaubten Bereich, die Kriterien I und II sind, im Gegensatz zu dem in Abb. 5.2 dargestellten Fall, nicht mit einander vereinbar. Damit liefern obige Rechnungen eine untere Grenze für die mittlere Größe der speckles sowie für das Gesamtvolumen aller speckles im Zellkern.

### 5.3 Relaxationen mit den ausgewählten Parameterwerten

Mit den in den vorigen beiden Abschnitten erhaltenen Werten für die Parameter  $\epsilon_1$ , die Stärke der Wechselwirkung zwischen R-Banden und speckles, sowie für die Größe der speckles wurden auf mehrere Rechner verteilt insgesamt 204 Relaxationen von jeweils 400000 Monte Carlo Schritten durchgeführt. Die Endkonfigurationen dieser Simulationsläufe wurden anhand von 5 Kriterien untersucht, zu denen auch die Kriterien I und II aus den vorhergehenden Abschnitten gehören.

- **Kriterium I.** Das Verhältnis von G-Banden-Kontakten zu R-Banden-Kontakten im Zellkern, **G2RContacts** wurde ermittelt und dem experimentell gemessenen Wert in Tabelle 5.1 gegenübergestellt.
- **Kriterium II.** Die mittlere Anzahl an kontaktierten Speckles pro Territorium, **contPerTer** wurde ermittelt und mit dem Wert verglichen, den ein biologisches Modell nahelegt. Werte ebenfalls in Tabelle 5.1.

Parameter	gemessen bzw. postuliert	berechnet
G2RContacts	$\frac{1}{8} = 0.125$	$0.1190 \pm 0.0290$
contPerTer	3 - 4	$3.4266 \pm 0.2080$

Tabelle 5.1: Auswertung der Kriterien I und II

Der Wert G2RContacts ist leicht verschoben gegenüber dem im vorigen Abschnitt eingestellten Wert. Dies liegt daran, daß im vorigen Abschnitt im Zuge der Parametervariation für jeden festen Wert von  $\epsilon_1$ , der Stärke der Wechselwirkung zwischen R-Banden und speckles, wesentlich weniger Konfigurationen berechnet werden konnten, als in diesem Abschnitt für das eingestellte  $\epsilon_{1,opt}$ . Der in diesem Abschnitt ermittelte Wert ist daher genauer. Innerhalb der Fehlergrenzen ergibt sich jedoch eine gute Übereinstimmung.

Der Wert contPerTer liegt vollständig im nach Kriterium II vorgeschriebenen Bereich.

- **Kriterien III und IV.** Untersucht wurden die Kontakte von Domänen der Genregionen 17q21 mit speckles (Kriterium III) und von Domänen der Genregion 7p21 mit speckles (Kriterium IV). Die Genregion 17q21 ist mehrheitlich eine R-Bande und umfasst 10 Chromatindomänen, die Genregion 7p21 ist mehrheitlich eine G-Bande und umfasst 9 Chromatindomänen (schematische Darstellungen dieser beiden Genregionen in Abbildung 5.4).



Abbildung 5.4: Schematische Darstellung der Genregionen 17q21 (links) und 7p21 (rechts). G-Banden sind schwarz, R-Banden weiß.

Für jede der 204 relaxierten Konfigurationen wird ermittelt, zu wievielen speckles die Genregion 17q21 Kontakt hat. Anschließend wird gezählt in wievielen der 204 Konfigurationen die Genregion 17q21 zu keinem speckle Kontakt hat, in wievielen sie Kontakt zu einem speckle hat und in wievielen sie Kontakt zu mehr als einem speckle hat. Mit den Kontakten von Genregion 7p21 wird analog verfahren. Die beiden Genregionen kommen in normalen

Genregion		Prozentsatz		
		0 Kontakte	1 Kontakt	>1 Kontakte
7p21 (G)	gemessen	37 ± 4.79	52 ± 6	11 ± 2.24
	berechnet	56.86 ± 9.26	30.88 ± 6.05	12.25 ± 3.31
17q21 (R)	gemessen	1 ± 0.56	34 ± 4.29	63 ± 6.46
	berechnet	10.29 ± 2.97	22.06 ± 4.83	67.65 ± 10.49

Tabelle 5.2: Auswertung der Kriterien III und IV; aufgelistet ist, bei welchem prozentualen Anteil der 204 relaxierten Kerne die untersuchten Genregionen Kontakt zu keinem, einem oder mehr als einem speckle haben.

Zellkernen doppelt vor, da sie sich auf homologen Chromosomen befinden. Die Genregion 17q21 gibt es je einmal in den Territorien 33 und 34, die Genregion 7p21 je einmal in den Territorien 13 und 14. Im experiment kann jedoch nicht zwischen den beiden Homologen unterschieden werden. Um die oben ermittelten Kontakte zu speckles mit experimentellen Daten vergleichen zu können, müssen also die oben gezählten Werte der beiden Homologen addiert werden. Diese Summen werden anschließend durch 204 geteilt, um relative Angaben zu erhalten. Experimentelle Untersuchungen dazu wurden von L. Shopland et al. in [SJB<sup>+</sup>03] durchgeführt. In Tabelle 5.2 sind die Ergebnisse des SCD-Modells den experimentellen Ergebnissen gegenübergestellt. Die Auswertung der Kriterien III und IV ist implementiert in der Funktion `bandContactComparison`.

Der Fehler einer Zählrate  $q = \frac{n}{N}$  wurde bestimmt gemäß

$$\frac{\Delta q}{q} = \frac{\Delta n}{n} + \frac{\Delta N}{N} = \frac{\sqrt{n}}{n} + \frac{\sqrt{N}}{N} = \frac{1}{\sqrt{n}} + \frac{1}{\sqrt{N}} \quad (5.11)$$

Innerhalb der Fehlergrenzen stimmen nur die Zählraten der letzten Spalte, das heißt für mehr als einen Kontakt zu speckles, überein. Qualitativ richtig wiedergegeben werden jedoch auch die übrigen Zählraten in der letzten Zeile insofern als die Genregion 17q21 sowohl im Experiment, als auch nach dem SCD-Modell, am seltensten keinen Kontakt zu speckles hat, öfter Kontakt zu einem speckle und deutlich am häufigsten Kontakt zu mehr als einem speckle.

Die ersten beiden Zählraten für die Genregion 7p21 werden auch qualitativ nicht korrekt wiedergegeben. Eine mögliche Ursache dafür könnte sein, daß die Rundungsoperation, die beim SCD-Modell erfolgt, indem Chromatindomänen der Größe 1 Mbp angenommen werden, die Größe und den Typ der randständigen Domänen der Genregion 7p21 verfälschen.

Weiterhin wurden die Territorien 13 und 14 (Chromosom 7) und 33 und 34 (Chromosom 17) mit dem Verfahren der virtuellen Mikroskopie dargestellt. Bei diesem Verfahren (siehe [Kre00]) werden die Bildinformationen einer Zellkernkonfiguration mit der Punktbildfunktion eines hochauflösenden Lichtmikroskops, zum Beispiel eines konfokalen Laser Scanning Mikroskops (CLSM) gefaltet und eventuell mit einem Rauschen versehen. Dadurch erhält man Bilder, welche direkt mit experimentellen Daten vergleichbar sind und welche mit bereits bestehenden Bildverarbeitungsprogrammen analysiert werden können. Die Berechnung spezieller physikalischer Werte mit diesem Verfahren erfordert jedoch wesentlich mehr Rechenleistung und Rechenzeit als die Berechnung derselben Werte direkt aus den Zellkernkonfigurationen, wie in dieser Arbeit durchgeführt. Abbildung 5.3 zeigt die G-Banden, Abbildung 5.3 die R-Banden der Territorien 13, 14, 33 und 34.

- **Kriterium V.** Es wurden die gemeinsame Kontakte der syntenischen<sup>1</sup> Genregionen 7q21.3 und 7p12-p15 untersucht. Beide Genregionen enthalten sowohl G-Banden als auch R-Banden, die Genregion 7q21.3 umfasst 7 Chromatindomänen und enthält das Gen COL1A2 (Kollagen Typ I,  $\alpha 2$ ), die Genregion 7p12-p15 umfasst 34 Chromatindomänen und enthält das

<sup>1</sup>auf demselben Chromosom liegend

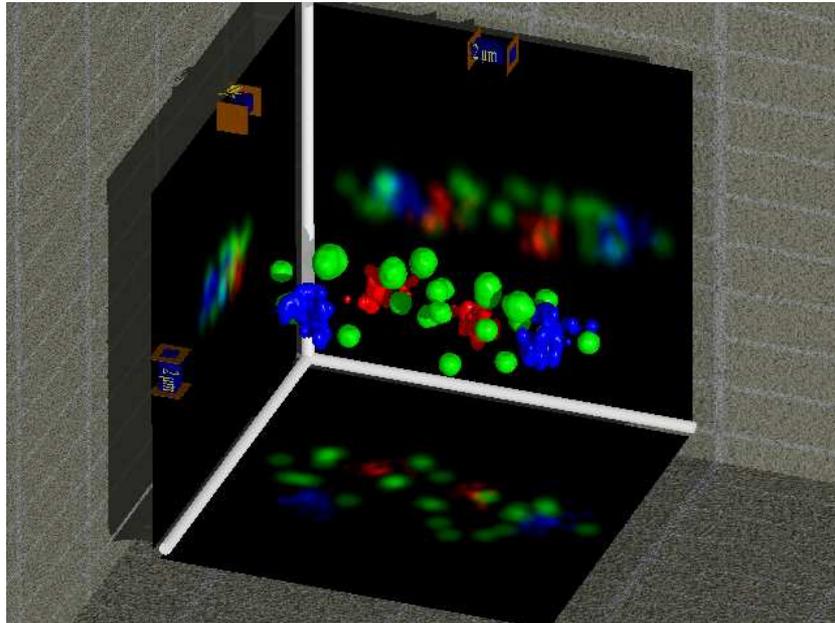


Abbildung 5.5: Visualisierung der G-Banden der Territorien 13 und 14 (Chromosom 7) in blau, 33 und 34 (Chromosom 17) in rot und der SC35-Domänen in grün mit dem Verfahren der virtuellen Mikroskopie

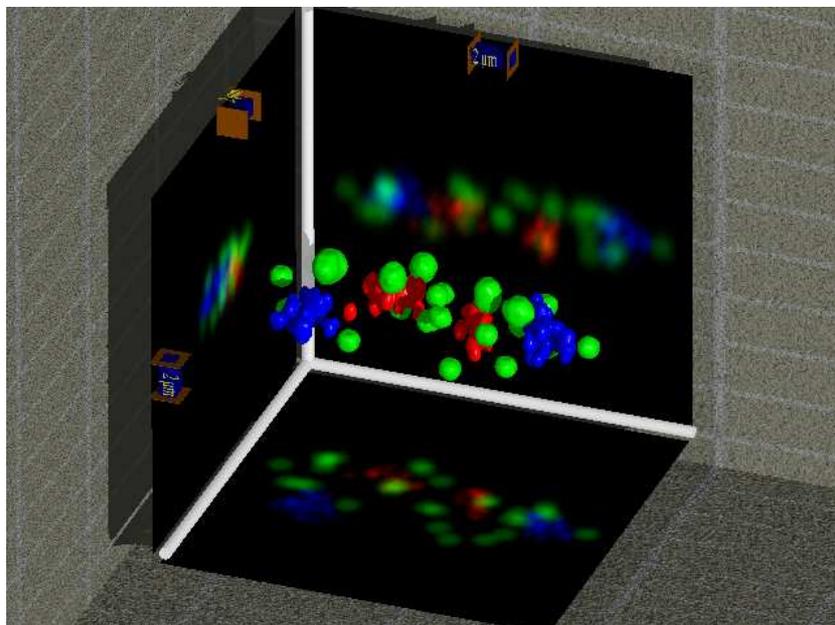


Abbildung 5.6: Visualisierung der R-Banden der Territorien 13 und 14 (Chromosom 7) in blau, 33 und 34 (Chromosom 17) in rot und der SC35-Domänen in grün mit dem Verfahren der virtuellen Mikroskopie

Gen ACTB ( $\beta$ -Actin). Ein gemeinsamer Kontakt liegt dann vor, wenn mindestens ein speckle zu beiden Genregionen gleichzeitig Kontakt hat. Dies wird auch Koassoziation genannt. Da auch hier wieder beide Genregionen auf jedem der beiden homologen Chromosomen 7 (Territorien 13 und 14) liegen, wird in der Auswertung als gemeinsamer Kontakt insgesamt gewertet, wenn ein gemeinsamer Kontakt innerhalb von Territorium 13 oder innerhalb von Territorium 14 vorliegt. Es wird jedoch nicht als gemeinsamer Kontakt gewertet, wenn ein speckle gleichzeitigen Kontakt hat z.B. zur Genregion 7p12-p15 auf Territorium 13 und zur Genregion 7q21.3 auf Territorium 14. Experimentelle Untersuchungen dazu wurden von L. Shopland et al. in [SJB<sup>+</sup>03] durchgeführt. In dieser Veröffentlichung wurde außerdem ermittelt, wie häufig die beiden Genregionen koassoziieren, wenn Die Domänen zufällig über das Zellkernvolumen verteilt sind. In Tabelle 5.2 sind die Ergebnisse des SCD-Modells den experimentellen Ergebnissen sowie der ermittelten Häufigkeit bei zufälliger Verteilung gegenübergestellt. Die Auswertung der Kriterien III und IV ist implementiert in der Funktion `bandContactComparison`.

Coassoziation der syntenischen Genregionen 7q21 und 7p12	
Vorhersage nach Zufallsverteilung	$11 \pm 3,5 \%$
Experiment	$28.7 \pm 8.2 \%$
berechnet mit SCD-Modell	$45.1 \pm 7.86 \%$

Tabelle 5.3: Auswertung des Kriteriums V

Auch hier wurde der Fehler des mit dem SCD-Modell ermittelten Wertes gemäß Gleichung 5.11 berechnet. Der nach dem SCD-Modell berechnete Wert stimmt zwar innerhalb der Fehlergrenzen nicht mit dem experimentellen Wert überein, gibt jedoch qualitativ richtig wieder, daß wesentlich mehr simultane Kontakte zwischen den syntenischen Genregionen 7q21 und 7p21 bestehen, als es rein zufällig der Fall wäre. Auch hier könnte die Ursache für die Abweichung wieder die oben beschriebene Rundungsoperation des SCD-Modells sein.

## 5.4 Geometrische Merkmale und Größen

Um den Einfluß der speckles und ihrer Wechselwirkungen auf bestimmte geometrische Merkmale der Chromatinanordnung im Zellkern zu untersuchen, wurden 204 weitere Fibroblasten-Zellkernkonfigurationen mit 400000 Monte-Carlo Schritten relaxiert, diesmal jedoch ohne speckles. Die Gyrationradien von bestimmten Genregionen sowie die Verteilung von G- und R-Banden auf Oberfläche und Inneres der Chromosomenterritorien wurden dann sowohl für die Kerne mit, als auch für die Kerne ohne speckles ausgewertet und einander gegenübergestellt.

### 5.4.1 Gyrationradien spezieller Genregionen

Die Gyrationradien der Genregionen 17q21 und 7p21, sowohl in Kernen mit, als auch ohne speckles, sind in Tabelle 5.4 aufgeführt.

Genregion	Gyrationradius in nm	
	mit speckles	ohne speckles
7p21 (G-Bande)	$44.13 \pm 13.66$	$46.58 \pm 14.93$
17q21 (R-Bande)	$50.05 \pm 15.17$	$50.88 \pm 15.53$

Tabelle 5.4: Gyrationradien der Genregionen 17q21 und 7p21

Innerhalb der Fehlergrenzen sind hier keine Unterschiede erkennbar, weder zwischen den beiden verschiedenen Genregionen, noch zwischen den beiden Fällen mit oder ohne speckles. Das liegt

wahrscheinlich daran, daß die hier betrachteten Genregionen zu klein und zu homogen sind, um signifikante Unterschiede in ihrem Gyrationradius, das heißt in der Größe des von ihnen eingenommenen Volumens, aufzuweisen. Man kann jedoch erwarten, daß eine große Genregion, welche sowohl viele G-Banden als auch viele R-Banden an verschiedenen Stellen enthält, in einem Kern mit speckles eine größere Spannung erfährt und daher auch einen größeren Gyrationradius hat als in einem Kern ohne speckles. Umgekehrt sollte sich eine große G-Banden-Region verhalten, da sie durch die Wechselwirkungen der verschiedenen Domänen im Zellkern in das Innere der Territorien, zu denen sie gehören, getrieben werden und ihnen dort nur ein kleineres Volumen zur Verfügung steht.

#### 5.4.2 Oberfläche und Inneres der Territorien

Anschließend wurde mit dem Algorithmus `surfaceFinder` (vergleiche 4.2.7) die Oberfläche und das Innere sowie die isolierten Domänen aller Chromosomenterritorien ermittelt. Dann wurde klassifiziert, ob es sich um eine G-, R- oder C-Bande handelt. Mit dieser Information wurden für jedes Territorium folgende Summen ermittelt:

- $GBandsSurfaceCounter_i$ : Anzahl der G-Banden auf der Oberfläche von Territorium  $i$ .
- $RBandsSurfaceCounter_i$ : Anzahl der R-Banden auf der Oberfläche von Territorium  $i$ .
- $CBandsSurfaceCounter_i$ : Anzahl der C-Banden auf der Oberfläche von Territorium  $i$ .
- $GBandsInteriorCounter_i$ : Anzahl der G-Banden im Inneren von Territorium  $i$ .
- $RBandsInteriorCounter_i$ : Anzahl der R-Banden im Inneren von Territorium  $i$ .
- $CBandsInteriorCounter_i$ : Anzahl der C-Banden im Inneren von Territorium  $i$ .

Damit wurden folgende Quotionen gebildet:

$$\begin{aligned} GBandsSrf2Int_i &= \frac{GBandsSurfaceCounter_i}{GBandsInteriorCounter_i} \\ RBandsSrf2Int_i &= \frac{RBandsSurfaceCounter_i}{RBandsInteriorCounter_i} \\ CBandsSrf2Int_i &= \frac{CBandsSurfaceCounter_i}{CBandsInteriorCounter_i} \end{aligned} \quad (5.12)$$

Diese Daten wurden für jedes Chromosom einzeln ausgewertet und in Abbildung 5.7 gegen die Chromosomengröße aufgetragen, Tabelle 5.6 listet die dazugehörigen Mittelwerte über die 204 Zellkernkonfigurationen sowie die Standardabweichungen auf.

Dann wurden die Informationen der verschiedenen Territorien addiert, um die Verteilung im gesamten Zellkern angeben zu können:

$$\begin{aligned} GBandsSurfaceTotalCounter &= \sum_{i=1}^{46} GBandsSurfaceCounter_i \\ RBandsInteriorTotalCounter &= \sum_{i=1}^{46} RBandsInteriorCounter_i \end{aligned} \quad (5.13)$$

Die oben für jedes Territorium einzeln ermittelten Quotienten wurden nun für den ganzen Zellkern ermittelt:

$$\begin{aligned} GBandsSrf2Int_{tot} &= \frac{GBandsSurfaceTotalCounter}{GBandsInteriorTotalCounter} \\ RBandsSrf2Int_{tot} &= \frac{RBandsSurfaceTotalCounter}{RBandsInteriorTotalCounter} \end{aligned} \quad (5.14)$$

Die Werte dieser Quotienten sind in Tabelle 5.5 aufgeführt.

Aus der Aufteilung eines Territoriums in Oberfläche und Inneres lassen sich weitere Informationen gewinnen, wie z.B. den Quotienten aus Anzahl der G-Banden auf der Oberfläche zu Anzahl der R-Banden auf der Oberfläche:

$$G2RSrf_i = \frac{GBandsSurfaceCounter_i}{RBandsSurfaceCounter_i} \quad (5.15)$$

und ebenso für das Innere:

$$G2RInt_i = \frac{GBandsInteriorCounter_i}{RBandsInteriorCounter_i} \quad (5.16)$$

Diese Daten wurden wieder für jedes Chromosom einzeln ausgewertet und in Abbildung gegen den R-Bandengehalt aufgetragen, Tabelle 5.7 listet die dazugehörigen Mittelwerte über die 204 Zellkernkonfigurationen sowie die Standardabweichungen auf. Dieselben Quotienten wurden nochmal mit den Gesamtzählraten des ganzen Zellkerns gebildet (siehe Tabelle 5.5):

$$\begin{aligned} G2RSrf_{tot} &= \frac{GBandsSurfaceTotalCounter}{RBandsSurfaceTotalCounter} \\ G2RInt_{tot} &= \frac{GBandsInteriorTotalCounter}{RBandsInteriorTotalCounter} \end{aligned} \quad (5.17)$$

Es werden die oben beschriebenen Quotienten ausgewertet und nicht die absoluten Zählraten. Das hat den Vorteil, daß sich verschiedene Einflüsse, welche einen Vergleich erschweren, bei diesen Quotienten wieder herausdividieren. So gibt es in einem menschlichen Zellkern 3123 R-Bandendomänen, aber nur 2419 G-Bandendomänen. Insgesamt gibt es also 1.297 mal so viele R-Bandendomänen wie G-Bandendomänen bzw. es gilt im Mittel über einen ganzen Zellkern

$$G2R = \frac{GBandsTotalCounter}{RBandsTotalCounter} = 0.770733.$$

In den Quotienten  $GBandsSrf2Int$  und  $RBandsSrf2Int$  werden aber nur G-Banden mit G-Banden und R-Banden mit R-Banden verglichen. Das obige Ungleichgewicht kann hier also ignoriert werden. Beim Vergleich der Quotienten  $G2R$  muß natürlich darauf geachtet werden, daß es mehr R-Bandendomänen als G-Bandendomänen gibt.

Möchte man die Verteilung der verschiedenen Bandentypen auf verschiedenen Chromosomen vergleichen, so ist auch hier der Vergleich der Quotienten von Vorteil, da sich die Chromosomen stark in ihrer Größe unterscheiden und deshalb absolute Zählraten in großen Chromosomen grundsätzlich größer ausfallen, als in kleinen Chromosomen.

Parameter		mit Speckles	ohne Speckles
Domänen Srf2Int	G-Banden	4.9767 ± 0.3286	5.3638 ± 0.3849
	R-Banden	5.5987 ± 0.3303	5.4672 ± 0.3183
Kontakte Srf2Int	G-Banden	8.4107 ± 3.6527	-
	R-Banden	8.5464 ± 1.4130	-
G2R	Srf	0.7563 ± 0.0092	0.7683 ± 0.0091
	Int	0.8529 ± 0.0544	0.7853 ± 0.0500

Tabelle 5.5: Auswertung der Verteilung von R- und G-Banden auf Oberfläche und Inneres der Territorien, die hier dargestellten Werte sind Quotienten aus den Zählraten für den gesamten Zellkern.

An der Verteilung von G- und R-Banden auf die Oberfläche und das Innere der Territorien (Zeilen 1 und 2 in Tabelle 5.5) fällt auf: In allen hier behandelten Fällen befinden sich auf der Oberfläche der Territorien im Zellkern ungefähr 5 mal so viele Chromatindomänen wie im Inneren der Territorien. Drei der Quotienten,  $GBandsSrf2Int_{tot}$  ohne speckles,  $RBandsSrf2Int_{tot}$  ohne speckles und  $RBandsSrf2Int_{tot}$  mit speckles stimmen innerhalb ihrer Fehlergrenzen überein. Für

die beiden erstgenannten ist das nicht verwunderlich, da in einem Zellkern ohne speckles G- und R-Banden denselben Kräften unterliegen, es gibt keine Präferenz des einen Bandentyps gegenüber dem anderen. Die Anwesenheit von speckles im Zellkern führt jedoch zu unterschiedlichen Verteilungen für G- und für R-Banden. Der Quotient  $GBandsSrf2Int_{tot}$  nimmt gegenüber dem Fall ohne speckles ab, was bedeutet, daß nun weniger G-Banden auf der Oberfläche und mehr im Inneren der Territorien lokalisiert sind. Der Quotient  $RBandsSrf2Int_{tot}$  nimmt im Gegensatz dazu leicht zu, was bedeutet, daß nun mehr R-Banden auf der Oberfläche und weniger im Inneren der Territorien lokalisiert sind - dieser Unterschied ist jedoch wesentlich kleiner als die Schwankungsbreite der verglichenen Werte  $GBandsSrf2Int_{tot}$ . Innerhalb der Fehlergrenzen sind die beiden Werte  $RBandsSrf2Int_{tot}$  also gleich. Der Effekt ist bei den G-Banden also stärker. Eine Ursache dafür könnte sein, daß die speckles mit den G-Bandendomänen ausschließlich repulsiv wechselwirken, mit den R-Bandendomänen jedoch sowohl attraktiv, als auch repulsiv. Die biologische Bedeutung dieser Beobachtung ist, daß an der Organisation des Chromatins im Zellkern zu Euchromatin und Heterochromatin die inaktiven Chromatinabschnitte stärker beteiligt sind, als das aktive Chromatin.

Ähnliche Information steckt in den Zeilen 5 und 6 von Tabelle 5.5. Während ohne speckles das Verhältnis von G-Bandendomänen zu R-Bandendomänen auf der Oberfläche einerseits und im Inneren andererseits ähnlich groß ist (Übereinstimmung innerhalb der Fehlergrenzen des zweiten Wertes), führt die Anwesenheit von speckles dazu, daß der Quotient  $G2RSrf_{tot}$  kleiner und der Quotient  $G2RInt_{tot}$  größer wird (G-Banden werden von der Oberfläche ins Innere verschoben, R-Banden vom Inneren zur Oberfläche), es liegt keine Übereinstimmung innerhalb der Fehlergrenzen mehr vor.

Kontakte zwischen Chromatindomänen und speckles treten auf der Oberfläche ca. 8.5 mal so häufig auf wie im Inneren. In diesem Quotienten unterscheiden sich G- und R-Banden nicht, wobei beachtet werden muß, daß die Gesamtzahl der R-Bandenkontakte ja das Achtfache der Gesamtzahl der G-Bandenkontakte ist. Es ist bemerkenswert, daß es überhaupt Kontakte zwischen speckles und Chromatindomänen im Inneren von Territorien gibt. Das bedeutet, daß die speckles die Oberfläche stellenweise durchdringen und Domänen in tieferen Schichten kontaktieren. Dies stützt die im ICD-Modell erhobene These, daß der ICD-Raum sich teilweise bis ins Innere der Territorien erstreckt, daß sich also ein kleiner Teil der aktiven Genregionen im Inneren der Territorien befinden kann. Die Verteilung der Kontakte kann selbstverständlich in Kernen ohne speckles nicht ermittelt werden, da es hier gar keine Kontakte zwischen Chromatindomänen und speckles gibt.

### 5.4.3 Auswertung von $GBandsSrf2Int$ und $RBandsSrf2Int$ für einzelne Chromosomen

Bei der Auftragung der Parameter  $Srf2Int$  gegen die Chromosomengröße in Abbildung 5.7 beobachtet man, daß alle Kurven im Bereich oberhalb von 80 Mbp monoton fallend sind. Dies ist dadurch zu erklären, daß grundsätzlich bei Körpern im Raum das Verhältnis Oberfläche zu Volumen mit steigender Größe abnimmt. Am Beispiel einer Kugel kann man ausrechnen:

$$\frac{Surface}{Volume} = \frac{4\pi r^2}{\frac{4}{3}\pi r^3} = \frac{3}{r} = \frac{3}{\left(\frac{3}{4\pi}V\right)^{\frac{1}{3}}} = (36\pi)^{\frac{1}{3}} \cdot V^{-\frac{1}{3}} = (36\pi\rho)^{\frac{1}{3}} \cdot N_{Dom}^{-\frac{1}{3}}$$

Dabei ist  $V = \frac{4}{3}\pi r^3$  das Volumen einer Kugel,  $N_{Dom}$  die Anzahl der Domänen im Volumen  $V$ , also die Chromosomengröße und  $\rho = \frac{N_{Dom}}{V}$  die Domänendichte. Die Quotienten  $Srf2Int$  skalieren also wie  $N_{Dom}^{-\frac{1}{3}}$ . Weiterhin ist zu beachten, daß die hier dargestellten Werte eine große Schwankungsbreite aufweisen (siehe Tabelle 5.6), die Fehlerbalken sind in Abbildung 5.7 zugunsten einer besseren Übersichtlichkeit nicht eingetragen.

Die Quotienten  $GBandsSrf2Int_{tot}$  und  $RBandsSrf2Int_{tot}$ , welche Informationen über den ganzen Kern enthalten und in Tabelle 5.5 aufgelistet sind, jeweils einmal für Kerne mit speckles und einmal für Kerne ohne speckles, sind als gerade Linien in Abbildung 5.7 eingefügt.

In einem Kern ohne speckles sind die Quotienten  $Srf2Int$  für G-Banden und  $Srf2Int$  für R-Banden auf fast allen Chromosomen sehr ähnlich, einzige Ausnahmen sind die Chromosomen Y

(50 Mbp) und 22 (49 Mbp). Die Kurven für G-Banden ohne speckles (schwarz) und für R-Banden ohne speckles (rot) liegen daher abgesehen von diesen beiden Werten sehr dicht bei einander.

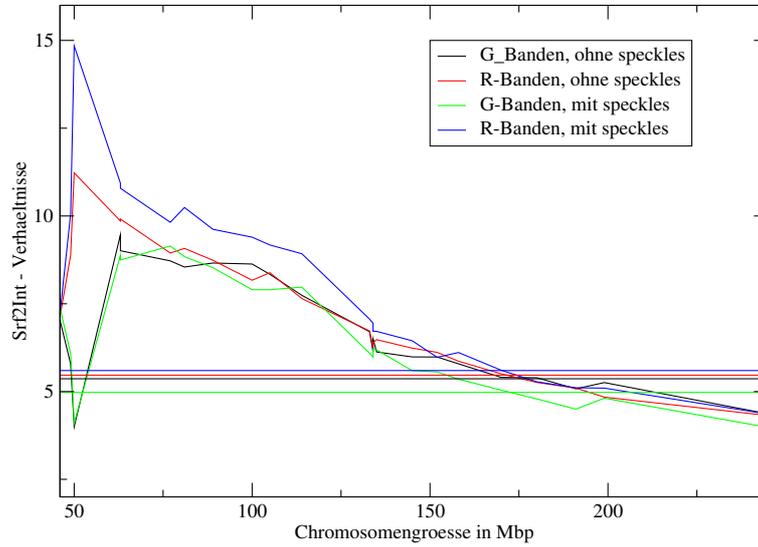


Abbildung 5.7: Verteilung von R- und G-Banden auf Oberfläche und Inneres der Territorien; der Parameter  $Srf2Int$  (siehe Text) ist gegen die Chromosomengröße aufgetragen, vergleiche Tabelle 5.6. Die geraden Linien stellen die Werte dar, die für den ganzen Zellkern erhalten werden, vergleiche Tabelle 5.5.

In einem Kern mit speckles liegen die Werte von  $Srf2Int$  für G-Banden (grün) und für R-Banden (blau) bei allen Chromosomen weiter aus einander, jedoch bei verschiedenen Chromosomen verschieden stark.

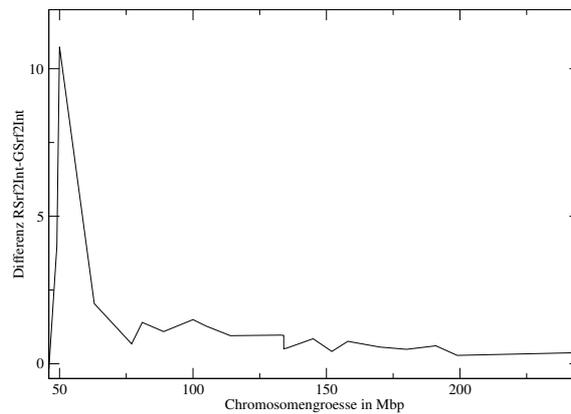


Abbildung 5.8: Differenz  $RSrf2Int-GSrf2Int$

Der Einfluß der speckles macht sich bemerkbar, indem G-Bandendomänen ins Innere der Territorien gedrängt werden und R-Bandendomänen bevorzugt auf der Oberfläche verbleiben. Dabei

ist darauf hinzuweisen, daß in großen Chromosomen bei der Anwesenheit von speckles vor allem eine Umordnung der G-Banden stattfindet (die grüne Kurve ist gegenüber der schwarzen Kurve ab einer Chromosomengröße von ungefähr 120 Mbp nach unten verschoben), während die Anwesenheit der speckles an den Quotienten für R-Banden wenig ändert. Dieses Verhalten weisen auch die Werte für den gesamten Zellkern auf. Die kleineren Chromosomen zeigen ein davon abweichendes Verhalten: hier sind es die R-Bandendomänen, die durch die Anwesenheit der speckles eine starke Umordnung erfahren (unterhalb von 130 Mbp ist die blaue Kurve gegenüber der roten nach oben verschoben), während sich an den Quotienten für G-Bandendomänen wenig ändert.

Die hier durchgeführte Untersuchung kann daher für zukünftige Experimente eine Orientierung liefern, auf welchen Chromosomen die Verteilung von aktiven und inaktiven Chromatinabschnitten am ehesten zu beobachten ist: das sind diejenigen, bei denen die Differenz zwischen der Wert der blauen Kurve und dem Wert der grünen Kurve besonders groß ist.

In Abbildung 5.8 ist diese Differenz gegen die Chromosomengröße aufgetragen. Kleine Chromosomen, d.h. Chromosomen mit weniger als 125 Mbp, die eine besonders hohe Differenz aufweisen, sind: Y, 22, 20, 15 und 17, sortiert nach der Größe der Differenz. Große Chromosomen, d.h. solche mit mehr als 125 Mbp, mit einer großen Differenz sind: 12, 8, 7 und 4.

Chromosom	Größe in Mbp	ohne speckles		mit speckles	
		G-Banden	R-Banden	G-Banden	R-Banden
1	245	4.5371 ± 1.7710	4.3896 ± 1.0615	4.152 ± 2.288	4.328 ± 1.219
2	243	4.4034 ± 1.4044	4.3409 ± 1.2084	4.016 ± 1.305	4.389 ± 1.345
3	199	5.2555 ± 2.0927	4.8429 ± 1.5122	4.813 ± 2.038	5.097 ± 1.980
4	191	5.0777 ± 1.9864	5.0976 ± 1.9146	4.499 ± 1.635	5.108 ± 1.779
5	180	5.3938 ± 2.3395	5.2602 ± 2.3893	4.789 ± 1.805	5.279 ± 1.767
6	170	5.4040 ± 2.2061	5.5057 ± 2.1983	5.042 ± 2.537	5.609 ± 2.360
7	158	5.7892 ± 2.9614	5.8631 ± 2.5270	5.352 ± 2.307	6.112 ± 2.961
X	152	5.9844 ± 3.1718	6.1164 ± 2.6640	5.558 ± 2.764	5.976 ± 2.426
8	145	5.9858 ± 3.0461	6.2368 ± 3.1885	5.597 ± 2.246	6.445 ± 2.971
10	135	6.1272 ± 3.1350	6.4773 ± 3.1367	6.188 ± 4.919	6.714 ± 4.422
9	134	6.4988 ± 4.9301	6.3915 ± 2.9181	6.211 ± 4.385	6.720 ± 3.781
11	134	6.1348 ± 3.3912	6.2771 ± 2.9806	5.983 ± 3.820	6.942 ± 3.415
12	133	6.7044 ± 4.6096	6.7266 ± 3.6942	6.092 ± 3.403	7.064 ± 4.114
13	114	7.7398 ± 5.7275	7.6470 ± 5.7053	7.973 ± 6.711	8.925 ± 7.342
14	105	8.3386 ± 6.5933	8.3840 ± 6.8505	7.900 ± 6.526	9.174 ± 7.425
15	100	8.6311 ± 6.6798	8.1705 ± 6.2276	7.901 ± 6.474	9.397 ± 7.527
16	89	8.6594 ± 6.5748	8.7384 ± 7.2954	8.529 ± 6.664	9.620 ± 8.026
17	81	8.5452 ± 5.9731	9.0763 ± 7.1493	8.840 ± 6.412	10.241 ± 8.141
18	77	8.7227 ± 6.5491	8.9457 ± 7.5233	9.146 ± 7.359	9.820 ± 7.448
19	63	9.0094 ± 5.8627	9.9121 ± 7.4353	8.747 ± 5.725	10.789 ± 8.000
20	63	9.4605 ± 6.5453	9.8585 ± 7.5766	8.881 ± 6.444	10.936 ± 8.576
Y	50	3.9987 ± 1.9401	11.2290 ± 9.2334	4.101 ± 1.803	14.844 ± 11.432
22	49	5.7848 ± 3.0356	8.8885 ± 5.5306	6.080 ± 3.108	10.064 ± 5.815
21	46	7.0404 ± 3.7221	7.1228 ± 3.5890	7.395 ± 3.707	7.209 ± 3.523

Tabelle 5.6: Wertetabelle zu Abbildung 5.7

#### 5.4.4 Auswertung von $G2R\text{Ratio}$ für einzelne Chromosomen

In Abbildung 5.9 sind die Quotienten aus Anzahl an G-Banden durch Anzahl an R-Banden,  $G2R\text{Ratio}$ , gegen den R-Bandengehalt der Chromosomen aufgetragen.

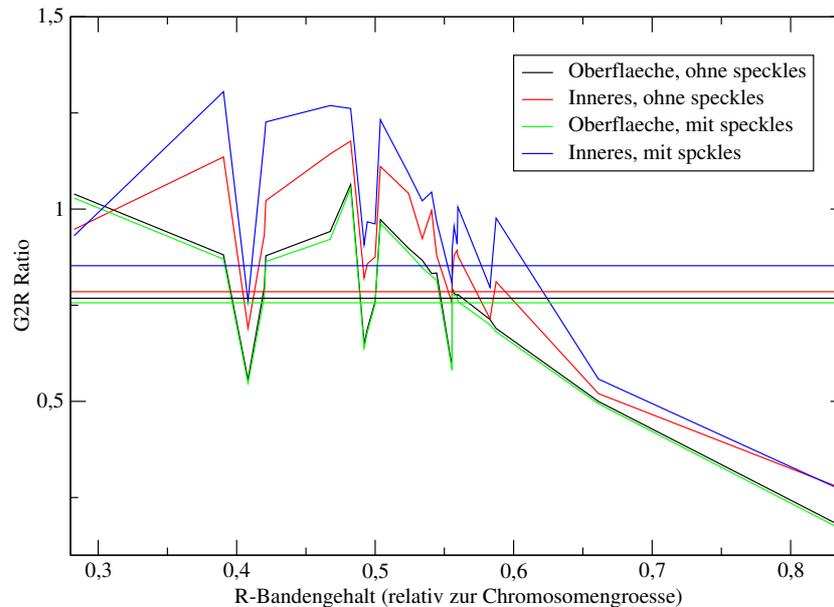


Abbildung 5.9: Verhältnis von G-Banden zu R-Banden,  $G2R\text{Ratio}$ , auf Oberfläche und Innerem der Territorien, aufgetragen gegen den relativen R-Bandengehalt der Chromosomen, vergleiche Tabelle 5.7. Die geraden Linien stellen die Werte dar, die für den ganzen Zellkern erhalten werden, vergleiche Tabelle 5.5.

Es wird unterschieden zwischen Oberfläche und Innerem der Territorien, jeweils in Kernen mit und ohne speckles. Daher werden für jedes Territorium vier verschiedene Quotienten  $G2R\text{Ratio}$  ermittelt, es gibt vier verschiedene Kurven. In dieser Abbildung wurden wie in Abbildung 5.9 die Fehlerbalken zugunsten einer besseren Übersicht weggelassen; die Schwankungsbreiten der einzelnen Werte sind jedoch in Tabelle 5.7 angegeben. Die relative Schwankungsbreite ist im Inneren wesentlich höher als auf der Oberfläche, dies liegt daran, daß mehr Domänen auf der Oberfläche lokalisiert sind als im Inneren. Insgesamt fallen alle Kurven mit steigendem R-Bandengehalt ab. Dies ist einsichtig, bei großem R-Bandengehalt der Nenner des Quotienten  $G2R\text{Ratio}$  groß und der Zähler klein, der ganze Quotient also klein ist. Die Kurven sind jedoch nicht monoton fallend, sondern es gibt Chromosomen, bei denen alle vier Kurven lokale Minima aufweisen: 22, 19 und 17.

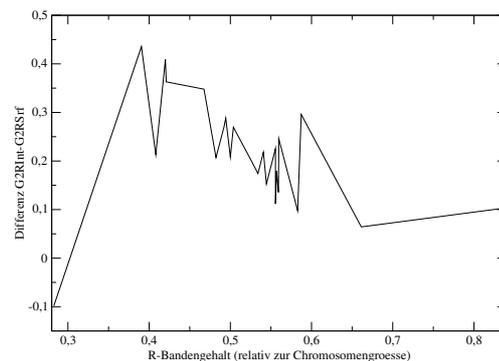


Abbildung 5.10: Differenz aus den Quotienten  $G2R\text{Int}$  und  $G2R\text{Srf}$ , Erläuterungen siehe Text.

Chromosom	R-Banden-Anteil	ohne speckles		mit speckles	
		Oberfläche	Inneres	Oberfläche	Inneres
Y	0.840	0.1691 ± 0.0363	0.2694 ± 0.4509	0.161 ± 0.034	0.264 ± 0.400
1	0.661	0.4999 ± 0.0353	0.5202 ± 0.1430	0.494 ± 0.038	0.558 ± 0.173
20	0.587	0.6897 ± 0.0932	0.8112 ± 0.6139	0.681 ± 0.089	0.977 ± 0.960
3	0.583	0.7129 ± 0.0535	0.7150 ± 0.2597	0.699 ± 0.057	0.797 ± 0.295
11	0.560	0.7777 ± 0.0682	0.8783 ± 0.4220	0.760 ± 0.072	1.006 ± 0.517
X	0.559	0.7757 ± 0.0662	0.8928 ± 0.4323	0.774 ± 0.064	0.909 ± 0.341
7	0.557	0.7809 ± 0.0663	0.8804 ± 0.3754	0.776 ± 0.063	0.955 ± 0.455
2	0.556	0.7948 ± 0.0546	0.8195 ± 0.2299	0.781 ± 0.057	0.893 ± 0.277
17	0.556	0.5883 ± 0.0791	0.7548 ± 0.6803	0.581 ± 0.071	0.806 ± 0.694
5	0.544	0.8337 ± 0.0666	0.8789 ± 0.4087	0.815 ± 0.069	0.968 ± 0.356
10	0.541	0.8327 ± 0.0706	0.9963 ± 0.5166	0.826 ± 0.076	1.044 ± 0.573
4	0.534	0.8670 ± 0.0642	0.9233 ± 0.3375	0.847 ± 0.069	1.021 ± 0.365
8	0.524	0.8978 ± 0.0804	1.0407 ± 0.5589	0.887 ± 0.079	1.093 ± 0.485
12	0.504	0.9726 ± 0.0834	1.1108 ± 0.5344	0.963 ± 0.084	1.232 ± 0.644
9	0.500	0.7599 ± 0.0675	0.8756 ± 0.4156	0.753 ± 0.072	0.961 ± 0.548
16	0.494	0.6880 ± 0.0850	0.8577 ± 0.7459	0.679 ± 0.080	0.967 ± 0.787
19	0.492	0.6505 ± 0.0969	0.8210 ± 0.8113	0.638 ± 0.101	0.908 ± 0.980
6	0.482	1.0639 ± 0.0890	1.1767 ± 0.5258	1.054 ± 0.086	1.262 ± 0.528
18	0.468	0.9418 ± 0.1234	1.1434 ± 0.9794	0.921 ± 0.112	1.269 ± 1.047
13	0.421	0.8786 ± 0.1481	1.0219 ± 0.4527	0.864 ± 0.148	1.227 ± 0.522
15	0.420	0.7980 ± 0.0882	0.9339 ± 0.7646	0.773 ± 0.088	1.182 ± 1.022
22	0.408	0.5569 ± 0.1117	0.6899 ± 0.8391	0.547 ± 0.105	0.760 ± 0.804
14	0.390	0.8809 ± 0.0940	1.1356 ± 1.0416	0.870 ± 0.108	1.305 ± 1.184
21	0.283	1.0392 ± 0.2130	0.9471 ± 0.9971	1.029 ± 0.200	0.931 ± 0.959

Tabelle 5.7: Wertetabelle zu Abbildung 5.9

Ähnlich wie im letzten Abschnitt liegen bei dieser Untersuchung die Kurven für Oberfläche und Inneres in einem Kern mit speckles weiter auseinander als in einem Kern ohne speckles. Ursache ist wieder, daß die speckles durch ihre Wechselwirkungen mit den Chromatindomänen die G-Banden ins Innere der Territorien drängen, während sich die R-Banden auf der Oberfläche anreichern. Wie im vorhergehenden Abschnitt wurde die Differenz aus den zwei Kurven für Kerne mit speckles ermittelt, jedoch diesmal gegen den R-Bandengehalt aufgetragen (Abbildung 5.10). Nach diesem Kriterium wäre also ein experimenteller Nachweis der unterschiedlichen Verteilung von G- und R-Banden auf den Chromosomen 13, 14, 15, 18 und 20 am ehesten möglich. Im Unterschied zu den Ergebnissen des vorigen Abschnitts gibt es hier kein unterschiedliches Verhalten für kleine und große Chromosomen.

In der Schnittmenge der in diesem und im vorigen Absatz ausgezeichneten Chromosomen befinden sich die Chromosomen 15 und 20.

# Kapitel 6

## Diskussion und Ausblick

### 6.1 Diskussion

#### 6.1.1 Weiterentwicklung des SCD-Modells

In diesem Abschnitt soll zunächst auf die technischen Veränderungen in den Programmen zum SCD-Modell eingegangen werden. In den folgenden Abschnitten werden die Ergebnisse der Auswertung diskutiert.

Es wurden Proteinkomplexe, sogenannte SC35-Domänen, speckles oder nuclear bodies in das SCD-Modell integriert. Ihre Anzahl und Größe sind variabel gehalten. Die Wechselwirkung zwischen den frühreplizierenden R-Banden und den speckles wurde durch ein Lennard-Jones-Potential, die Wechselwirkungen zwischen den spätreplizierenden G-Banden und den speckles sowie zwischen zwei speckles unter einander durch den repulsiven Teil eines Lennard-Jones-Potentials modelliert. Die übrigen Wechselwirkungen, denen eine Chromatindomäne in einer SCD-Zellkernkonfiguration unterworfen ist (Chromatinlinker, Excluded Volume, Tensegrity), wurden nicht verändert. Zum ersten Mal wurde ein funktioneller Zusammenhang zwischen den oben genannten Proteinkomplexen und Chromatindomänen im SCD-Modell realisiert. Damit kann der Einfluß von SC35-Domänen oder speckles auf die Kernstruktur quantitativ untersucht werden.

Das Programm zur Generierung von Startkonfigurationen für das SCD-Modell wurde erweitert. Es wurde eine größenabhängige Verteilung der Chromosomenterritorien, wie sie bei Fibroblasten vorliegt, implementiert. Daher können nun Startkonfigurationen für Fibroblasten generiert werden. In diese können speckles eingefügt werden. Dabei sind verschiedene Anordnungen möglich, die verschiedene biologische Bedeutungen haben.

Die Funktion `updatePosition` wurde neu geschrieben. Sie wird immer dann aufgerufen, wenn bei der Monte Carlo Relaxation einer Zellkernkonfiguration im SCD-Modell nach einem Monte-Carlo-Schritt die Verrückung einer Domäne akzeptiert wurde und die neue Position der Domäne in die Zellkernkonfiguration geschrieben werden muß. Es wurde mit der neuen Funktion eine Reduktion der Laufzeitordnung von quadratisch auf linear erreicht und die Rechenzeit auf ein Elftel der ursprünglichen Rechenzeit verkürzt. Dies machte Parametervariationen sowie numerische und quantitative Untersuchungen im einem wesentlich größeren Umfang als zuvor möglich.

Es wurden Algorithmen implementiert, welche Cluster von Chromatindomänen oder Cluster von speckles finden. Darüber hinaus kann mit zusätzlichen Funktionen die Oberfläche der Chromosomenterritorien ermittelt werden.

### 6.1.2 Variation der Stärke der Wechselwirkung zwischen speckles und R-Banden $\epsilon_1$

Durch die Integration der SC35-Domänen oder speckles in das SCD-Modell mussten vier neue Parameter mit sinnvollen Werten belegt werden:  $\epsilon_1$ , die Stärke der Wechselwirkung zwischen speckles und R-Banden,  $\epsilon_2$ , die Stärke der Wechselwirkung zwischen zwei speckles,  $\epsilon_3$ , die Stärke der Wechselwirkung zwischen speckles und G-Banden sowie  $r_{speckle}$ , der Radius der speckles. Für die drei letztgenannten Parameter wurden die Festlegungen getroffen:  $r_{speckle} = 3 \cdot r_{Chr}$  bei 30 speckles in einer Fibroblasten-Zellkernkonfiguration ( $r_{speckle}$  ist der Radius der speckles und  $r_{Chr}$  der Radius der Chromatindomänen) sowie  $\epsilon_2 = \epsilon_3 = 1.5k_B T$ , damit sind die mit diesen Parametern assoziierten Wechselwirkungen genauso stark wie die Wechselwirkung zwischen zwei Chromatindomänen, haben jedoch nicht dieselbe Form; zwei Chromatindomänen wechselwirken über das MLS-Potential, während es sich bei den mit  $\epsilon_2$  und  $\epsilon_3$  assoziierten Wechselwirkungen um die repulsiven Anteile von Lennard-Jones-Potentialen handelt. Der Parameter  $\epsilon_1$  wurde variiert und folgende Zahlenwerte mit experimentellen Daten und biologischen Modellvorstellungen von L. Shopland et al. aus [SJB<sup>+</sup>03] verglichen.

- **Kriterium I:** Der Quotient aus allen Kontakten zwischen G-Banden und speckles im Zellkern durch alle Kontakte zwischen R-Banden mit speckles im Zellkern, **G2RContacts**. Nach diesem Kriterium ergibt sich bei der Variation für  $\epsilon_1$  der Zahlenwert  $\epsilon_{1,opt}$ :

$$\epsilon_{1,opt} = 1.90961^{+0.4248}_{-0.4616}$$

- **Kriterium II:** Die mittlere Anzahl an kontaktierten Speckles pro Territorium, **contPerTer**. Nach diesem Kriterium ergibt sich bei der Variation für  $\epsilon_1$  ein erlaubter Bereich:

$$\epsilon_{1,min} = 1.35377 \pm 0.03799$$

$$\epsilon_{1,max} = 4.06977 \pm 0.09477$$

Es gilt also:  $\epsilon_{1,min} < \epsilon_{1,opt} < \epsilon_{1,max}$ . Der Wert  $\epsilon_{1,opt}$  liegt folglich im erlaubten Bereich (auch innerhalb der Fehlergrenzen), die Kriterien I und II sind mit einander vereinbar.

Dieselbe Variation wie oben beschrieben wurde noch einmal durchgeführt, nur wurden die speckles als etwas kleiner angenommen:  $r_{speckle} = 2 \cdot r_{Chr}$  bei ebenfalls 30 speckles in einer Fibroblasten-Zellkernkonfiguration. Es galt ebenfalls  $\epsilon_2 = \epsilon_3 = 1.5k_B T$ . Mit den oben beschriebenen Kriterien I und II ergibt sich:

$$\epsilon_{1,opt} = 1.56139$$

$$\epsilon_{1,min} = 2.03069 \pm 0.04801$$

Es gilt also hier:  $\epsilon_{1,opt} < \epsilon_{1,min}$ . Im Falle der kleineren speckles liegt der Wert  $\epsilon_{1,opt}$  folglich nicht im laut Kriterium II erlaubten Bereich, die Kriterien I und II sind nicht mit einander vereinbar. Damit liefert diese Rechnung eine untere Grenze für die mittlere Größe der speckles sowie für das Gesamtvolumen aller speckles im Zellkern.

### 6.1.3 Kontaktverhalten von früh- und spät replizierenden Chromatindomänen

Mit den in den vorigen Abschnitten ermittelten Werten  $\epsilon_{1,opt} = 1.90961^{+0.4248}_{-0.4616}$  und  $r_{speckle} = 3 \cdot r_{Chr}$  wurden 204 Fibroblasten-Zellkernkonfigurationen relaxiert. Folgende Informationen wurden ausgewertet und mit experimentellen Daten von L. Shopland et al. (vergleiche [SJB<sup>+</sup>03]) verglichen:

- Der Quotient aus allen Kontakten zwischen G-Banden und speckles im Zellkern durch alle Kontakte zwischen R-Banden mit speckles im Zellkern, **G2RContacts**, vergleiche Kriterium I oben. Wegen der in den vorigen Abschnitten durchgeführten Parametervariation gibt es hier eine gute Übereinstimmung mit dem Experiment.

- Die mittlere Anzahl an kontaktierten Speckles pro Territorium, **contPerTer**, vergleiche Kriterium II oben. Hier gibt es ebenfalls aufgrund der Parametervariation eine gute Übereinstimmung mit dem Experiment.
- Kontaktverhalten der Genregionen 17q21 (mehrheitlich R-Banden, umfasst 10 Chromatindomänen, Kriterium III) und 7p21 (mehrheitlich G-Banden, umfasst 9 Chromatindomänen, Kriterium IV). Es wurde gezählt, wie oft Domänen dieser Genregionen mit keinem, mit einem oder mit mehr als einem speckle Kontakt hatten. Von den sechs ermittelten Zählraten (je drei Werte für zwei Genregionen) stimmen zwei innerhalb der Fehlergrenzen mit den experimentellen Daten überein, zwei geben die experimentellen Ergebnisse qualitativ richtig wieder und zwei sind nicht mit dem Experiment in Einklang zu bringen.
- Die Coassoziaton der syntenischen<sup>1</sup> Genregionen 7q21.3 und 7p12-p15. Coassoziatonen zwischen zwei Genregionen A und B liegt dann vor, wenn mindesten ein speckle zu mindestens einer Domäne der Genregion A und mindestens einer Domäne der Genregion B simultan Kontakt hat. Die hier ermittelte Zählrate stimmt qualitativ mit dem Experiment überein.

Bei knapp der Hälfte der berechneten Größen ergab sich folglich Übereinstimmung mit den experimentellen Daten, bei einem weiteren Drittel gaben die berechneten Daten die experimentellen zumindest qualitativ richtig wieder. Die in dieser Arbeit am SCD-Modell vorgenommenen Erweiterungen führen zu simulierten Datensätzen, welche mit experimentellen Daten vergleichbar sind und diese zum Großteil richtig wiedergeben.

#### 6.1.4 Geometrische Eigenschaften der Chromosomenterritorien

Um den Einfluß der speckles und ihrer Wechselwirkungen auf bestimmte geometrische Merkmale der Chromatinanordnung im Zellkern zu untersuchen, wurden 204 weitere Fibroblasten-Zellkernkonfigurationen relaxiert, jedoch ohne speckles. Diese wurden mit den 204 Zellkernkonfigurationen mit speckles verglichen.

Es wurden die Gyrationenradien der ähnlich großen Genregionen 17q21 (R-Bande, s.o.) und 7p21 (G-Bande, s.o.) ausgewertet. Innerhalb der Fehlergrenzen ergab sich kein Unterschied zwischen den vier Gyrationenradien. Dies liegt wohl daran, daß die untersuchten Genregionen zu klein und zu homogen sind, um Unterschiede in dem von ihnen eingenommenen Volumen aufzuweisen, wofür der Gyrationradius ja ein Maß ist. Eine große Genregion, welche viele G- und viele R-Banden enthält, müsste jedoch durch die Wechselwirkung mit den speckles in einen Spannungszustand versetzt werden und in einem Kern mit speckles einen größeren Gyrationradius aufweisen, als in einem Kern ohne speckles.

Anschließend wurden Oberfläche und Inneres von jedem Territorium in jedem der Zellkerne mit und ohne speckles bestimmt. Mit diesen Informationen wurden mehrere Quotienten berechnet:  $GBandsSrf2Int$ , die Anzahl aller G-Banden auf der Oberfläche dividiert durch die Anzahl aller G-Banden im Inneren,  $RBandsSrf2Int$  analog für R-Banden,  $G2RSrf$ , die Anzahl aller G-Banden auf der Oberfläche dividiert durch die Anzahl aller R-Banden auf der Oberfläche und  $G2RInt$  analog für das Innere der Territorien. Der Vergleich zwischen Kernen mit speckles und Kernen ohne speckles besagt bei allen Quotienten, daß die Anwesenheit von speckles zu einer Konzentration von G-Banden im Inneren der Territorien und zu einer Anreicherung von R-Banden auf der Oberfläche der Territorien führt. Dies ist an sich nicht weiter verwunderlich, denn so waren ja die Wechselwirkungen gewählt. Von Bedeutung ist jedoch, daß die Umordnung, welche die G-Banden erfahren, auf den ganzen Zellkern bezogen stärker ist als diejenige, welche die R-Banden erfahren. Eine biologische Interpretation dieser Beobachtung ist, daß an der Organisation des Chromatins im Zellkern zu Euchromatin und Heterochromatin die inaktiven Chromatinabschnitte stärker beteiligt sind als das aktive Chromatin. Dies gilt jedoch nicht für kleine Chromosomen (s.u.).

---

<sup>1</sup>auf demselben Chromosom liegenden

Kontakte zwischen Chromatindomänen und speckles treten auf der Oberfläche ca. 8.5 mal so häufig auf wie im Inneren. Es ist bemerkenswert, daß es überhaupt Kontakte zwischen speckles und Chromatindomänen im Inneren von Territorien gibt. Die speckles durchdringen stellenweise die Oberfläche und kontaktieren Domänen in tieferen Schichten. Dies stützt die im ICD-Modell erhobene These, daß der ICD-Raum sich teilweise bis ins Innere der Territorien erstreckt, daß sich also aktive Genregionen im Inneren der Territorien befinden können.

Die oben beschriebenen Quotienten  $GBandsSrf2Int$  und  $RBandsSrf2Int$  wurden für jedes Chromosom einzeln ausgewertet und gegen die Chromosomengröße aufgetragen. Bei großen Chromosomen beobachtet man bei Anwesenheit von speckles bei G-Bandendomänen eine stärkere Umordnung als bei R-Bandendomänen. Dieses Verhalten entspricht den Ergebnissen, die mit ganzen Zellkernen erzielt wurden (s.o.). Bei kleinen Chromosomen beobachtet man dagegen bei Anwesenheit von speckles bei R-Bandendomänen eine stärkere Umordnung als bei G-Bandendomänen.

Die Quotienten  $G2RSrf$  und  $G2RInt$  wurden gegen den R-Bandengehalt der Chromosomen aufgetragen.

Diese Auftragungen verdeutlichen, bei welchen Chromosomen der Einfluß der speckles am stärksten ist. Dies sind bezüglich des Parameters  $Srf2Int$  die Chromosomen Y, 22, 20, 15, 17, 12, 8, 7 und 4 und bezüglich des Parameters  $G2R$  die Chromosomen 14, 15, 18, 13, 20 (jeweils geordnet nach der Stärke des Einflusses der speckles). In der Schnittmenge der nach diesen Kriterien ausgezeichneten Chromosomen befinden sich die Chromosomen 15 und 20. Diese Information kann zur Orientierung bei der Planung von Experimenten dienen, da eine Vorhersage gemacht wird, auf welchen Chromosomen unter den gemachten Annahmen (Fibroblastenkerne, größenabhängige Chromosomenverteilung) die Verteilung der G- und R-Banden auf Oberfläche und Inneres des Territorien am ehesten beobachtet werden kann.

## 6.2 Ausblick

Mit den in dieser Arbeit geschaffenen technischen Voraussetzungen und Computerprogrammen lassen sich weitere interessante Fragestellungen bearbeiten und die bereits durchgeführten Untersuchungen vertiefen.

So könnte analog zu der in Abschnitt 6.1.2 beschriebenen unteren Grenze für die Größe der speckles mit den bereits bestehenden Algorithmen und einiger Rechenzeit eine obere Grenze für die Größe der speckles ermittelt werden. Beide Grenzen könnten im Rahmen einer Optimierung näher an den tatsächlichen Wert herangeführt werden und so eine genaue Angabe über die Größe und das Gesamtvolumen der speckles liefern.

Die in Abschnitt 6.1.4 durchgeführten Analysen der Gyrationenradien spezieller Genregionen könnten an anderen Genregionen durchgeführt werden, welche größer und heterogener bezüglich ihrer Zusammensetzung aus G- und R-Banden sind. Aufgrund ihrer unterschiedlichen Wechselwirkungen mit den speckles würden die G-Banden dieser Genregion in das Innere ihres Chromosomenterritoriums gedrückt, die R-Banden auf die Oberfläche. Damit könnte der Einfluß der speckles auf lokale Verformungen und Spannungen im Chromatin untersucht werden.

Die Analyse der Verteilung von G- und R-Banden könnte neben Oberfläche und Innerem der Chromosomenterritorien auch für die sogenannten isolierten Domänen durchgeführt werden. Isolierte Domänen sind diejenigen, welche zu keiner anderen Chromatindomäne Kontakt haben. Man erwartet hier, daß bei den isolierten Domänen G-Banden unterrepräsentiert und R-Banden überrepräsentiert sind.

Durch die variabel gestaltete Anzahl der speckles in einer SCD-Zellkernkonfiguration könnte untersucht werden, ob und wenn ja wie es zu einer Aggregation vieler kleiner speckles zu großen

Proteinkomplexen wie den SC35-Domänen kommt. Die Programme für die Generierung von Startkonfigurationen mit vielen kleinen speckles liegen vor, das Relaxationsprogramm kann mit solchen Konfigurationen arbeiten. Es müsste bei der Analyse die Wechselwirkung zwischen speckles untereinander genauer untersucht und eventuell variiert werden. Viele der in dieser Arbeit vorgestellten Kriterien zur Analyse von relaxierten Konfigurationen, insbesondere die Untersuchungen zum Kontaktverhalten von G- und R-Banden in 6.1.3 könnten bei diese Fragestellung ebenfalls ausgewertet werden.

Interessant wäre ebenfalls die Integration von genaueren Daten über die Gendichte einzelner Chromatin-Domänen in SCD-Zellkernkonfigurationen. Damit könnte man eine Implementierung von Chromatin-Domänen variabler Größe erreichen und dem unterschiedlichen Kompaktierungsgrad zwischen R-Banden und G-Banden Rechnung tragen. Dies erfordert im Gegensatz zu den oben beschriebenen Vorschlägen jedoch zusätzliche Programmierfähigkeit. Eine weitere umfangreiche Erweiterung wäre die Implementierung als Brownsche Dynamik Simulation.

# Literaturverzeichnis

- [ABJ<sup>+</sup>99] B. Alberts, D. Bray, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. *Lehrbuch der molekularen Zellbiologie*. VCH Verlagsgesellschaft mbH, 1999.
- [AT89] M.P. Allen and D.J. Tildesley. *Computer Simulation of Liquids*. Oxford University Press, 1989.
- [BC74] R. Berezney and D.S. Coffey. Identificaion of nuclear protein matrix. *Biochem. Biophys. Res. Commun.*, 60:1410–1417, 1974.
- [BH97] K. Binder and D.W. Heermann. *Monte Carlo Simulation in Statistical Physics*. Springer Verlag, Berlin, Heidelberg, New York, Tokyo, third edition, 1997.
- [CCB<sup>+</sup>82] T. Cremer, C. Cremer, H. Baumann, E.K. Luedtke, K. Sperling, V. Teubner, and C. Zorn. Rabl’s model of the interphase chromosome arrangement, tested in chinese hamster cells by premature chromosome condensation and Laser-UV-Microbeam experiments. *Hum. Genet.*, 60:46–56, 1982.
- [CKK<sup>+</sup>00] T. Cremer, G. Kreth, H. Koester, R.H.A. Fink, R. Heintzmann, M. Cremer, I. Solovei, D. Zink, and C. Cremer. Chromosome territories, interchromatin domain compartment, and nuclear matrix: an integrated view of the functional nuclear architecture. *Critical Reviews in Eukaryotic Gene Expression*, 12(2):179–212, 2000.
- [CKZ<sup>+</sup>93] T. Cremer, A. Kurz, R. Zirbel, S. Dietzel, B. Rinke, E. Schröck, M.R. Speicher, U. Mathieu, A. Jauch, P. Emmerich, H. Scherthan, T. Ried, C. Cremer, and P. Lichter. Role of chromosome territories in the functional compartmentalization of the cell nucleus. *Symp. Quant. Bio. Cold Spring Harb.*, pages 777–792, 1993.
- [CLB<sup>+</sup>88] T. Cremer, P. Lichter, J. Borden, D.C. Ward, and L. Manuelidis. Detection of chromosome aberrations in metaphase and interphase tumor cells by in situ hybridization using chromosome specific library probes. *Hum. Genet.*, 80:235, 1988.
- [Doi97] M. Doi. *Introduction to polymer physics*. Oxford University Press, 1997.
- [HHB<sup>+</sup>93] P. Hahnfeld, J.E. Hearst, D.E. Brenner, R.K. Sachs, and L.H. Hlatky. Polymer models for interphase chromosomes. *Proc. Natl. Acad. Sci. USA*, 90:7854–7858, 1993.
- [HHLS<sup>+</sup>89] J.S. Heslop-Harrison, A.R. Leitch, T. Schwarzacher, J.B. Smith, M.D. Atkison, and M.D. Bennet. The volumes and morphology of human chromosomes in mitotic reconstructions. *Human Genetics*, 84(1):27–34, 1989.
- [JP98] D.A. Jackson and A. Pombo. 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.
- [KPSF90] R. Knippers, P. Philippsen, K.P. Schäfer, and E. Fanning. *Molekulare Genetik*. Georg Thieme Verlag, Stuttgart, New York, 1990.

- [Kre00] Gregor Kreth. *Simulation von Chromosomen und virtuelle Mikroskopie*. PhD thesis, Fakultät für Physik und Astronomie, Universität Heidelberg, 2000.
- [LCB<sup>+</sup>88] P. Lichter, T. Cremer, J. Borden, L. Manuelidis, and D.C. Ward. Delineation of individual human chromosomes in metaphase and interphase cells by in situ suppression hybridization using recombinant DNA libraries. *Hum. Genet.*, 80:224, 1988.
- [LL86] L.D. Landau and E.M. Lifshitz. *Theory of Elasticity*. Pergamon Press, Oxford, New York, 1986.
- [ML79] M.P.F. Marsden and U.K. Laemmli. Metaphase chromosome structure: evidence for a radial loop modell. *Cell*, 17:849–858, 1979.
- [ML98] Ch. Münkler and J. Langowski. Chromosome structure predicted by a polymer model. *Physical Review E*, 57:5888–5897, 1998.
- [NB89] H. Nakayasu and R. Berezney. Mapping replicational sites in the eukaryotic cell nucleus. *Journal of Cell Biology*, 108:1–11, 1989.
- [NMS86] H. Nakamura, T. Morita, and C. Sato. Structural organization of replicon domains during dna synthetic phase in the mammalian nucleus. *Experimental Cell Research*, 165:291–297, 1986.
- [PLC<sup>+</sup>88] D. Pinkel, J. Landegent, C. Collins, J. Fuscoe, R. Segarves, J. Lucas, and J.W. Gray. Fluorescence in situ hybridization with human chromosome-specific libraries: Detection of trisomy 21 and translocation of chromosome 4. *Proc. Natl. Acad. Sci.*, 85:9138, 1988.
- [RL85] J.B. Rattner and C.C. Lin. Radial loops and helical colis coexist in metaphase chromosomes. *Cell*, 42:291–296, 1985.
- [SJB<sup>+</sup>03] L.S. Shopland, C.V. Johnson, M. Byron, J. McNeil, and J.B. Lawrence. Clustering of multiple specific genes and gene-rich r-bands around sc-35 domains: evidence for local euchromatic neighbourhoods. *Journal of Cell Biology*, 162:981–990, 2003.
- [Str97] G. Strobel. *The Physics of Polymers*. Springer Verlag, Berlin, Heidelberg, New York, second edition, 1997.
- [VV95] D. Voet and J.G. Voet. *Biochemisty*. John Wiley & Sons Inc., second edition, 1995.
- [ZCCZ79] C. Zorn, C. Cremer, T. Cremer, and J. Zimmer. Unscheduled DNA synthesis after partial UV irradiation of the cell nucleus. *Exp. Cell Res.*, 124:111, 1979.



# Danksagung

Mein Dank gilt allen, die diese Arbeit möglich gemacht haben, die mir geholfen oder ihre Hilfe angeboten haben.

Prof. Dr. Dr. Christoph Cremer hat mich durch seine Begeisterung für diese Arbeit immer aufs Neue motiviert.

Insbesondere möchte ich meinem Betreuer Gregor Kreth danken, für all die Mühe und Arbeit, die er mit mir hatte und für alle Hilfe, ohne die diese Arbeit nicht möglich gewesen wäre.

Der ganzen Arbeitsgruppe möchte ich für die freundliche Aufnahme und gute Zusammenarbeit danken.

Meiner Familie möchte ich, um es kurz zu machen, einfach für alles danken.

Zum Schluß will ich all denen danken, die mir sonst noch geholfen haben, unter anderem Nick für seine Tipps bei Computer- und  $\text{\LaTeX}$ -Problemen.



## Erklärung

Ich versichere, daß ich diese Arbeit selbständig verfaßt und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Heidelberg, den .....

.....