

# **Simulation von Struktur und Strukturregulation von Chromatin**

Dissertation  
zur Erlangung der Doktorwürde  
des Fachbereichs Physik  
der Bergischen Universität–Gesamthochschule Wuppertal

vorgelegt von  
Martin Gero Wedemann  
aus Karlsruhe

Oktober 1999  
WUB–DIS 99–23



# Computer Simulation of Chromatin Structure and Regulation

## Abstract

DNA in the nucleus of higher organisms is complexed with proteins into highly compact structures designated as chromatin. It is built of nucleosomes that are connected by linker DNA resulting in a bead-on-a-string like structure. Nucleosomes are formed by DNA wrapped around special proteins, the histones.

A new model for the structure of chromatin is presented here. It is based on the zigzag model as well as previous work on superhelical DNA and polynucleosomes. DNA is modeled as a flexible elastic chain, neglecting atomic details. The electrostatic interaction between DNA segments is described by a Debye-Hückel-approximation. The nucleosomes are represented by flat disks which interact with each other through a Gay-Berne-potential. DNA and nucleosomes are linked either at the cylinder of the nucleosome-core-particle or at the nucleosome stem. Representative ensembles of configurations are created by a Metropolis Monte Carlo-algorithm. The simulation program was implemented in C++ for use with parallel computers.

Simulations have been performed on polynucleosomes containing 100 nucleosomes at physiological ionic strength. For models with a linkage of the DNA at the nucleosome stem and a repeat length of 200 bp the characteristic experimental properties of chromatin have been reproduced: The calculated fiber diameter of 32 nm and the mass density of 6.1 nucleosomes per 11 nm fiber length are in agreement with experimental values from the literature. The experimental value of the inclination of DNA and nucleosomes to the fiber axis is also reproduced. The persistence length of the simulated fibers of 260 nm is compatible with the values of 50–150 nm estimated by experimental data. For random tilt angles between two nucleosomes chosen according to a Gaussian distribution the persistence length decreases with the increase of the width of the distribution. The experimentally observed decrease of the mass density with decreasing concentration of salt in the surrounding medium is reproduced qualitatively. Polynucleosomes with repeat lengths of 212 bp also form fibers with the expected experimental properties. Systems with larger repeat length form fibers but the mass density is significantly lower than the measured value. The theoretical characteristics of a fiber with a repeat length of 192 bp where DNA and nucleosomes are connected at the core-particle are in agreement with the experimental values. Systems with the same type of linkage and a repeat length of 217 bp do not form fibers.



# Inhaltsverzeichnis

<b>Vorwort</b>	<b>vi</b>
<b>1 Biologische Grundlagen</b>	<b>1</b>
1.1 DNA Struktur . . . . .	1
1.1.1 Primärstruktur . . . . .	1
1.1.2 Struktureigenschaften . . . . .	2
1.2 Chromatinfiber . . . . .	4
1.2.1 Nukleosom . . . . .	5
1.2.2 Eigenschaften der Chromatinfiber . . . . .	9
1.2.3 Modelle der Chromatinfiber . . . . .	18
1.2.4 Weitere chromosomale Proteine . . . . .	20
<b>2 Physikalische Grundlagen</b>	<b>23</b>
2.1 Grundlagen der Polymerphysik . . . . .	23
2.1.1 Gaußsche Kette . . . . .	23
2.1.2 Kette mit Biegeelastizität . . . . .	25
2.1.3 Kette mit Torsionselastizität . . . . .	26
2.2 Flüssigkristalle . . . . .	26
2.3 Monte Carlo–Verfahren . . . . .	28
2.3.1 Mastergleichung . . . . .	29
2.3.2 Simple Sampling . . . . .	29
2.3.3 Importance Sampling — Metropolis Algorithmus . . . . .	30
<b>3 Modell für DNA und Chromatin</b>	<b>33</b>
3.1 Diskretisierung . . . . .	34
3.1.1 Diskretisierung der DNA . . . . .	34
3.1.2 Diskretisierung der Nukleosomen . . . . .	35
3.2 Wechselwirkungen . . . . .	36
3.2.1 Elastische Energien . . . . .	37
3.2.2 Elektrostatische Energie der DNA . . . . .	38
3.2.3 Internukleosomale Kräfte . . . . .	41

---

3.3	Monte Carlo-Schritte . . . . .	43
3.4	Parametrisierung der Wechselwirkungspotentiale . . . . .	44
3.4.1	Parametrisierung der elastischen Wechselwirkungen . . . . .	46
3.4.2	Parametrisierung des Gay-Berne-Potentials . . . . .	46
3.4.3	Länge der Verbindungs-DNA . . . . .	48
<b>4</b>	<b>Aufbau und Verifikation des Rechen- und Auswerteprogramm-</b>	
	<b>packets</b> . . . . .	<b>51</b>
4.1	Überblick objektorientierte Programmierung . . . . .	52
4.2	Programmstruktur . . . . .	53
4.3	Einheitensystem . . . . .	54
4.4	Zufallszahlengenerator . . . . .	55
4.5	Parallelisierung . . . . .	56
4.6	Verifikation . . . . .	58
4.6.1	Formale Verifikation zur Compile-Zeit . . . . .	59
4.6.2	Überprüfung zur Laufzeit . . . . .	59
4.6.3	Inhaltliche Verifikation anhand von Resultaten . . . . .	60
<b>5</b>	<b>Algorithmen zur Auswertung</b> . . . . .	<b>61</b>
5.1	Konturlänge und Massenbelegungsichte . . . . .	61
5.2	Durchmesser der Fiber . . . . .	63
5.3	Neigung der Nukleosomen und der DNA zur Fiberachse . . . . .	63
5.4	Persistenzlänge . . . . .	63
5.5	Visualisierung . . . . .	64
<b>6</b>	<b>Ergebnisse</b> . . . . .	<b>65</b>
6.1	Ankopplung der Verbindungs-DNA am Nukleosomen-Stamm . . . . .	65
6.1.1	Korrelation der simulierten Daten . . . . .	66
6.1.2	Anzahl der Simulationschritte . . . . .	67
6.1.3	Öffnungswinkel Linker-DNA . . . . .	69
6.1.4	Verdrehungswinkel zwischen Nukleosomen . . . . .	69
6.1.5	Relaxierung aus gestrecktem Zustand . . . . .	73
6.1.6	Einfluß der Stärke der internukleosomalen Wechselwir- kung . . . . .	75
6.1.7	Einfluß der Salzkonzentration . . . . .	75
6.1.8	Persistenzlänge . . . . .	78
6.1.9	Internukleosomaler Abstand . . . . .	81
6.1.10	Einfluß der Länge der Linker-DNA . . . . .	81
6.2	Ankopplung der Verbindungs-DNA am Core-Zylinder . . . . .	85
<b>7</b>	<b>Diskussion und Zusammenfassung</b> . . . . .	<b>87</b>

---

<b>Ausblick</b>	<b>94</b>
<b>A Momente höherer Ordnung der idealen Zufallskette</b>	<b>97</b>
<b>B Formale Programmstruktur</b>	<b>99</b>
B.1 Übersicht UML . . . . .	99
B.1.1 Use-Case-Diagramm . . . . .	100
B.1.2 Packetdiagramm . . . . .	100
B.1.3 Sequenzdiagramm . . . . .	101
B.1.4 Klassenstrukturdiagramm . . . . .	101
B.2 Struktur des Simulations- und Analysepackets . . . . .	102
B.2.1 Use-Case-Diagramm . . . . .	102
B.2.2 Packetdiagramm . . . . .	102
B.2.3 Klassendiagramme . . . . .	106
B.2.4 Sequenzdiagramme . . . . .	109
<b>Abbildungsverzeichnis</b>	<b>114</b>
<b>Tabellenverzeichnis</b>	<b>117</b>
<b>Literaturverzeichnis</b>	<b>119</b>



# Vorwort

Well we know what make the flowers grow  
But we don't know why  
And we all have the knowledge of DNA  
But we still die

*Sullivan/Heaton/Harris: White Coats  
New Model Army, Thunder and  
Consolation. EMI. 1987.*

Das Erbgut fast aller Organismen ist auf DNA-Molekülen codiert. Die Struktur dieser Moleküle, die berühmte Watson-Crick-Doppelhelix, ist inzwischen in das Allgemeinwissen eingegangen. Künstliche Veränderungen der DNA sind Laboralltag. Ein neuer Industriezweig, *Biobusiness*, beschäftigt sich ausschließlich mit der Erforschung und Nutzung dieses Wissens. Erbgut ist heute patentierbar. In der Presse werden bereits Themen wie das Klonen berühmter Zeitgenossen diskutiert — Es scheint, als sei das Erbgut beherrschbar geworden.

Eine genaue Analyse des Wissenstands der molekularen Biochemie beleuchtet jedoch eine andere Seite. Viele Techniken wurden rein heuristisch gewonnen, und es ist häufig unbekannt, warum sie funktionieren oder warum auch nicht. Die vorliegende Arbeit liefert einen Beitrag zum Grundlagenwissen der molekularen Biochemie und hilft, grundlegende Eigenschaften des Erbguts zu verstehen.

In den letzten Jahren hat sich herausgestellt, daß die Vorstellung, die im Erbgut enthaltene Information werde nur in der Abfolge der Basenpaare der DNA codiert, unvollständig ist, denn auch der räumliche Aufbau spielt eine entscheidende Rolle. Durch die Verpackung der DNA in kompakten Strukturen wird verhindert, daß darauf befindliche Information abgelesen werden kann. Eine günstige räumliche Orientierung verschiedener Abschnitte der DNA kann hingegen die Aktivität von Genen verstärken. Um diese Effekte besser zu verstehen, ist es notwendig, den Aufbau und die Regulation von komplexeren DNA-Strukturen zu klären.

Bei höheren Lebewesen wie dem Mensch ist die DNA in Chromatin organisiert. Im Chromatin ist die DNA um Proteine, die Histone, gewickelt und bildet zylinderförmige Einheiten. Diese Einheiten, die Nukleosomen, sind durch DNA verbunden und bilden so eine Struktur, die an eine Perlenkette erinnert. Sie faltet sich bei physiologischen Bedingungen in eine Fiber. Die Struktur dieser Fiber beeinflusst die Aktivität von Genen entscheidend. Sie wird von der Zelle aktiv kontrolliert. Modelle der inneren Struktur der Fiber sind bis heute umstritten.

Basierend auf älteren, einfachen Modellen wird in dieser Arbeit ein neuartiges physikalisches Modell für die Chromatinfiber vorgestellt. Mit Hilfe von Computersimulationen werden die Eigenschaften dieses physikalischen Modells untersucht. Das Modell gibt die Eigenschaften der realen Fiber quantitativ wieder. Hiermit ist es nun möglich, den Einfluß beliebiger Einflußparameter, insbesondere Modifikationen der Histone, auf die Struktur der Fiber zu untersuchen und damit die Regulation der Struktur besser verstehen und nutzen zu können.

Die vorliegende Arbeit wurde am Deutschen Krebsforschungszentrum (DKFZ) in Heidelberg in der Arbeitsgruppe „Biophysik der Makromoleküle“ unter der Leitung von Prof. J. Langowski durchgeführt. Den Mitgliedern der Arbeitsgruppe möchte ich für die gute Zusammenarbeit danken. Herrn Prof. Langowski schulde ich Dank für den vertrauensvollen Umgang; seine interdisziplinäre Kompetenz hat zum Erfolg der Arbeit entscheidend beigetragen. Besonderen Dank möchte ich Herrn Prof. P. Grassberger von der BUGH Wuppertal aussprechen; ohne seine Unterstützung, nicht nur in Form von zahlreichen Anregungen und konstruktiver Kritik, wäre diese Arbeit nicht entstanden. Dr. C. Münkel, nun bei der SAP AG Walldorf, danke ich für die Betreuung in der Anfangsphase der Arbeit; viele entscheidende Ideen sind in dieser Zeit entstanden. Dr. K. Rippe (DKFZ) danke ich für die Beantwortung vieler Fragen zu biologischen Themen und die Durchsicht des Manuskripts.

Weiterhin möchte ich mich beim DKFZ für die Bereitstellung der benötigten großen Rechnerkapazität bedanken. Das Institut für wissenschaftliches Rechnen (IWR) der Universität Heidelberg und die Universität Mannheim hat den Linux-Cluster *Comyc* zur Verfügung gestellt, auf dem ebenfalls ein wesentlicher Teil der hier vorgestellten Ergebnisse berechnet wurde. Andere Benutzer werden sicher aufatmen, daß die Maschinen nun wieder weniger ausgelastet sind.

Die vorliegende Arbeit wurde ermöglicht durch Mittel des Bundesministeriums für Bildung und Forschung (BMBF) im Rahmen des Humanen Genom Projekts (HUGO) und durch eine Abschlußförderung des Graduiertenkollegs am IWR Heidelberg und der Arbeitsgruppe Biocomputing von Prof. J. Smith.

# Kapitel 1

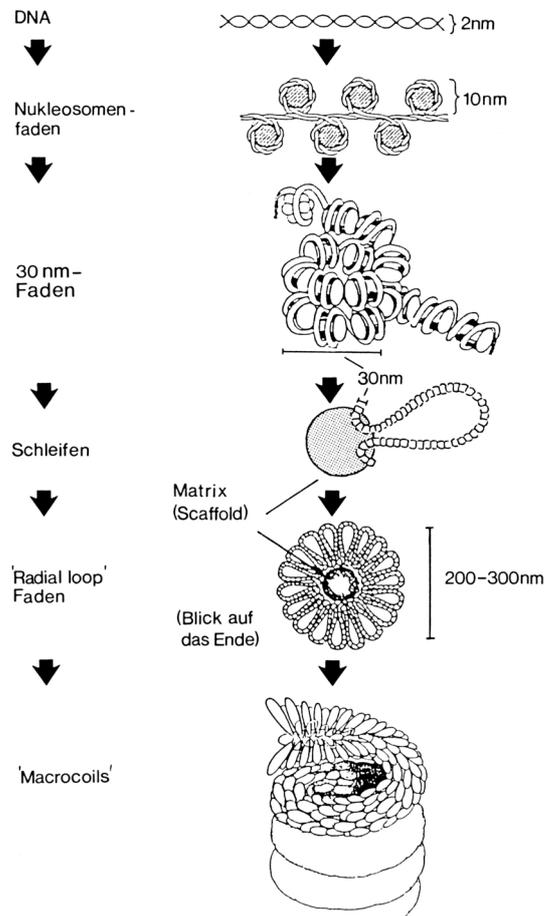
## Biologische Grundlagen

Die Erbinformation von Zellen befindet sich auf einem einem langen, fadenförmigen Makromolekül, der Desoxyribonukleinsäure (DNA). Bei Zellen höher entwickelter Lebewesen schwimmt die DNA nicht frei im Zell-Plasma, sondern ist im Zellkern lokalisiert. Bei Eukaryonten, zu denen der Mensch gehört, kann man verschiedene Organisationsstufen der DNA unterscheiden (Abbildung 1.1): DNA ist gefaltet in die Chromatinfiber; Schleifen aus Chromatin bilden Chromosomen; die Chromosomen bilden den Zellkern.

### 1.1 DNA Struktur

#### 1.1.1 Primärstruktur

Desoxyribonukleinsäure ist ein Polymer, das üblicherweise aus vier verschiedenen Nukleotiden zusammengesetzt ist. Ein Nukleotid besteht aus einem Zucker ( $\beta$ -D-2'-Desoxyribose), einer (negativ geladenen) Phosphatgruppe und einer von vier heterozyklischen Basen: Adenin (A), Guanin (G), Thymin (T), Cytosin (C) (siehe Abb. 1.2). Die Sequenz der Basen trägt die genetische Information, während die Zucker- und Phosphatgruppen eine strukturelle Aufgabe erfüllen [Str91]. Die Polymerisierung findet durch 3'-5' Phosphodiester-Bindung statt. Die Kette der so verbundenen Monomere bezeichnet man als Einzelstrang mit einer Richtung vom 5'-Ende zur 3'-Ende. Ein Doppelstrang kann sich aus zwei Einzelsträngen durch Ausbildung von Wasserstoffbrücken zwischen Adenin und Thymin bzw. Guanin und Cytosin bilden (siehe Abb. 1.3). Die Basenpaarung ist dabei hochspezifisch, da sich bei der A-T Bindung zwei, bei der G-C Bindung drei Wasserstoffbrücken ausbilden. Jedes Basenpaar ist aufgrund der geladenen Phosphatgruppen zweifach geladen.

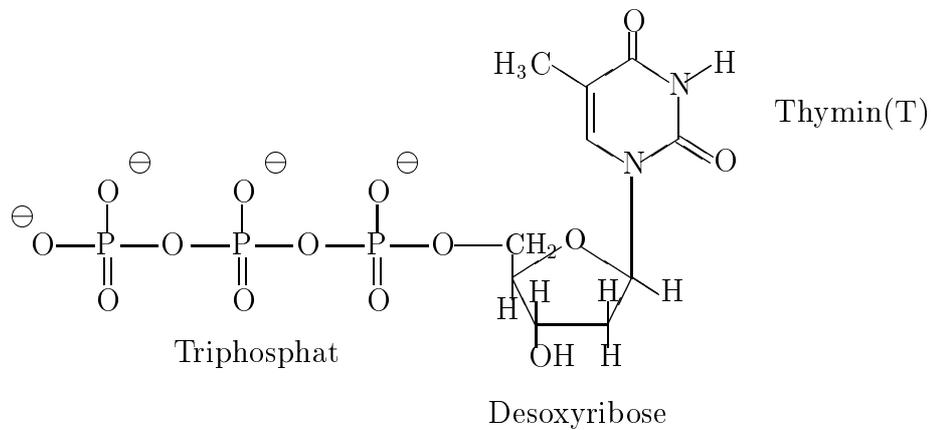


**Abbildung 1.1:** Klassische Lehrbuchansicht der Kompaktierung der DNA im Zellkern in Metaphase. Die abgebildeten *Macrocoils* bilden nach dieser Ansicht im wesentlichen die Chromosomen. Aus [Tra91].

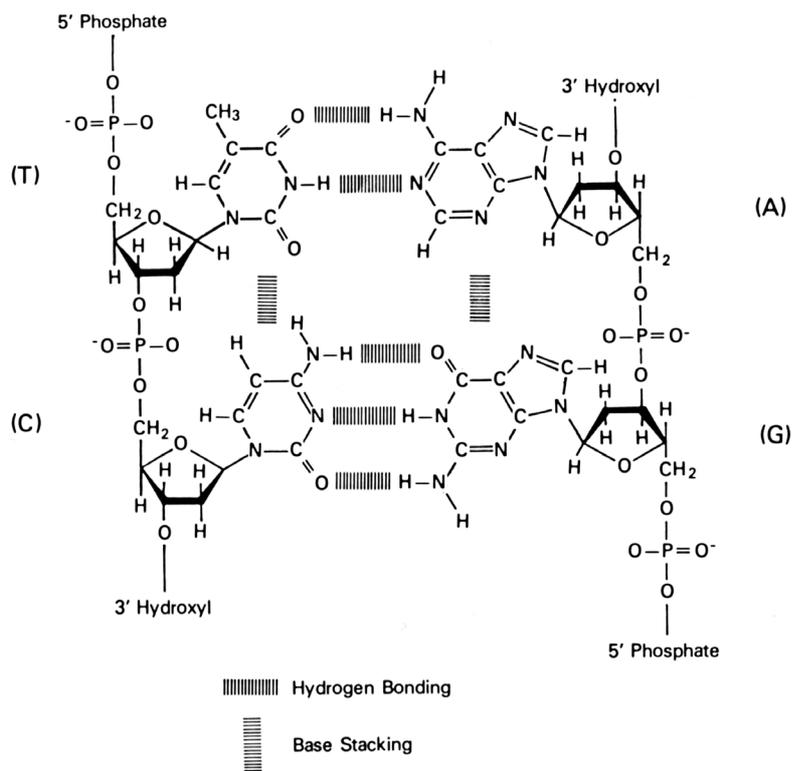
### 1.1.2 Struktureigenschaften

Die räumliche Struktur der DNA wurde 1953 von Watson und Crick aus Ergebnissen von Röntgenstreuexperimenten ermittelt:

- Der Doppelstrang ist um eine gemeinsame Achse helikal gewunden (Doppelhelix). Die Einzelstränge laufen entgegengesetzt.
- Die Basenpaare sind zum Inneren der Helix orientiert, das Zucker-Phosphat-Rückgrat liegt außen.
- Der Durchmesser der Helix beträgt ca. 2 nm.



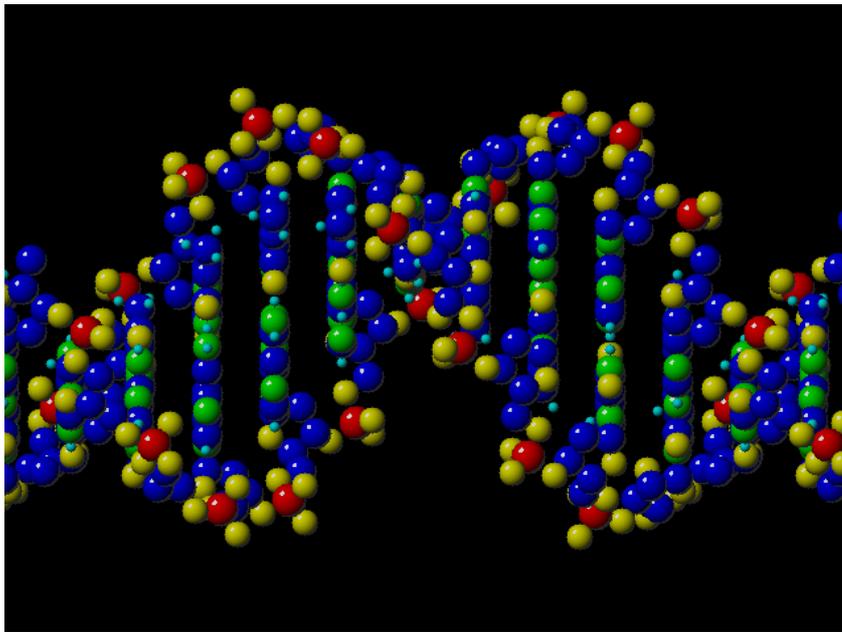
**Abbildung 1.2:** Beispiel eines Nucleotids (DNA-Monomer): Desoxythymidin-5'-triphosphat (dTTP)



**Abbildung 1.3:** Zwei Basenpaare einer DNA (aus [Wol95])

Die DNA kann dabei in verschiedenen Konformationen vorliegen. In biologischen Systemen wird die DNA meist in der B-Form beobachtet. Diese ist durch folgende Eigenschaften gekennzeichnet:

- Rechtshändigkeit der Doppelhelix
- 10,4 Basenpaare (bp) pro Helix-Windung
- 0,34 nm Ganghöhe pro bp.



**Abbildung 1.4:** Räumliches Modell einer DNA-Doppelhelix. Die jeweilige Farbe entspricht dem Typ des Atoms: blau – C, rot – P, gelb – O, grün – N, hellblau – H.

Die Länge der DNA ist für verschiedene Organismen sehr unterschiedlich. Die DNA des SV40-Virus besitzt beispielsweise 5243 Basenpaaren (bp), ein Hefechromosom 350 kbp, das *E. coli*-Genom 4.6 Mbp und das menschliche Genom 6 Gbp.

## 1.2 Chromatinfiber

Als Chromatin bezeichnet man die mit basischen Farbstoffen anfärbbare Substanz der Chromosomen und der Zellkerne höherentwickelter Lebewesen. Man beobachtet, daß Chromatin in Fäden mit einem Durchmesser von

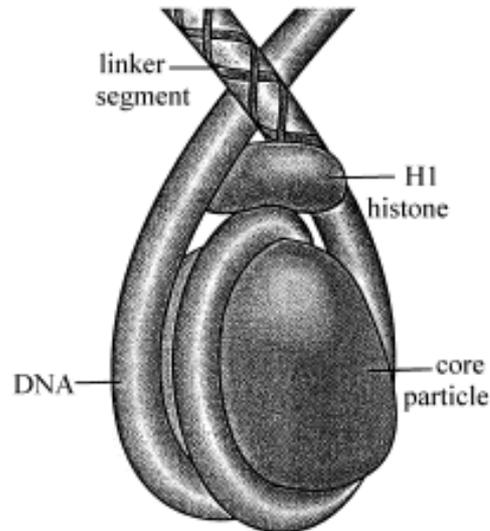
30 nm organisiert ist, die man als Chromatinfibern bezeichnet. Sie bestehen aus DNA, die mit einer Gruppe basischer Proteine, den Histonen, fest verbunden ist [Hol89, Wol95]. Diese Proteine sind ein wesentlicher Bestandteil der Chromosomen. Sie machen etwa die Hälfte der Masse von Eukaryontenchromosomen aus. Die Struktur von Chromatin kann die Transkriptionsaktivität, also das Ablesen der Erbinformation von DNA, regulieren [Fel96, Wol94]. Transkriptionsaktivität kann beispielsweise unterbunden werden, indem die DNA so verpackt wird, daß Polymerasen — Proteine, die die Transkription durchführen — nicht in Kontakt mit der DNA gelangen können und so die dort enthaltene Information nicht ablesen können.

### 1.2.1 Nukleosom

Beobachtet man Chromatin in Pufferlösungen mit niedrigem Salzgehalt, erkennt man Fäden, die mit ca. 10 nm großen, rundlichen Gebilden, den Nukleosomen, besetzt sind. Nukleosomen bestehen aus den Histonen H2A, H2B, H3 und H4 sowie DNA-Abschnitten von ungefähr 200 nm Länge. Meistens ist noch ein sogenanntes Linker-Histon vorhanden, z.B. H1 oder H5. Je zwei Moleküle von H2A, H2B, H3 und H4 bilden den Histonkern, ein Oktamer, um den sich die DNA-Helix in etwa zwei Links-Windungen herumlegt (Abbildung 1.5). Man unterscheidet begrifflich zwischen dem Nukleosomen-Core-Partikel, dem Chromatosom und dem Nukleosom: Das Nukleosomen-Core-Partikel besteht aus dem Histon-Oktamer und der daran gebundenen DNA. Zusammen mit dem Linker-Histon und der daran gebundenen DNA, bildet es das Chromatosom. Das Nukleosom besteht aus dem Chromatosom und der Linker-DNA, die die Chromatosomen verbindet. Häufig wird der Begriff Nukleosom synonym mit dem Begriff Chromatosom verwendet. Dies wird auch im folgenden so gehandhabt, soweit dies nicht zu Unklarheiten führt.

Das Nukleosomen-Core-Partikel besteht aus einem Histon-Oktamer, um das 146 bp DNA in 1,65 Windungen in einer flachen Superhelix gewunden sind. Es ist zylinderförmig mit einem Durchmesser von 11 nm und einer Dicke von 5,5 nm. Die Position der Atome des Histonkerns und der DNA sind aus Röntgenstreuexperimenten an Kristallen des Nukleosomen-Core-Partikel in einer Auflösung von 2,8 Å bekannt (Abbildung 1.6) [LMR<sup>+</sup>97]. Die Lage der N-terminalen Protein-Schwänze der Histone läßt sich auf diese Weise nicht bestimmen, da sie anscheinend keine feste Position einnehmen.

Das Chromatosom besteht aus dem Nukleosomen-Core-Partikel und weiteren 20 bp DNA. Es besitzt eine Gesamtmasse von 230 kDa. Die 20 bp zusätzliche DNA sind an das Linker-Histon gebunden. Die genaue Lage des Linker-Histons ist nicht bekannt. Es wird aber eine zentrale, symmetrische Position wie in Abbildung 1.5 für wahrscheinlich gehalten [Tra99].



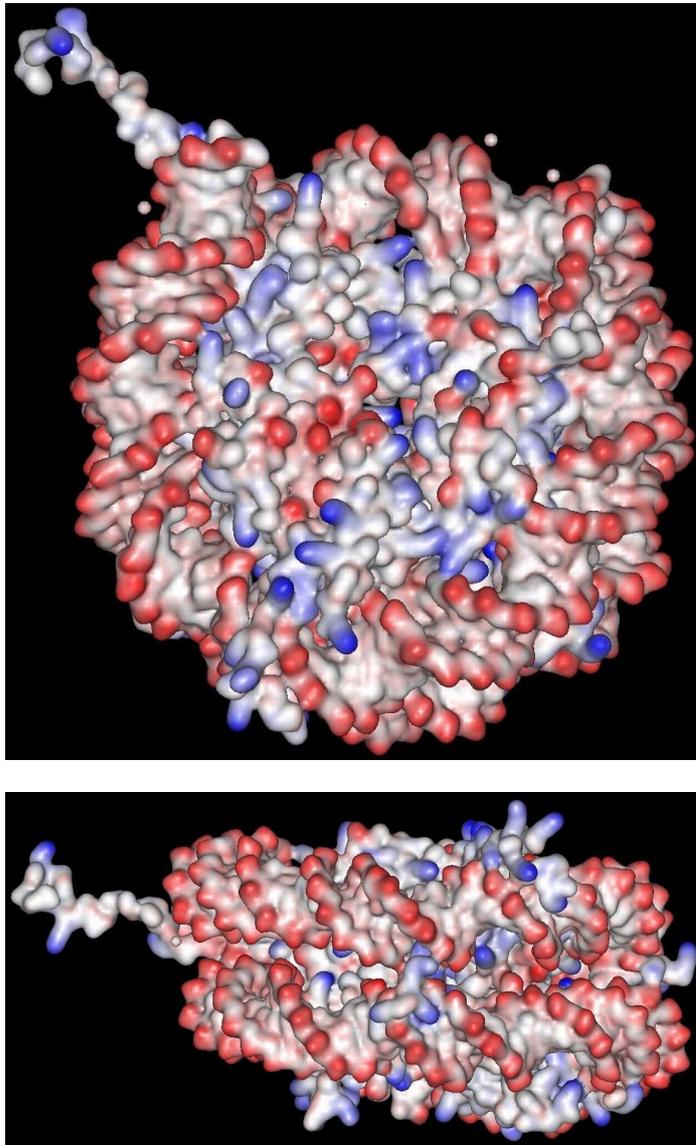
**Abbildung 1.5:** Skizze des Nukleosoms

Die Linker-Histone können sich bei verschiedenen Spezies und Zelltypen unterscheiden. So besitzt beispielsweise der Mensch verschiedene Varianten des Histons H1 und das Huhn Varianten von H5.

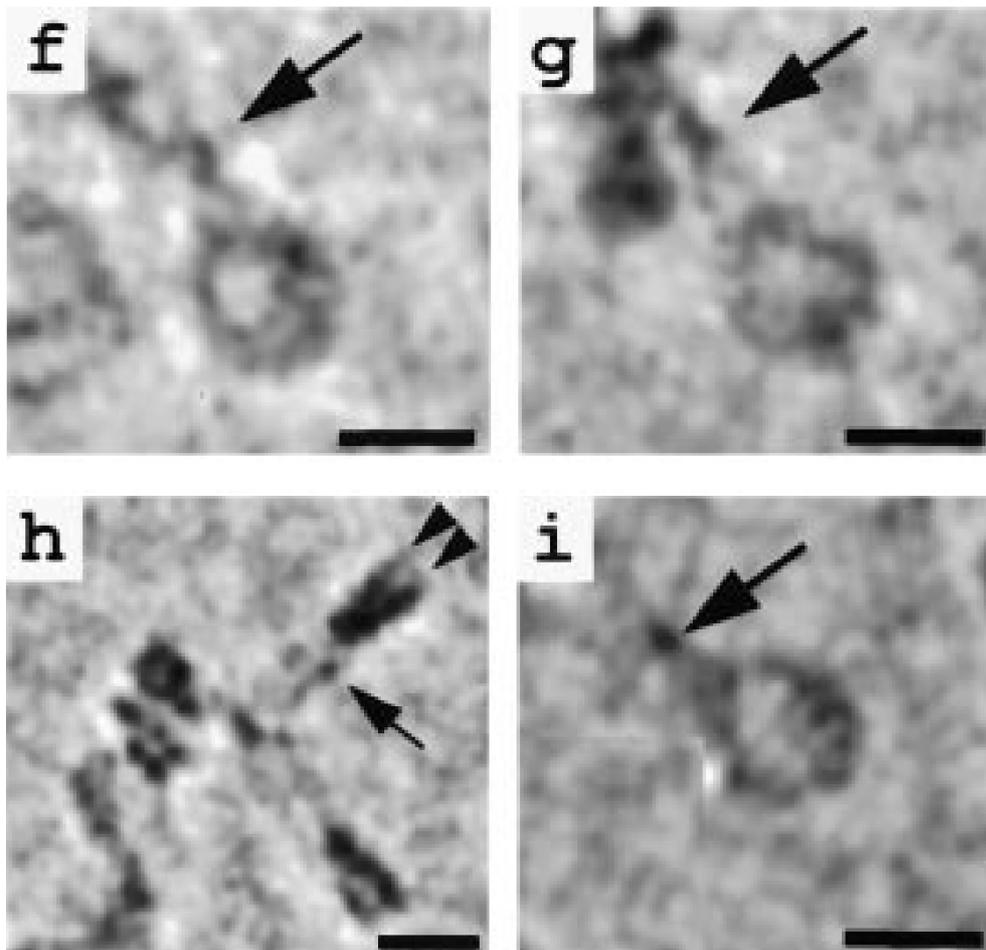
Verschiedene experimentelle Untersuchungen ergaben, daß das Linker-Histon den Winkel zwischen einlaufender und auslaufender DNA festlegt (Abbildung 1.12). Neuere Daten weisen außerdem darauf hin, daß das Linker-Histon bei Hühnerythrozyten zusammen mit einem Abschnitt der DNA eine stammförmige Struktur bildet (*stem-motif*) [ZLH98, BHG<sup>+</sup>98]. Die aus dem Nukleosomen-Core auslaufende DNA und das Linker-Histon bilden bis zu einer Entfernung von etwa 8 nm von der Achse des Nukleosomen-Core-Partikels eine zusammenhängende Einheit, den Nukleosomen-Stamm. Erst dann laufen die beiden DNA-Doppelstränge auseinander (Abbildung 1.7).

Der Winkel zwischen einlaufender und auslaufender DNA ist abhängig von der Salzkonzentration des Lösungsmittels in dem sich die Nukleosomen befinden. Orientierende Daten finden sich bei Woodcock et al. [WGHW93]: Hier wird ein Bild aus einem Übersichtsartikel über die Methode der Cryo-Elektronen-Mikroskopie<sup>1</sup> von Dubochet et al. [DAC<sup>+</sup>88] ausgewertet und

<sup>1</sup>Bei Cryo-Elektronenmikroskopie werden die Proben sehr schnell in flüssigen Stickstoff abgekühlt und anschließend mit konventionellen elektronenmikroskopischen Techniken beobachtet. Man geht davon aus, daß durch den schnellen Abkühlungsprozess keine Präparationsartefakte auftreten und die Struktur der Probe im wesentlichen erhalten bleibt.



**Abbildung 1.6:** Verschiedene Ansichten der atomaren Struktur eines Nukleosomen-Core-Partikels [LMR<sup>+</sup>97]. Die Visualisierung erfolgte mit dem Programm WebLab von Molecular Simulations Inc. (<http://www.msi.com/weblab>). Dargestellt ist die Oberfläche der Histone und der DNA. Die Färbung entspricht der Ladungsdichte an dieser Stelle. Die rote Farbe entspricht negativer, die blaue positiver Ladungsdichte. Man erkennt das stark negativ geladene Phosphat-Rückgrat der DNA-Doppelhelix am Rand des Zylinders und die positiv geladenen Histone in der Mitte. Links oben ist der einzige Histon-Schwanz zu sehen, dessen Lage und Geometrie bestimmt werden konnte.



**Abbildung 1.7:** Struktur von Nucleosomen aus Hühner-Erythrozyten in einer cryo-elektronenmikroskopischen Aufnahme [BHG<sup>+</sup>98]. Die großen Pfeile zeigen auf das Ende des Nucleosomen-Stamms. Die kleinen Pfeile markieren die zwei Windungen der DNA im Chromosom, das von der Seite zu sehen ist.

der mittlere Winkel auf  $20^{\circ}$ - $60^{\circ}$  abgeschätzt. Eine genauere, sehr sorgfältige Bestimmung mit Cryo-Elektronen-Mikroskopie erfolgte 1995 von Bednar et al. [BHDW95]:  $56^{\circ} \pm 19^{\circ}$  bei 5 mM NaCl und  $39^{\circ} \pm 11^{\circ}$  bei 20 mM NaCl (Definition: mM: =  $10^{-3}$  mol/l). Das fast identische Autorenkollektiv veröffentlichte 1998 folgende Werte [BHG<sup>+</sup>98]:  $85^{\circ}$  bei 5 mM NaCl,  $45^{\circ}$  bei 15 mM NaCl,  $35^{\circ} \pm 9^{\circ}$  bei 80 mM NaCl. Der Wert bei 80 mM NaCl wurde mit Elektron-Cryotomographie und anschließender 3D-Rekonstruktion ermittelt. Die Werte sind kritisch zu beurteilen. Die Diskrepanz bei 5 mM

zu der vorhergehenden Publikation wird nicht diskutiert. Von den Werten bei 5 mM und 20 mM wird keine Fehlergrenze angegeben. Es ist weiterhin auffällig, daß die angegebenen Werte zu dem ebenfalls in diesem Artikel vorgestellten Modell sehr gut passen.

Eine weitere Eigenschaft der Linker-Histone besteht in der Reduktion der Ladung der DNA. Um diese zu bestimmen, kann man die Ladung der Linker-Histone bestimmen. Sie kann grob abgeschätzt werden, indem man die Ladungen der Aminosäuren addiert<sup>2</sup>. Für menschliches H1 (H1F1) errechnet man so 51 Elementarladungen, für H5 vom Huhn (*gallus gallus*) 54 Ladungen. Eine detailliertere Analyse mit dem Programm SEDNTERP [Phi97] von fünf häufigen menschlichen Varianten von H1 ergibt im Mittel eine Ladung von 56,9 Ladungen und von H5 vom Huhn 61,6 Ladungen. Für die globuläre Domäne von H5 bestimmt man analog eine Ladung von 25. Dies bedeutet, daß mehr als die Hälfte der Ladungen auf den Schwänzen lokalisiert sind. Man kann also davon ausgehen, daß die 40 Ladungen der an das Linker-Histon gebundenen 20 bp DNA vollständig neutralisiert sind. Eine andere, detaillierte Analyse zeigt, daß das Linker-Histon 58 bis 59 Ladungen der Linker-DNA neutralisiert [CK90]. Diese Daten sind mit den Ergebnissen von Zlatanova et al. konsistent [ZLH98], die zeigen, daß die globuläre Domäne von H5 allein im Gegensatz zu vollständigem H5 nicht fähig ist, den Öffnungswinkel zwischen einlaufender und auslaufender DNA zu stabilisieren.

### 1.2.2 Eigenschaften der Chromatinfiber

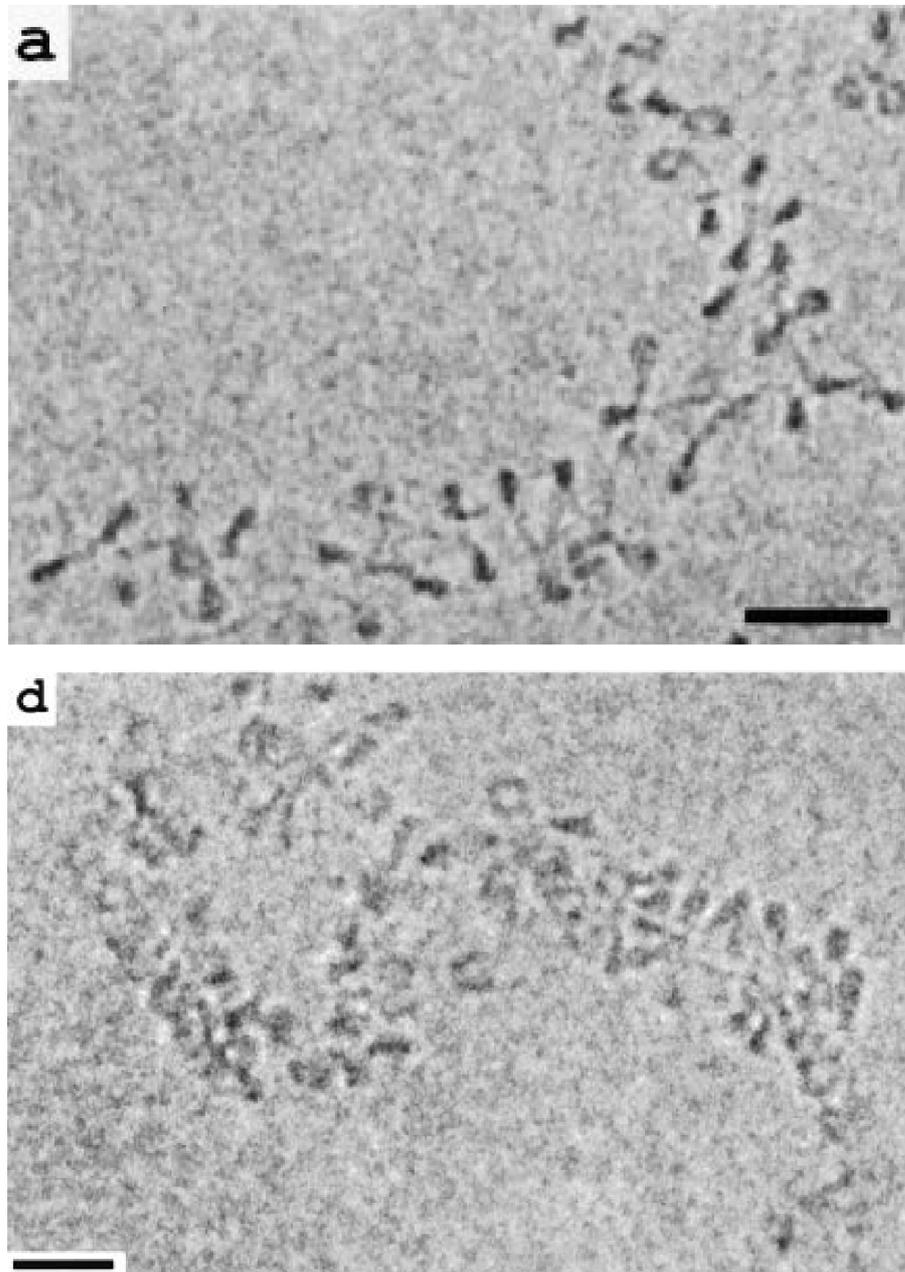
In Chromatin sind die Nukleosomen wie auf einer Perlenkette aufgereiht. Sie nehmen in vitro bei niedriger Salzkonzentration eine offene Struktur ein, die sich durch einen Durchmesser von ungefähr 10 nm auszeichnet (Abbildung 1.8 oben). Bei höherer Salzkonzentration kompaktiert die Struktur immer mehr, bis sich bei 80–100 mM NaCl eine fiberartige Struktur mit einem Durchmesser von 30 nm bildet (Abbildung 1.8 unten). Man bezeichnet diese als 30 nm-Fiber oder einfach als Chromatinfiber.

Das beschriebene Chromatin wird auch als Euchromatin bezeichnet. Im Gegensatz dazu gibt es auch Heterochromatin, das nach der Mitose<sup>3</sup> kondensiert bleibt und weitgehend transkriptionsinaktiv ist. Bei elektronenmikroskopischen Aufnahmen kann man hier ebenfalls 30 nm-Fibern nachweisen. Was den Unterschied auf der Ebene der Feinstruktur ausmacht, ist ungeklärt. In der vorliegenden Arbeit werden nur die Eigenschaften von Euchromatin untersucht.

---

<sup>2</sup>ARG und LYS tragen die Ladung 1, ASP und GLU tragen die Ladung -1, HIS trägt je nach Protonierungsgrad die Ladung 0 oder 1, bei  $p_H$  7 im Durchschnitt die Ladung 0,5.

<sup>3</sup>Mitose bezeichnet den Zustand der Zelle während der Zellteilung.



**Abbildung 1.8:** Struktur von Chromatinfibern aus Hühner-Erythrozyten in einer cryo-elektronenmikroskopischen Aufnahme [BHG<sup>+</sup>98]. Bei niedriger Salzkonzentration (5 mM NaCl) entsteht eine aufgefaltete Kette (oben), bei höherer Salzkonzentration (40 mM NaCl) eine kompaktierte Fiber. Die Balken entsprechen einer Länge von 30 nm.

Im folgenden wird ein Überblick über Eigenschaften der Chromatinfiber gegeben, die experimentell gewonnen wurden. Weiterhin werden die wichtigsten Chromatin-Modelle diskutiert.

### Länge der Linker-DNA

Die Länge der Linker-DNA ist bei verschiedenen Spezies und verschiedenen Zellen eines Organismus unterschiedlich. Sie besitzt häufig eine Länge von durchschnittlich 30 bp. Üblicherweise wird nicht die Länge der Linker-DNA angegeben, sondern die sogenannte Repeat-Länge. Sie gibt an, in welchen Abständen Nukleosomen auf der DNA angeordnet sind. Kurze Repeat-Längen, die auf eine Linker-DNA Länge deutlich unter 20 bp schließen lassen, kommen vor allem in besonders transkriptionsaktiven<sup>4</sup> Zellen vor, wie z.B. Cortex-Neuronen von Hasen mit einer Repeat-Länge von 162 bp ([TT77, Bro78] wie zitiert in [Hol89]). Es handelt sich hierbei um Chromatin, das sich in der Zusammensetzung von gewöhnlichem Chromatin unterscheidet. So ist nur eine geringe Menge vom Histon H1 vorhanden, Histon H2A ist zudem meist stark ubiquitiniert<sup>5</sup> und man findet dort große Mengen von HMG-Proteinen (siehe Abschnitt 1.2.4). Welchen Einfluß auf die Struktur diese Unterschiede zu gewöhnlichem Chromatin haben, ist Gegenstand der aktuellen Forschung. An sehr stark transkribierten Genen findet man keine Nukleosomen, man kann aber immunologisch Histone in diesen Bereichen nachweisen. Längere Linker-DNA findet man üblicherweise in nicht-aktivem Chromatin wie z.B. in Seeigel-Sperma (242 bp Repeat-Länge), an dem keine Transkription stattfindet. Hier sind allerdings die Histone häufig ersetzt durch Protamine<sup>6</sup>. Außerdem sind meist andere Proteine wie Spermin oder Spermidin vorhanden, die das Chromatin weiter kompaktieren.

### Massenbelegungsdichte

Ein wichtiger Strukturparameter der Chromatinfiber ist neben ihrem Durchmesser die Massenbelegungsdichte. Ihr Wert kann mit verschiedenen biophysikalischen Methoden bestimmt werden: Neutronen-, Röntgen- und Lichtstreuung, sowie analytische Ultrazentrifugation und Scanning-Transmissions-Elektronenmikroskopie (STEM). Man findet bei physiologischer Salzkonzentration Werte im Bereich von 6 Nukleosomen pro 11 nm Fa-

---

<sup>4</sup>Transkription ist der Prozeß im Zellkern, bei dem Information von der DNA abgelesen wird um anschließend anhand dieser Information Proteine zu synthetisieren.

<sup>5</sup>An ubiquitinierten Proteinen ist das Protein Ubiquitin gebunden.

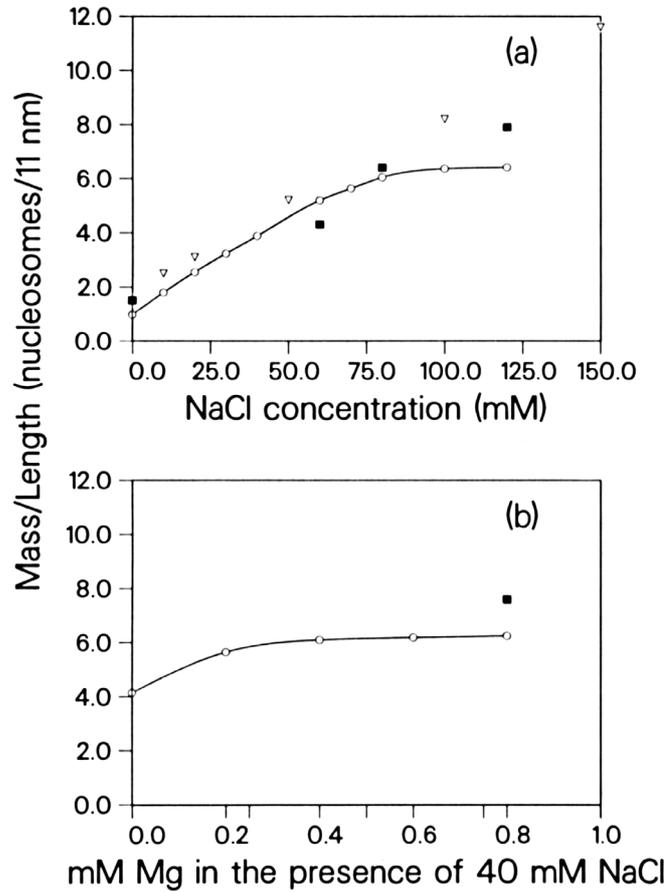
<sup>6</sup>Protamine sind Proteine, die an DNA in Sperma binden. Sie ersetzen die Histone und erlauben Chromosomen mehr zu kondensieren als dies mit Histonen möglich wäre.

denlänge. Messungen, die auf optischer Anisotropie beruhen, werden aufgrund der großen Unsicherheit dieser Methode nicht diskutiert (siehe Seite 14). Da die Ergebnisse vieler älterer Arbeiten aus heutiger Sicht mit Präparationsartefakten belastet sind (siehe unten), soll im folgenden ein Überblick über die vertrauenswürdigen Daten gegeben werden. Einen ausführlichen Überblick über die Arbeiten bis 1988 gibt das Lehrbuch von van Holde [Hol89].

Beim Isolationsvorgang von Chromatin kann die Struktur durch die Präparation verändert werden. Bei frühen Arbeiten wurden beispielsweise die langen Chromatinfibern durch hydrodynamische Scherungskräfte zerschnitten. Es wurde jedoch nachgewiesen, daß diese Scherungskräfte die Anordnung der Nukleosomen im Chromatin verändern. Bei neueren Arbeiten wird deswegen die Probe kurz mit Micrococcus-Nuclease behandelt und damit DNA teilweise verdaut. Dabei werden allerdings bevorzugt offene Strukturen verdaut. Es geschieht also eine künstliche Isolierung kompakterer Strukturen. Weiterhin werden häufig Reagenzien zur Stabilisierung von Chromatin eingesetzt, die vermutlich die Fiber weiter kondensieren. Besonders prominent ist beispielsweise in [GR87] die Verwendung von Spermin und Spermidin, positiv geladenen, drei- und vierwertigen ionischen Proteinen, die Ionenbrücken zwischen geladenen Teilen der Fiber bewirken.

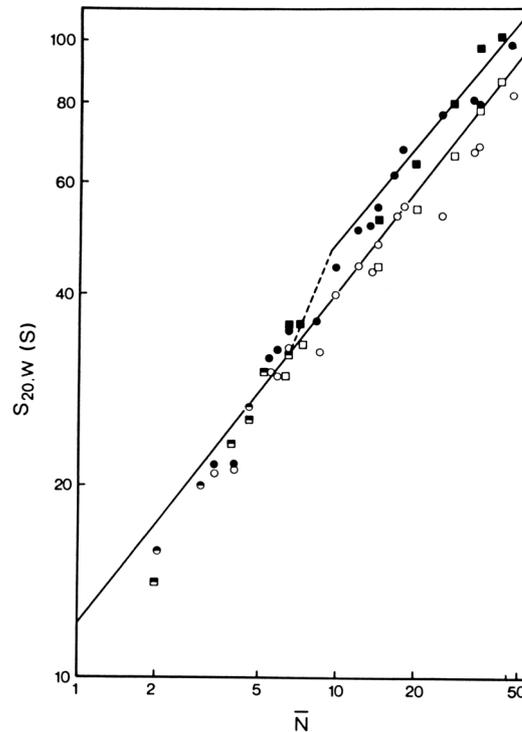
Eine sehr sorgfältige vergleichende Messung wurde von Gerchman und Ramakrishnan durchgeführt [GR87]. Fragmente von Chromatin aus Hühner-Erythrozyten mit ca. 100 Nukleosomen wurden bei verschiedenen NaCl- und MgCl<sub>2</sub>-Konzentrationen mit STEM und Neutronenstreuung vermessen. Die Ergebnisse sind in Abbildung 1.9 dargestellt. Übereinstimmend mit anderen STEM-Experimenten ergibt sich oberhalb einer Salzkonzentration von 80 mM NaCl bzw. 0.4 mM MgCl<sub>2</sub> eine Massenbelegungsichte von etwa 6 Nukleosomen pro 11 nm. Bei der Präparation wurde, wie bereits oben erwähnt, Spermin und Spermidin verwendet. Dies ist vermutlich auch der Grund, warum es anderen Arbeitsgruppen nicht gelungen ist, Chromatinfibern bei ähnlich hoher Ionenstärke noch zu beobachten. Sie erhalten bei diesen Ionenstärken Aggregate. Experimente mit Neutronenstreuung ohne die Verwendung von Spermin oder Spermidin ergaben bei 70 mM NaCl eine Massenbelegungsichte von lediglich 4.2 Nukleosomen pro 11 nm während Gerchman et al. einen Wert von etwa 5.5 angeben [SBB79].

Mit Lichtstreuung wurden Chromatinfibern mit 50–150 Nukleosomen vermessen [CCP78]. Aus den Gyrationradien kann man für eine Fiber von 30 nm-Durchmesser die folgenden Massenbelegungsdichten abschätzen: 3,1 Nukleosomen/11 nm bei 10 nM Tris, 4,4 Nukleosomen/11 nm bei 40 mM NaCl und 10 nM Tris, 5,5 Nukleosomen/11 nm bei 60 mM NaCl und 10 nM Tris. Tris ist wie Na<sup>+</sup> einfach positiv geladen. Diese Daten stimmen mit den Daten aus Neutronenstreuung und STEM überein.



**Abbildung 1.9:** Massenbelegungsichte von Chromatinfibern gemessen mit Neutronenstreuung ( $\circ$ ), und STEM ( $\blacksquare$ ,  $\nabla$ ) in Abhängigkeit von der Salzkonzentration. Aus [GR87]

Mit analytischer Ultrazentrifugation kann man die Sedimentationskoeffizienten von Chromatinfibern verschiedener Längen bestimmen (Abbildung 1.10). Ein Vergleich mit den Sedimentationskoeffizienten von starren Stäben ergibt eine Übereinstimmung mit einem Stabdurchmesser von  $22 \pm 7$  nm und einer Massenbelegungsichte von etwa 2,5 Nucleosomen/11 nm bei 60 mM NaCl [OKV90]. Der letztere Wert unterscheidet sich erheblich von den anderen hier dargestellten Ergebnissen. Der Grund für diese Abweichung ist vermutlich die Näherung als solider, starrer Stab.



**Abbildung 1.10:** Abhängigkeit des Sedimentationskoeffizienten  $S_{20,w}$  von der Anzahl der Nukleosomen  $\bar{N}$  von Nukleosomen–Oligomeren bei verschiedenen Ionenstärken. Die Oligomere wurden aus Rattenlebern ( $\circ$ ) und Hühnererythrozyten ( $\square$ ) isoliert. Daten der unteren Linie sind bei 25 mM NaCl (offene Symbole), der oberen Linie bei 65 mM (gefüllte Symbole) gemessen worden. Der Sprung bei  $\bar{N} \approx 6$  zeigt, daß mindestens 6 Nukleosomen notwendig sind, um bei höherer Ionenstärke ein Struktur auszubilden. Aus [But84].

### Orientierung der Nukleosomen und Linker–DNA innerhalb der Fiber

Die räumliche Orientierung der Nukleosomen und der Linker–DNA innerhalb der Chromatinfiber gibt Informationen über die Organisation der Fiber. Die Orientierungen werden mit Hilfe von Experimenten abgeschätzt, die den Effekt des linearen Dichroismus (LD) ausnutzen. Hierbei werden Chromatinfibern mit Hilfe von Strömungen des Lösungsmittels („Flow Linear Dichroism“ — FLD) oder einem starken elektrischen Feld („Electric Linear Dichroism“ — ELD) räumlich ausgerichtet. Man mißt die Absorption von Licht, das parallel ( $A_{\parallel}$ ) und senkrecht ( $A_{\perp}$ ) zur Ausrichtung der Moleküle polarisiert ist.

Man nennt  $LD^r = (A_{\parallel} - A_{\perp}) / A$  den reduzierten linearen Dichroismus, wobei  $A$  die Absorption von nicht polarisiertem Licht bezeichnet. Bei Chromatin wird bei Messungen im nahem UV (260 nm) das LD-Signal vom intensiven  $\pi \rightarrow \pi^*$ -Übergang der Basen der DNA dominiert. Die Absorption ist besonders groß, wenn das Licht parallel zur Ebene der Basen polarisiert ist, d.h. wenn die DNA senkrecht zur Polarisations Ebene des Lichts liegt. In der Literatur findet sich eine Vielzahl von Messungen, die von Dimitrov et al. in einem Überblicksartikel zusammengefaßt sind [DMP90]. Die Autoren weisen darauf hin, daß es sich bei negativen Werten von  $LD^r$  bei hoher Salzkonzentration vermutlich um Präparationsartefakte handelt. Beide Verfahren FLD und ELD liefern Ergebnisse, die bei niedriger und hoher Salzkonzentration von annähernd gleicher Größe sind, sich aber bei einer mittleren Konzentration unterscheiden. Der Unterschied wurde von Kubista et al. durch eine neue Interpretation der vorhandenen Daten geklärt [KHNN90]. Es wurde gezeigt, daß sich die Unterschiede erklären lassen, wenn man annimmt, daß bei mittlerer Salzkonzentration die Chromatin-Achse senkrecht zu dem angelegtem elektrischen Feld orientiert ist. Die nun konsistente Interpretierung der Daten erlaubt eine Abschätzung des möglichen Bereichs der Orientierung der Linker-DNA und der Nukleosomen: Man erhält für die Linker-DNA einen Bereich von  $50^\circ$  bis  $90^\circ$  und für die Achse der nukleosomalen DNA-Superhelix einen Bereich von  $45^\circ$  bis  $60^\circ$  Neigung zur Achse der Chromatinfiber.

Alternativ wurde auch die Doppelbrechung von Chromatinfibern untersucht [DMP90]. Die Interpretation dieser Daten ist jedoch sehr umstritten, so daß auf eine Diskussion verzichtet wird.

### Lage der Linker-Histone

Weitere Information über die innere Struktur der Chromatinfiber erhält man aus Experimenten, die die Position der Linker-Histone in der Fiber bestimmen. Durch verschiedene Experimente konnte nachgewiesen werden, daß die Linker-Histone innerhalb der Fiber und nicht in den Außenbereichen liegen.

Durch immunologische Reaktionen wurde von Leuba et al. gezeigt, daß H1 und H5 besonders geschützt in der Fiber liegen und nicht durch proteolytische Enzyme<sup>7</sup> hydrolysiert werden [LZH93]. Dieser Nachweis ist im Gegensatz zu vorhergehenden Arbeiten überzeugend, da die proteolytischen Enzyme an eine Oberfläche gebunden waren und dadurch sichergestellt werden konnte, daß nur Moleküle an der Oberfläche der Fiber angegriffen wurden und die Enzyme nicht in das Innere der Fiber diffundieren konnten.

Graziano et al. untersuchten mit Röntgen-Kleinwinkel-Streuung rekon-

---

<sup>7</sup>Ein proteolytisches Enzym hydrolysiert komplexe Proteine in kleinere Proteine.

stituiertes Chromatin mit normalem H1 und deuteriertem H1 [GGSR94]. Sie zeigten zunächst, daß die Streukurven des rekonstituierten Chromatins sich nicht von denen des nativen Chromatins unterscheiden. Bei einem bestimmten Grad der Deuterierung des Lösungsmittels wird nur die Streuung von H1 gemessen. Es ergibt sich ein Gyrationradius von 73 Å. Für die nicht-deuterierten Fiber wird der Gyrationradius von 122 Å bestimmt. Damit kann man abschätzen, daß H1 etwa 60–65 Å und der Mittelpunkt der Nukleosomen etwa 115 Å von der Fiberachse entfernt sind. H1 liegt also deutlich im Innern der Fiber.

### Flexibilität

Beobachtung von Chromatinfibern mit elektronenmikroskopischen und rasterkraftmikroskopischen Techniken zeigen, daß die Chromatinfiber nicht sehr steif ist (siehe z.B. Abbildung 1.8). Die Steifigkeit in Bezug auf die Biegung kann durch die Persistenzlänge charakterisiert werden (Kapitel 2.1).

Zur Bestimmung der Persistenzlänge der Chromatinfiber wurden zwei Wege beschritten: Die Bestimmung des räumlichen Abstands von Markern mit bekanntem genomischem Abstand<sup>8</sup> und die direkte Messung der End-zu-End-Abstands-Verteilung einer Chromatinfiber. Die Persistenzlänge kann dann aus der bekannten Verteilung des Porod-Kratky-Modells (Gleichung 2.14) oder des Flory-Modells<sup>9</sup> (z.B. in [DE86]) berechnet werden.

Bei Fluoreszenz-in situ-Hybridisierung (FISH) werden fluoreszierende Marker an definierten Stellen mit bekanntem genomischen Abstand in das Erbgut eingebracht. Es werden die projizierten geometrischen Abstände einer großen Anzahl dieser Marker mit Hilfe von Lichtmikroskopie bestimmt. Bei Längen unter etwa 1 Mbp ergibt sich ein linearer Zusammenhang zwischen dem Mittelwert des geometrischen Abstands und dem genomischen Abstand. Eine Auswertung der Daten von Engh et al. [EST92] durch Ostashevsky und Lange [OL94] auf Basis des Flory-Modells ergab eine Persistenzlänge von 100–140 nm, eine Auswertung auf Basis des Porod-Kratky-Modells durch Mehring ergab etwa 130 nm [Meh98]. Die Analyse der neueren Daten von Yokota et al. [YEH<sup>+</sup>95], bei denen die verwendeten Zellen wesentlich schonender präpariert wurden, ergeben eine Persistenzlänge von ungefähr 70 nm.

Ein direkte Bestimmung wurde von Castro im Labor von Bustamante durchgeführt [Cas94]. Die Arbeit ist nicht öffentlich verfügbar. Die Resultate

<sup>8</sup>Unter *genomischem Abstand* versteht man die Anzahl von Basenpaaren zwischen zwei Stellen einer DNA.

<sup>9</sup>Die Aussage des Flory-Modells ist, daß der Erwartungswert des End-zu-End-Abstands  $\vec{R}$  von  $N$  selbstvermeidenden Segmente der Länge  $b$  sich berechnen läßt wie  $\langle \vec{R}^2 \rangle = b^2 N^{2\nu}$ .  $\nu$  ist typischerweise etwa 3/5.

tate wurden aber in einer Arbeit von Houchmandzadeh et al. [HMCL97] zitiert; Details wurden von John Marko persönlich berichtet. Chromatin wurde aus Hühner-Erythrozyten isoliert und der End-zu-End-Abstand mit Rasterkraft-Mikroskopie bestimmt. Daraus wurde eine Persistenzlänge von 30-50 nm berechnet. Zur Beobachtung mit Rasterkraftmikroskopie muß die Fiber auf einem Glimmer immobilisiert werden. Es ist unklar, welchen Einfluß die bei der Immobilisierung auftretenden Kräfte auf die Struktur und damit auf die Persistenzlänge haben.

Zusammenfassend scheinen also Werte zwischen 50 nm und 150 nm mit den experimentellen Daten vereinbar zu sein. Weitere Daten werden von Arbeiten erwartet, bei denen einzelne Chromatinfibern mit Hilfe von Laser-Pinzetten gestreckt werden. Bei den diesen Daten scheint es sich aber um einen Flexibilitätsbereich zu handeln, bei dem die Chromatinfiber bereits so weit auseinandergezogen wurde, daß hauptsächlich die Flexibilität der DNA gemessen wurde [KBO98, Kat99].

### Strahlenschäden

Wenn Chromatin starker ionisierender Strahlung ausgesetzt werden, treten Brüche des DNA-Doppelstrangs auf. Aus der Verteilung der Fragmentlängen kann man auf die Struktur der Fiber zurückschließen. Rydberg et al. setzten menschliche Fibroblasten<sup>10</sup> ionisierender Strahlung aus und analysierten anschließend die Verteilung der Längen der DNA-Fragmente mit Gelen [RHMC98]. Diese Daten wurden mit simulierten Daten von verschiedenen geometrischen Modellen der Chromatinfiber verglichen. Es ergab sich, daß das Solenoid-Modell (siehe Abschnitt 1.2.3) nicht mit den Daten verträglich ist, wohl aber das Zick-Zack-Modell (siehe Abschnitt 1.2.3).

### Internukleosomaler Abstand

Die Abstände zwischen Nukleosomen, die direkt durch die Linker-DNA verbunden sind, wurde in einer Vielzahl von Experimenten bestimmt. Es wurden Di-, Trinukleosomen oder lange Chromatinfibern untersucht. Die eingesetzten experimentellen Techniken sind Lichtstreuung, Rasterkraftmikroskopie [ZLH98], analytische Ultrazentrifugation [YLW90, YLW91], Cryo-Elektronen-Mikroskopie [BHDW95] und elastische Neutronenstreuung [Ham99]. Die gemessenen Abstände stimmen bei geringen Salzkonzentrationen (0-20 mM NaCl) annähernd überein. Es ergeben sich 200-220 Å für Polynukleosomen aus Hühnererythrozyten und 180 Å bei Polynukleosomen aus COS-7-Zellen, was auf 64 bp bzw. 44 bp gestreckte Linker-DNA

---

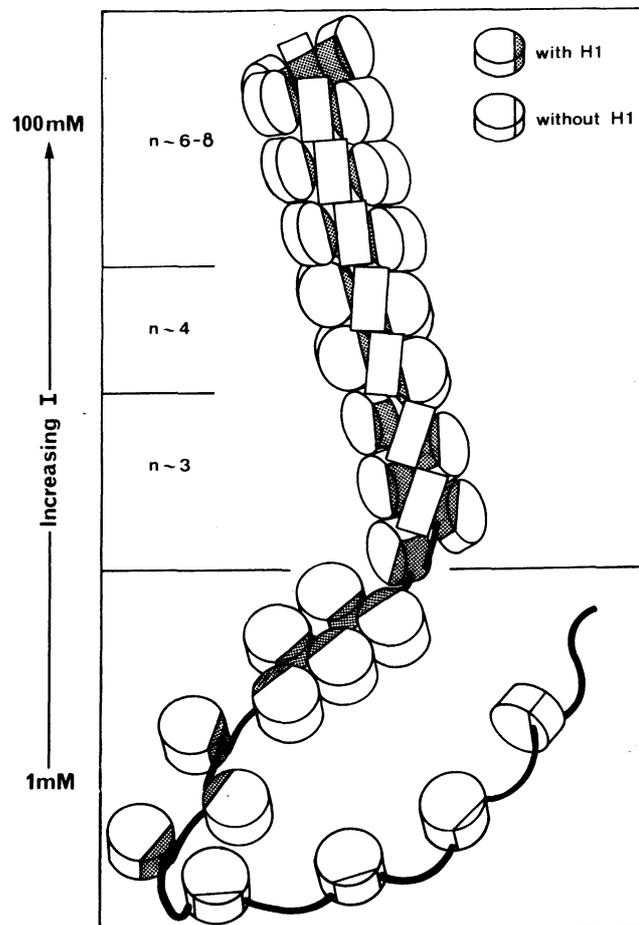
<sup>10</sup>Fibroblasten sind ein spezieller Zelltyp.

schließen läßt. Bei höherer Salzkonzentration sind aufgrund experimenteller Probleme bisher nur Messungen an Di- und Trinukleosomen möglich. Es ergibt sich ein Abstand von ungefähr 140 Å. Lediglich die Ergebnisse aus Cryo-Elektronen-Mikroskopie ergeben einen Wert von 300 Å, was aber im Vergleich mit den anderen Ergebnisse als Präparationsartefakt einzustufen ist. Der Grund für die Verringerung des internukleosomalen Abstands bei höherer Salzkonzentration ist unklar.

### 1.2.3 Modelle der Chromatinfiber

Zur Beschreibung der Struktur von Chromatin existiert ein Vielzahl von Modellen. Viele Modelle beruhen auf Ergebnissen, die sich später als Artefakt der Messung herausgestellt haben. Einen Überblick über Modelle bis 1988 gibt das Lehrbuch von van Holde [Hol89] oder der Review-Artikel von Dimitrov et al. [DMP90]. Das bisher wohl am weitesten akzeptierte Modell für die Chromatinfiber ist das Solenoid-Modell [FK76, TKK79] (Abbildung 1.11). Es beruht auf Ergebnissen, die mit Elektronenmikroskopie und Lichtstreuung gewonnen wurden. In diesem Modell bilden die Nukleosomen mit zunehmender Salzkonzentration eine Superhelix, wobei die Scheiben der Nukleosomen kaum gegen die Achse der Fiber geneigt sind. Eine Windung enthält 6 Nukleosomen und besitzt eine Ganghöhe von 11 nm. Der genaue Verlauf der Linker DNA zwischen den Nukleosomen wird nicht spezifiziert. Diese Struktur wird bereits von den Autoren dieses Modells als sehr schematisch bzw. idealisiert bezeichnet. Die häufig angenommene perfekte Regelmäßigkeit des Solenoids ist daher nicht gerechtfertigt. Daten über die Ausrichtung der Linker DNA und Nukleosomen sowie neuere Ergebnisse aus Strahlenschäden sind wie oben beschrieben nicht mit diesem Modell vereinbar. Zur Erklärung der Dichroismus-Daten wurden einige Varianten des Solenoid-Modells vorgeschlagen, die aber nicht befriedigend sind. Bei allen vorgeschlagenen Modellen ist die Linker-DNA sehr stark gebogen. Was diese Krümmung stabilisiert oder wie ein im Verhältnis zur Persistenzlänge so kurzes Stück DNA gebogen werden kann, bleibt ungeklärt.

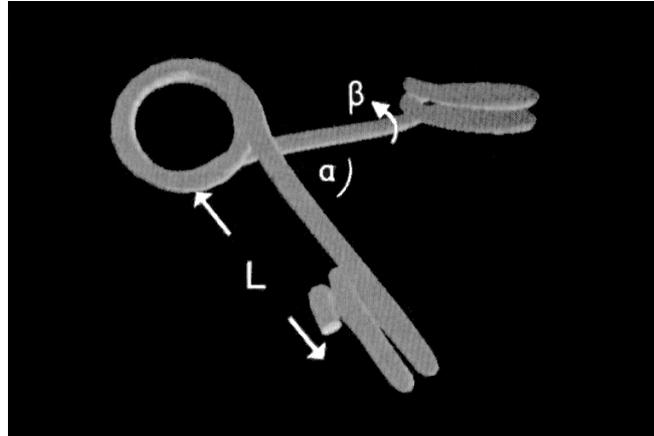
Ausgehend von einer Neuanalyse der bestehenden Daten und dem Vergleich mit neuen elektronenmikroskopischen Bildern entwarfen Woodcock et al. ein neues Modell der Chromatinfiber [WGHW93]. Dieses Modell stellt eine Weiterentwicklung der Crosslinker-Modelle von Bordas [BPGK<sup>+</sup>86] und Kubista [KHNN90] dar. Die Crosslinker-Modelle postulieren im wesentlichen eine Zick-Zack-Struktur. Die Linker-DNA läuft in etwa durch das Zentrum der Fiber. Woodcock führt im wesentlichen zwei Neuerungen dieser Modelle ein: Erstens findet eine mathematische Parametrisierung der Chromatinfiber statt. Zweitens wird eine im allgemeinen irreguläre Struktur angenommen,



**Abbildung 1.11:** Skizze des Solenoid-Modells aus [TKK79]. Die Nucleosomen (dargestellt als Zylinder) bilden bei zunehmender Salzkonzentration eine Helix-förmige Struktur.  $n$  bezeichnet die Anzahl der Nucleosomen pro Helixwindung.

da in elektronenmikroskopischen Bildern in der Regel geordnete und ungeordnete Bereiche zu beobachten sind.

Die Chromatinfiber kann durch drei Größen parametrisiert werden (Abbildung 1.12): Der Winkel  $\alpha$  beschreibt den Winkel zwischen einlaufender und auslaufender Linker-DNA am Nucleosom. Er ist experimentell gut bekannt (siehe Abschnitt 1.2.1). Ein weiterer Winkel  $\beta$  beschreibt die Rotation zwischen zwei aufeinanderfolgenden Nucleosomen.  $\beta$  konnte bisher nicht experimentell beobachtet werden. Der dritte Parameter ist die Länge der Linker-DNA.



**Abbildung 1.12:** Die Nukleosom-Linker-Nukleosom-Einheit:  $\alpha$ : Winkel zwischen einlaufender und auslaufender DNA.  $\beta$ : Verdrehungswinkel zwischen jeweils zwei aufeinanderfolgenden Nukleosomen.  $L$ : Länge des Linkers. (Aus [WGHW93])

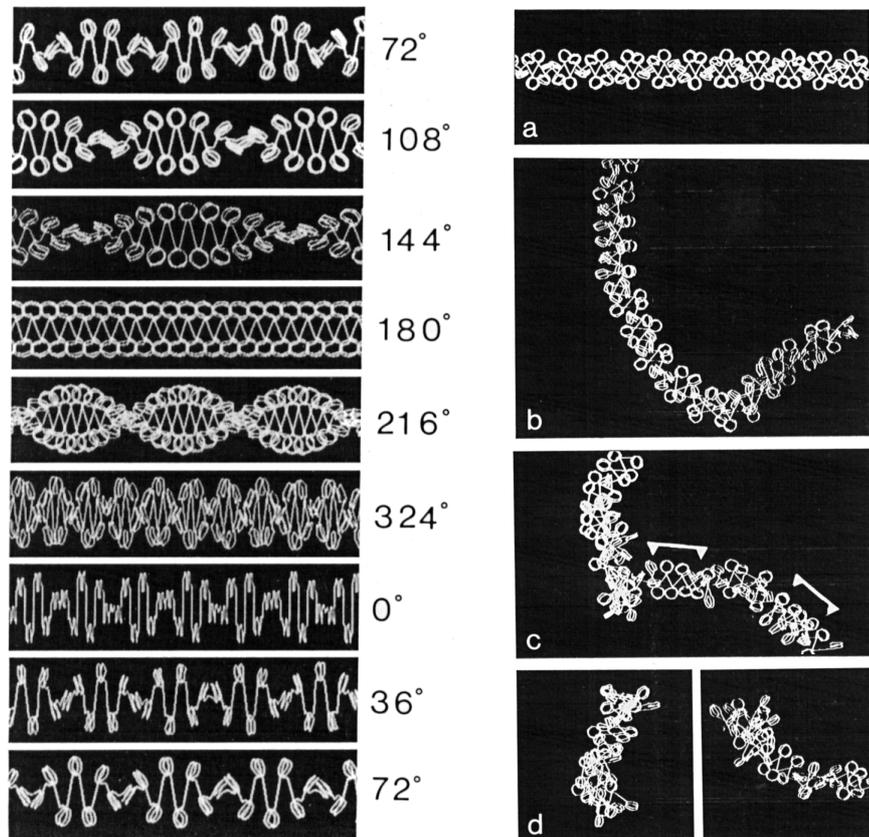
Bei konstanten Werten von  $\alpha$  und  $\beta$  erhält man reguläre Strukturen, die von der Größe der Parameter abhängen (Abbildung 1.13 links). Bei zufälliger Variation des Parameters  $\beta$  erhält man Strukturen, die umso unregelmäßiger aussehen, je größer die Variation von  $\beta$  ist (Abbildung 1.13 rechts).

Besonders einfache und suggestive Strukturen erhält man bei  $\beta = 0$  und  $\alpha = \frac{2\pi}{2n+1}$ ,  $n = 1, 2, 3, \dots$ . Der Fall  $n = 2$  (entsprechend 5 durch die Nukleosomen gebildeten Stränge) zeigt auch in etwa realistische Durchmesser und Massenbelegungsdichten. Eine Visualisierung dieser Anordnung ist in Abbildung 1.14 dargestellt.

In neueren Reviews [HZ95, HZ96, WH95] wird von verschiedenen Autoren darauf hingewiesen, daß in älteren Arbeiten häufig nur geordnete Bereiche der Fiber in elektronenmikroskopischen Bildern analysiert wurden und ungeordneten Bereiche als Präparationsartefakte dargestellt wurden. Die oben dargestellten Ergebnisse, die mit schonenderen Präparationsverfahren gewonnen wurden, zeigen jedoch, daß wohl auch die ungeordneten Bereiche keine Artefakte sind.

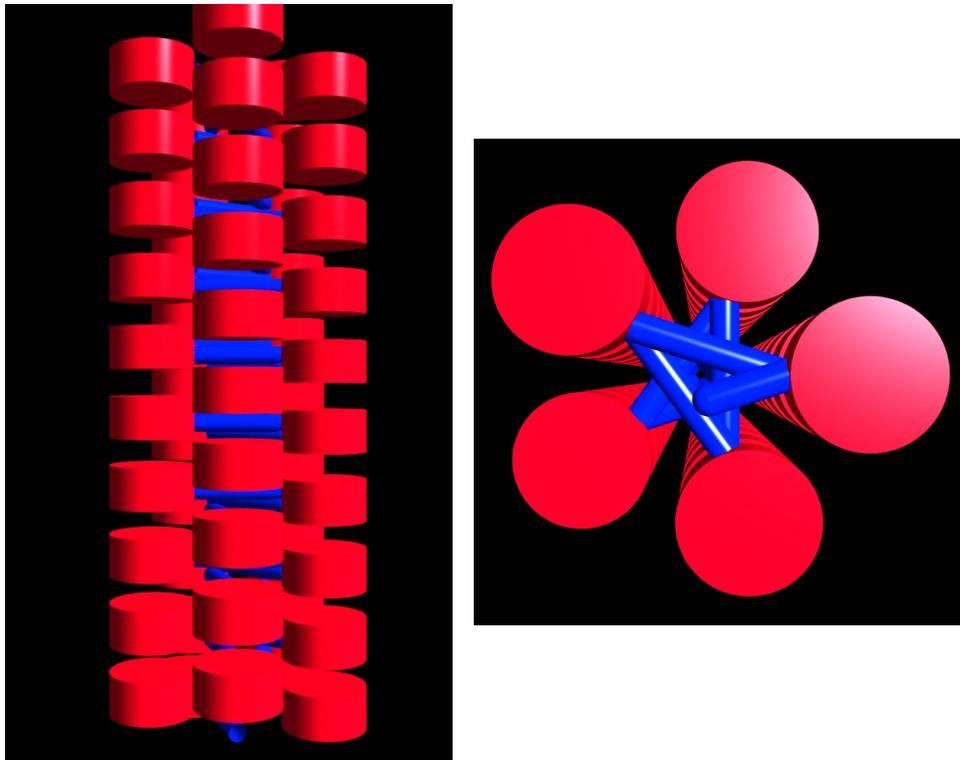
#### 1.2.4 Weitere chromosomale Proteine

Im Chromatin können neben Histonen weitere Proteine vorliegen, die Änderungen von Struktureigenschaften bewirken können. Viele Proteine wie z.B. transkriptionsregulierende Proteine scheinen nur kurzfristig an Chromatin zu



**Abbildung 1.13:** Strukturen der Chromatinfiber nach dem Zick-Zack-Modell von Woodcock et al. [WGHW93]. Linkes Bild: Strukturen bei  $\alpha=30^\circ$  und jeweils konstantem  $\beta$ . Rechtes Bild: Strukturen mit  $\alpha=60^\circ$  und  $\langle\beta\rangle=46^\circ$ .  $\beta$  schwankt um maximal  $0^\circ$  (a),  $36^\circ$  (b),  $72^\circ$  (c), und  $108^\circ$  (d)

binden, nehmen Modifikationen an Histonen oder DNA vor und lösen sich dann wieder vom Chromatin. Dauerhaft gebunden sind im wesentlichen nur Proteine der sogenannte *High Mobility Group* (HMG). Die wichtigsten Vertreter sind HMG 1, 2, 14 und 17. HMG 1 und 2 haben eine Masse von 29 kDa, HMG 14 und 17 von 10–12 kDa. Sie können bis zu 10% der Chromatin-Masse ausmachen. Sie sind dafür bekannt, Transkriptionstätigkeit regulieren zu können. Es wird vermutet, daß HMG-Proteine die Chromatin-Struktur verändern können. Funktion und Funktionsweise der HMG-Proteine sind nicht definitiv geklärt. Die HMG-Proteine sind nicht in allen Zellen präsent



**Abbildung 1.14:** Zwei verschiedene Ansichten der Struktur der Chromatinfiber nach dem Zick-Zack-Modell mit  $\alpha=37^\circ$ ,  $\beta=0^\circ$  und einer Linker-DNA Länge von 35 bp. Der Abstand zwischen einlaufender und auslaufender DNA an jedem Nukleosom ist 1,9 nm. (Für Definition dieses Abstands siehe Abbildung 3.4)

und sind damit im allgemeinen für die Faltung der Chromatinfiber auch nicht notwendig. Aus diesen Gründen wird der Einfluß der HMG-Proteine im folgenden nicht berücksichtigt.

# Kapitel 2

## Physikalische Grundlagen

In diesem Kapitel werden einige Grundlagen der Polymerphysik und der Flüssigkristalle dargestellt, soweit sie zum Verständnis dieser Arbeit notwendig sind. Weiterhin werden die Prinzipien der Monte Carlo-Methode und des zur Simulation benutzten Metropolis-Algorithmus behandelt.

### 2.1 Grundlagen der Polymerphysik

Dieser Abschnitt stellt die wesentlichen Grundbegriffe der Polymerphysik vor. Die Darstellung folgt im wesentlichen dem Lehrbuch von Doi und Edwards [DE86]. Alternative Herangehensweisen findet man in den Lehrbüchern von de Gennes [Gen79] oder Grossberg und Khoklov [GK94].

#### 2.1.1 Gaußsche Kette

Polymere bestehen aus langen Ketten von molekularen Einheiten. Das einfachste Modell eines Polymers ist die ideale Zufallskette (*freely jointed chain*). Sie besteht aus  $N$  starren Segmenten der Länge  $b_0$ , die in den Kopplungspunkte frei drehbar miteinander verbunden sind. Die Kette kann durch einen Satz von  $N + 1$  Ortsvektoren  $\{\vec{R}_n\} = (\vec{R}_0 \dots \vec{R}_N)$  oder alternativ durch einen Satz von  $N$  Verbindungsvektoren  $\{\vec{r}_n\} = (\vec{r}_1 \dots \vec{r}_N)$  beschrieben werden. Dabei gilt

$$\vec{r}_n = \vec{R}_n - \vec{R}_{n-1} \quad (2.1)$$

Die Verteilungsfunktion  $\Psi(\{\vec{r}_n\})$  gibt die Wahrscheinlichkeit einer Konfiguration der Kette an, die durch den Satz der Verbindungsvektoren  $\{\vec{r}_n\}$  beschrieben wird. Da alle Verbindungsvektoren  $\vec{r}_n$  voneinander unabhängig sind, kann  $\Psi(\{\vec{r}_n\})$  als Produkt der Verteilungsfunktion  $\psi(\vec{r}_n)$  der einzelnen

Segmente beschrieben werden:

$$\Psi(\{\vec{r}_n\}) = \prod_{n=1}^N \psi(\vec{r}_n) \quad (2.2)$$

Da jede Orientierung eines Segments gleich wahrscheinlich ist, entspricht die Verteilungsfunktion  $\psi(\vec{r})$  eines Segmentes der Verteilung eines zufälligen Vektors der konstanten Länge  $b_0$ :

$$\psi(\vec{r}) = \frac{1}{4\pi b_0^2} \delta(|\vec{r}| - b_0) \quad (2.3)$$

Diese Verteilung ist normiert:

$$\int d\vec{r} \psi(\vec{r}) = 1 \quad (2.4)$$

Der End-zu-End-Abstand  $\vec{R}$  der Kette ist wie folgt definiert:

$$\vec{R} = \vec{R}_N - \vec{R}_0 = \sum_{n=1}^N \vec{r}_n \quad (2.5)$$

Der Erwartungswert von  $\vec{r}_n$  bei Mittelung über alle Orientierungen ist 0:  $\langle \vec{r}_n \rangle = 0$ . Hieraus folgt  $\langle \vec{R} \rangle = 0$ . Der Erwartungswert des Quadrats des End-zu-End-Abstands  $\langle \vec{R}^2 \rangle$  hat einen von 0 verschiedenen Wert:

$$\langle \vec{R}^2 \rangle = \sum_{n,m=1}^N \langle \vec{r}_n \vec{r}_m \rangle = \sum_{n=1}^N \langle \vec{r}_n \rangle^2 + 2 \sum_{n>m}^N \langle \vec{r}_n \vec{r}_m \rangle = Nb_0^2 \quad (2.6)$$

Hierbei eintfallen die gemischten Glieder  $\langle \vec{r}_n \vec{r}_m \rangle = \langle \vec{r}_n \rangle \langle \vec{r}_m \rangle = 0$  für  $n \neq m$ , da  $\vec{r}_n$  und  $\vec{r}_m$  unkorreliert sind. Analog gilt für  $\langle \vec{R}^2 \vec{R}^2 \rangle$  (siehe Anhang A)

$$\langle \vec{R}^2 \vec{R}^2 \rangle = b_0^4 \left( \frac{3}{2} N^2 - \frac{N}{2} \right) \quad (2.7)$$

Bei einer Gaußschen Kette sind die Segmente nicht steif. Sie bestehen aus elastischen Federn, die Streck-Elastizität besitzen. Die Längenverteilung dieser Segmente genügt einer Gauß-Verteilung. Für die End-zu-End-Abstände ergeben sich die gleichen Ausdrücke wie für die ideale Zufallskette.

### 2.1.2 Kette mit Biegeelastizität

Um die Eigenschaften eines Polymers detaillierter zu beschreiben, wird das Polymer als kontinuierliche Fiber beschrieben. Im folgenden wird eine Fiber mit Biege-Elastizität und konstanter Länge betrachtet. Sei  $\vec{R}(s)$  die Position eines Punktes auf der Fiber der Länge  $L$  an dem Punkt mit der Konturlänge  $s$ . Dann ist

$$\vec{u}(s) = \frac{\partial \vec{R}}{\partial s} \quad (2.8)$$

der normierte Tangentenvektor an die Kette. Die Biegeenergie ist quadratisch in  $\partial \vec{u}/\partial s$ :

$$U_{bend} = \frac{1}{2} E_b \int_0^L ds \left( \frac{\partial \vec{u}}{\partial s} \right)^2 \quad (2.9)$$

Da die Fiber mit ihrer Umwelt in Wärmekontakt steht, ist die Verteilungsfunktion  $\Psi[\vec{u}]$  durch die Boltzmann-Verteilung gegeben:

$$\Psi[\vec{u}] \sim \exp\left(-\frac{U_{bend}}{k_B T}\right) = \exp\left[-\frac{1}{4\lambda} \int_0^L ds \left(\frac{\partial \vec{u}}{\partial s}\right)^2\right] \quad (2.10)$$

mit

$$\lambda = \frac{k_B T}{2E_b} \quad (2.11)$$

Man nennt  $\tilde{l}_b = (2\lambda)^{-1} = E_b/(k_B T)$  die Persistenzlänge. Dieses Modell wird auch Porod-Kratky-Modell genannt. Aus Gleichung 2.10 kann man für kleine  $s$  ableiten [DE86], daß

$$\langle (\vec{u}(s' + s) - \vec{u}(s'))^2 \rangle = 4\lambda s, \quad (2.12)$$

woraus wiederum folgt, daß

$$\langle \vec{u}(s' + s) \vec{u}(s') \rangle = \exp(-2\lambda s) = \exp(-s/\tilde{l}_b). \quad (2.13)$$

Die Korrelation zwischen den Orientierungen an zwei Punkten der Fiber zerfällt also exponentiell mit ihrem Abstand. Man betrachtet i.a. einen Abschnitt mit der Länge der Persistenzlänge  $\tilde{l}_b$  als steif.

Der End-zu-End-Abstand berechnet sich zu

$$\begin{aligned} \langle \vec{R}^2 \rangle &= \left\langle \left( \vec{R}(L) - \vec{R}(0) \right)^2 \right\rangle = \int_0^L ds \int_0^L ds' \langle \vec{u}(s) \vec{u}(s') \rangle \\ &= 2 \int_0^L ds \int_0^s ds' \exp(-2\lambda(s - s')) \\ &= \frac{L}{\lambda} - \frac{1}{2\lambda^2} [1 - \exp(-2\lambda L)]. \end{aligned} \quad (2.14)$$

Für  $L\lambda \gg 1$  bzw.  $L \gg \tilde{l}_b$  ergibt sich

$$\langle \vec{R}^2 \rangle = L/\lambda = 2L\tilde{l}_b \quad (2.15)$$

Dies ist das gleiche Ergebnisse wie für eine Gaußsche Kette der Länge  $L$  und Gliederlänge  $2\tilde{l}_b$ . Eine lange Kette mit Biegeelastizität verhält sich also wie eine Gaußsche Kette mit Gliederlänge  $2\tilde{l}_b$ . Man bezeichnet  $2\tilde{l}_b$  auch als Kuhn-Länge.

### 2.1.3 Kette mit Torsionselastizität

Polymerketten können Deformationsenergie bei Verdrehung aufnehmen. Die Gesamtverdrehung einer Kette ist die Summe der Verdrehungen der einzelnen Glieder. Man beschreibt sie als Twist  $Tw$ :

$$Tw = \frac{1}{2\pi} \int_0^L \varphi(s) ds \quad (2.16)$$

Hierbei bezeichnet  $\varphi(s)$  die Verdrehung pro Einheitslänge an der Stelle mit der Bogenlänge  $s$ . Unter Verwendung des Vektors  $\vec{a}(s)$  senkrecht zur Fibrachse bzw. senkrecht zu  $\vec{u}(s)$ , kann Gleichung 2.16 wie folgt umgeformt werden:

$$Tw = \frac{1}{2\pi} \int_0^L ds \vec{a}'(s) [\vec{u}'(s) \times \vec{a}(s)] \quad (2.17)$$

mit den Bezeichnungen  $\vec{u}' := d\vec{u}/ds$  und  $\vec{a}' := d\vec{a}/ds$ . Die Deformationsenergie der Fiber bei Torsion ergibt sich zu

$$U_{tors} = (E_t/L) 2\pi Tw \quad (2.18)$$

mit  $E_t$  als Torsionsmodul der Kette.

Mit diesem Ausdruck für  $U_{tors}$  können die gleichen Schritte durchgeführt werden, die zu den Gleichungen 2.10–2.13 des vorangehenden Abschnitt führen. Es ergibt sich

$$\langle \phi(s' + s)\phi(s') \rangle = \exp(-s/\tilde{l}_t) \quad (2.19)$$

mit der Persistenzlänge

$$\tilde{l}_t = \frac{E_t}{k_B T}. \quad (2.20)$$

## 2.2 Flüssigkristalle

Nukleosomen können bei genügend hoher Konzentration in Lösung Flüssigkristalle bilden [LL97]. Aus den Eigenschaften dieser Flüssigkristalle können

Eigenschaften der Wechselwirkungen zwischen Nukleosomen abgeschätzt werden. Deswegen wird in diesem Abschnitt eine kurze Einführung in die Physik der Flüssigkristalle gegeben. Für eine detailliertere Darstellung empfiehlt sich der einführende Lehrbuchartikel von Heppke und Bahr [HB92] oder das Lehrbuch von de Gennes und Prost [GP93].

In einem Kristall sind die Bestandteile (Moleküle oder Molekülgruppen) regulär angeordnet und besitzen eine Fernordnung. In einer Flüssigkeit besitzen die Bestandteile lediglich eine Nahordnung. Zwischen diesen zwei Phasen existieren Zwischenzustände, die Flüssigkristalle oder Mesophasen<sup>1</sup> genannt werden. Sie besitzen wie Flüssigkeiten einen großen Grad an Fluidität, d.h. sie nehmen in kürzester Zeit die Form des sie beinhaltenden Gefäßes an. Andererseits sind viele ihrer physikalischen Eigenschaften richtungsabhängig. Beispielsweise kann wie bei Kristallen optische Anisotropie wie z.B. Doppelbrechung auftreten. Je nach Art der Anisotropie unterscheidet man im wesentlichen zwischen den folgenden Phasen:

**nematisch** Die nematische Phase unterscheidet sich von der isotropen Phase durch eine bevorzugte Parallelorientierung der Molekülachsen. Die Molekülschwerpunkte weisen wie in der isotropen Phase keine Positionsfornordnung auf.

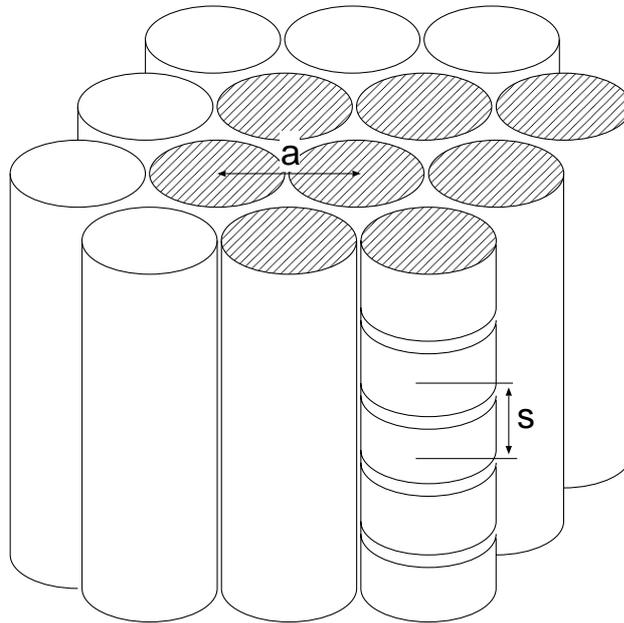
**smektisch** In smektischen Phasen tritt zusätzlich ein Positionsfornordnung der Molekülschwerpunkte auf. Es entsteht eine Schichtstruktur. Man kann die Schichten als zweidimensionale Flüssigkeiten ansehen, die sich in einem wohldefinierten Abstand zueinander befinden.

**columnar** Columnare Phase besitzen eine Positionsfornordnung der Moleküle in zwei Dimensionen. Die Moleküle sind in einer Säulenstruktur angeordnet. Man kann das System als zweidimensionale Anordnung von Flüssigkeitsröhren betrachten.

Smektische Flüssigkristalle bestehen meist aus gestreckten Objekten, während columnare Phasen meist aus scheibenförmigen Teilchen bestehen. Der häufig benutzte Begriff diskotisch für columnare Phasen bezieht sich auf die Form der Bestandteile. In hexagonal columnaren Phasen sind die Säulen hexagonal angeordnet (Abbildung 2.1). Man bezeichnet diese Phase auch mit  $D_h$ . Weiterhin existieren rektangulare Phasen, bei denen die zweidimensionale Einheitszelle der Säulen rechteckig ist. Man bezeichnet sie mit  $D_r$ .

---

<sup>1</sup>griechisch μέσος — in der Mitte gelegen



**Abbildung 2.1:** Skizze einer diskotisch columnar hexagonalen Flüssigkristallphase.  $a$  bezeichnet den mittleren Abstand der benachbarten Flüssigkeitsröhren,  $s$  den mittleren Abstand zwischen Teilchen einer Flüssigkeitsröhre.

## 2.3 Monte Carlo–Verfahren

Zur Berechnung der Eigenschaften von Chromatinfibern wurde ein Monte Carlo–Verfahren eingesetzt<sup>2</sup>. Mit Hilfe von Monte Carlo–Rechnungen lassen sich Erwartungswerte eines physikalischen Systems im thermodynamischen Gleichgewicht bestimmen. Man unterscheidet zwischen statischen und dynamischen Monte Carlo–Algorithmen [Sok96a]. Statische Methoden generieren statistisch unabhängige Konfigurationen. Manche Algorithmen zur Erzeugung einer Gesamtheit von selbstvermeidenden Zufallswegen gehören zu dieser Klasse. Bei dynamischen Methoden sind aufeinanderfolgende Konfigurationen nicht statistisch unabhängig. Ausgehend von einer beliebigen

<sup>2</sup>In der Theoretischen Informatik werden Algorithmen, die auf Zufälligkeit basieren („Randomisierte Algorithmen“) und deren Ergebnis nicht notwendigerweise exakt, sondern mit hoher Wahrscheinlichkeit nahe dem exakten Ergebnis liegt, als Monte Carlo–Algorithmen bezeichnet ([MR95]). In der Theoretischen Physik hingegen ist es üblich, als Monte Carlo–Algorithmus nur die Berechnung repräsentativer Stichproben eines Ensembles von physikalischen Zuständen zu bezeichnen, was aber auch im Sinne der Informatik ein Monte Carlo–Algorithmus ist.

Startkonfiguration werden durch einen stochastischen Prozeß neue Zustände generiert. Der Metropolis–Algorithmus stellt eine solche dynamische Methode dar.

Im folgenden wird zunächst die Mastergleichung dargestellt. Darauf aufbauend wird der in dieser Arbeit verwendete Metropolis–Algorithmus begründet. Weiterführende Diskussionen finden sich in der Literatur [AT87, AT93, BH88].

### 2.3.1 Mastergleichung

Die Zeitentwicklung eines Systems in den Zuständen  $r$  mit den Wahrscheinlichkeiten  $P_r$  läßt sich durch die Mastergleichung beschreiben (siehe z.B. [RF94, Rei65]):

$$\frac{dP_r}{dt} = \sum_s P_s W_{sr} - \sum_s P_r W_{rs} \quad (2.21)$$

$W_{rs}$  ist dabei die Übergangswahrscheinlichkeit von Zustand  $r$  nach Zustand  $s$  pro Zeiteinheit.

Das System ist im Gleichgewicht, wenn  $dP_r/dt = 0$ . Diese Bedingung ist erfüllt, wenn die Bedingung des detaillierten Gleichgewichts erfüllt ist:

$$P_{s,gl} W_{sr} = P_{r,gl} W_{rs} \quad (2.22)$$

Weiterhin soll das System im thermischen Gleichgewicht sein. Dann gilt  $P_r \sim e^{-\beta E_r}$ . Daraus folgt

$$W_{sr} = e^{\beta(E_s - E_r)} W_{rs}. \quad (2.23)$$

### 2.3.2 Simple Sampling

Für ein System, das durch eine Hamiltonfunktion (bzw. den Hamiltonoperator)  $H(\mathbf{x})$  im Phasenraum  $\mathbf{x}$  beschrieben werden kann, errechnet sich der Erwartungswert einer Observable  $A$  wie folgt:

$$\langle A \rangle = \frac{1}{Z} \int e^{-\frac{H(\mathbf{x})}{k_B T}} A(\mathbf{x}) d\mathbf{x} = \int p(\mathbf{x}) A(\mathbf{x}) d\mathbf{x} \quad (2.24)$$

$$Z = \int e^{-\frac{H(\mathbf{x})}{k_B T}} d\mathbf{x} \quad (2.25)$$

Der normalisierte Boltzmann–Faktor

$$p(\mathbf{x}) = \frac{1}{Z} e^{-\frac{H(\mathbf{x})}{k_B T}} \quad (2.26)$$

gibt die Wahrscheinlichkeit an, mit welcher der Zustand  $\mathbf{x}$  im thermodynamischen Gleichgewicht vorkommen kann.

Beim *Simple Sampling* werden  $N$  Konfigurationen  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$  zufällig aus dem Phasenraum  $\mathbf{x}$  ausgewählt. Die Erwartungswerte werden durch eine approximierete Form von Gleichung (2.24) berechnet:

$$\langle A \rangle \approx \frac{1}{G} \sum_{i=1}^N e^{-\beta H(\mathbf{x}_i)} A(\mathbf{x}_i) \quad (2.27)$$

$$G = \sum_{i=1}^N e^{-\beta H(\mathbf{x}_i)} \quad (2.28)$$

Der große Nachteil dieser Methode liegt darin, daß alle Zustände mit gleicher Wahrscheinlichkeit ausgewählt werden, also auch solche, die aufgrund ihrer Gewichtung mit  $p(\mathbf{x})$  nur einen geringen Einfluß auf  $\langle A \rangle$  haben.

### 2.3.3 Importance Sampling — Metropolis Algorithmus

Generiert man Zustände  $\mathbf{x}$  durch einen Monte Carlo-Algorithmus, so daß sie mit der normierten Wahrscheinlichkeitsdichte  $P(\mathbf{x})$  erzeugt werden, so muß (2.27) durch

$$\langle A \rangle \approx \frac{1}{N} \sum_{i=1}^N \frac{p(\mathbf{x}_i)}{P(\mathbf{x}_i)} A(\mathbf{x}_i) \quad (2.29)$$

ersetzt werden. Bei einer Wahl von  $P(\mathbf{x}) \propto e^{-\frac{H(\mathbf{x})}{k_B T}}$  entsprechen die Erwartungswerte der Observablen den arithmetischen Mittelwerten und (2.29) läßt sich zu

$$\langle A \rangle \approx \frac{1}{N} \sum_{i=1}^N A(\mathbf{x}_i) \quad (2.30)$$

vereinfachen. Ein solcher Monte Carlo-Algorithmus wird *Importance Sampling* genannt, da die Zustände gemäß ihrer Wahrscheinlichkeit erzeugt werden.

Die Idee von Metropolis et al. [MRR<sup>+</sup>53] besteht darin, den Zustand  $\mathbf{x}_{i+1}$  aus dem Zustand  $\mathbf{x}_i$  mit der Übergangswahrscheinlichkeit  $W_{\mathbf{x}_i \rightarrow \mathbf{x}_{i+1}}$  zu konstruieren. Ausgehend von einem Ausgangszustand werden gemäß dieser Iterationsregel neue Zustände erzeugt. Dies kann als eine künstliche, in der Zeit diskrete Dynamik gemäß Abschnitt 2.3.1 angesehen werden. Wie dort gezeigt wurde, gehorchen die erzeugten Zustände einer Verteilung bei der

Temperatur  $T$ , wenn das detaillierte Gleichgewicht (Gleichung 2.22) erfüllt ist und gilt

$$\frac{W_{s \rightarrow s'}}{W_{s' \rightarrow s}} = e^{-\beta \Delta H} \quad (2.31)$$

Man wählt

$$\begin{aligned} W_{s \rightarrow s'} &= 1 && \text{für } \Delta H \leq 0 \\ W_{s \rightarrow s'} &= e^{-\beta \Delta H} && \text{für } \Delta H > 0 \end{aligned} \quad (2.32)$$

Der Metropolis–Algorithmus läßt sich wie folgt formulieren:

1. Ein beliebiger Startzustand  $\mathbf{x}_0$  wird ausgewählt.
2. Nach dem Zufallsprinzip wird ein neuer Zustand ausgewählt (Monte Carlo–Schritt).
3. Ist die Energiedifferenz  $\Delta H$  zwischen dem alten und dem neuen Zustand größer 0, so wird der neue Zustand mit der Wahrscheinlichkeit  $e^{-\frac{\Delta H}{k_B T}}$  akzeptiert. Dazu wird eine Zufallszahl  $\xi$  aus dem Intervall  $[0, 1]$  generiert und mit  $e^{-\frac{\Delta H}{k_B T}}$  verglichen. Im Falle von  $\xi < e^{-\frac{\Delta H}{k_B T}}$  wird der neue Zustand angenommen, ansonsten der alte Zustand wiederhergestellt. Wenn  $\Delta H < 0$  ist, wird der neue Zustand immer angenommen. Das bedeutet, daß ein neu generierter Zustand mit der Wahrscheinlichkeit  $\min(1, e^{-\frac{\Delta H}{k_B T}})$  akzeptiert wird.
4. Die Schritte 2–3 werden so lange wiederholt, bis eine ausreichende Anzahl von Zuständen vorliegt. Anschließend können nach Gleichung (2.30) die Erwartungswerte berechnet werden.

Bei der Wahl der Monte Carlo–Schritte ist darauf zu achten, daß alle Konfigurationen mit gleicher Wahrscheinlichkeit erzeugt werden, da sonst eine weiterer Faktor zur Übergangswahrscheinlichkeit  $W_{s \rightarrow s'}$  hinzukommen würde. Dies ist gleichbedeutend mit der Bedingung, daß die Wahrscheinlichkeit, daß aus dem Zustand  $s$  der Zustand  $s'$  erzeugt wird, genauso groß ist, wie die Wahrscheinlichkeit für den umgekehrten Vorgang. Weiterhin ist darauf zu achten, daß der Algorithmus ergodisch ist. Dies bedeutet, daß die Wahrscheinlichkeit, einen beliebigen Zustand von einem anderen beliebigen Zustand in endlich vielen Schritten zu erreichen, ungleich Null ist.



# Kapitel 3

## Modell für DNA und Chromatin

In diesem Kapitel wird ein Computer-Modell der Chromatinfiber vorgestellt, mit dem es möglich ist, Eigenschaften der Fiber quantitativ zu reproduzieren und vorherzusagen. Die Berechnung der Chromatinfiber ist mit herkömmlichen Methoden unmöglich, da die Fiber eine viel größere Anzahl von Atomen besitzt als Moleküle, deren Eigenschaften üblicherweise berechnet werden<sup>1</sup>. Ausgehend von Modellen für DNA [CL94, CL96, Chi96, KFKL95, KML98] und Polynukleosomen [Ehr96, EMCL97] wurde deswegen ein Modell entwickelt, das auf die Beschreibung der atomaren Details verzichtet. DNA und Chromatosomen werden durch Einheiten approximiert, die aus vielen Atomen bestehen. Die Wechselwirkungen zwischen DNA und Nukleosomen werden durch effektive Potentiale beschrieben, unter die die interatomaren Potentiale subsumiert werden. Mit einem Monte Carlo-Verfahren werden Ensembles dieses Modells bei einer bestimmten Temperatur berechnet.

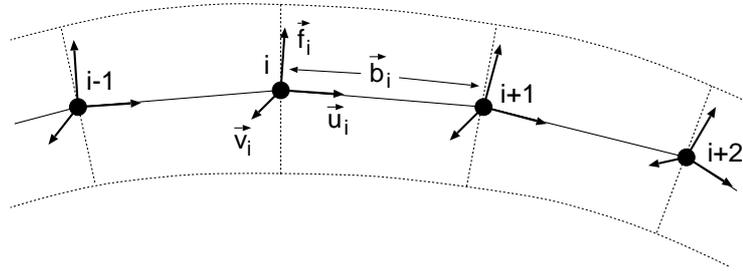
---

<sup>1</sup>Ein einzelnes Nukleosom besteht aus etwa 20000 Atomen; eine Polynukleosom mit 50 Nukleosomen besitzt also etwa 1 Million Atome. Das umgebende Wasser hat einen großen Einfluß auf die Struktur der Fiber. Um diesen im Modell angemessen zu berücksichtigen, wäre es notwendig, mindestens die gleiche Anzahl von Wassermolekülen zu dem zu berechnenden System hinzuzunehmen. Die Berechnungen von Systemen mit 2 Millionen Atomen sind mit heutigen Mitteln nicht möglich. Die Grenze liegt für Berechnungen mit Methoden der Molekulardynamik bei ca. 20000 Atomen.

## 3.1 Diskretisierung

### 3.1.1 Diskretisierung der DNA

Die DNA-Doppelhelix wird durch Segmente beschrieben. Die Segmentlänge wird deutlich kürzer als die Persistenzlänge der DNA bezüglich der Biegung gewählt, um ein einzelnes Segment vereinfacht als gerade beschreiben zu können.



**Abbildung 3.1:** Zur Diskretisierung der DNA

Die geometrische Lage eines DNA-Segments wird durch einen Ortsvektor und ein lokales Koordinatensystem  $(\vec{f}_i, \vec{u}_i, \vec{v}_i)$ ;  $\vec{v}_i = \vec{f}_i \times \vec{u}_i$  definiert, welches die Orientierung des Segments im Raum beschreibt (Abbildung 3.1). Der Segmentvektor  $\vec{u}_i$  ist der normierte Verbindungsvektor zwischen  $\vec{r}_i$  und  $\vec{r}_{i+1}$ . Man bezeichnet den Abstand  $b_i := |\vec{r}_{i+1} - \vec{r}_i|$  als Segmentlänge.  $\vec{f}_i$  bezeichnet einen normierten Vektor, der senkrecht auf  $\vec{u}_i$  steht. Die Winkel  $\alpha_i$ ,  $\beta_i$  und  $\gamma_i$  sind die Eulerwinkel der Transformation des Koordinatensystems am Segment  $i$  in das Koordinatensystem am Segment  $i+1$ .  $\beta_i$  gibt also den Biegewinkel zwischen den zwei Segmenten an, die Summe von  $\alpha_i$  und  $\gamma_i$  den Torsionswinkel. Falls zwischen zwei Segmenten eine Biegung vorhanden ist, gibt der normierte Vektor  $\vec{B}_i$  den Wert von  $\vec{u}_{i+1}$  im elastischen Gleichgewichts bezüglich dem an das Segment  $i$  angehefteten Koordinatensystems an (Abbildung 3.2). Man definiert  $\vec{B}_i$  über die Winkel  $\vartheta_i$  und  $\varphi_i$ :

$$\vec{B}_i = \vec{f}_i \sin \vartheta_i \cos \varphi_i + \vec{v}_i \sin \vartheta_i \sin \varphi_i + \vec{u}_i \cos \vartheta_i \quad (3.1)$$

Im Falle keiner Verbiegung ist  $\vartheta_i = 0$  und  $\varphi_i = 0$  und daher  $\vec{B}_i = \vec{u}_i$ .

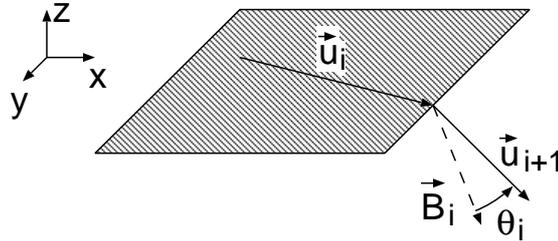


Abbildung 3.2: Zur Definition des Biegungsvektors  $\vec{B}_i$

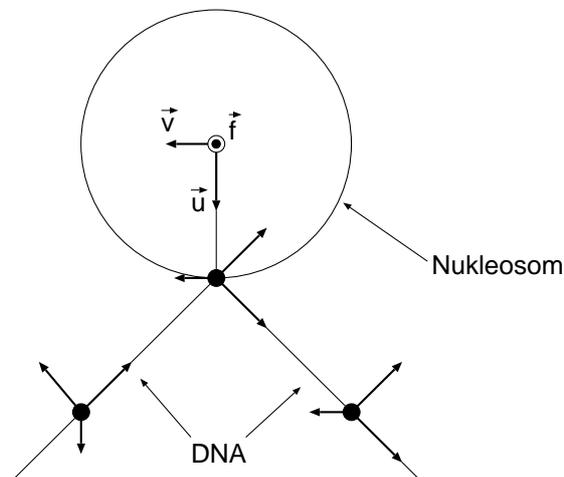


Abbildung 3.3: Zur Diskretisierung des Nukleosoms. Der Richtungsvektor  $\vec{f}$  des Nukleosoms steht senkrecht zur Zeichenebene.

### 3.1.2 Diskretisierung der Nukleosomen

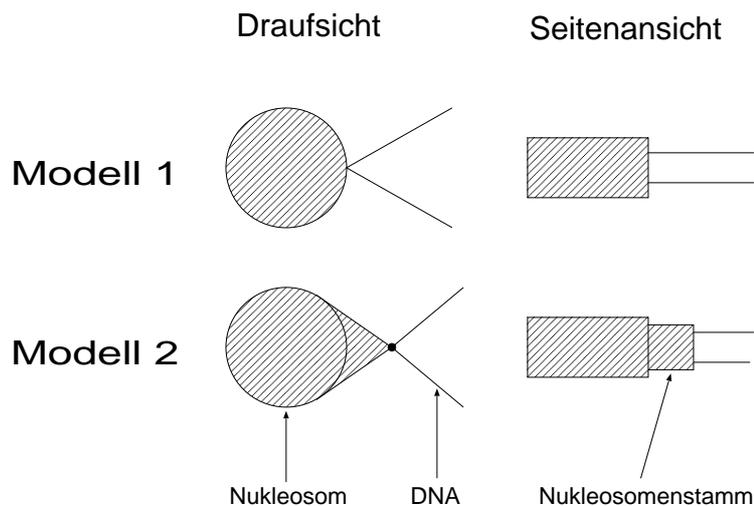
Die Lage eines Nukleosoms<sup>2</sup> wird durch einen Vektor beschrieben, der die Position des Schwerpunktes angibt. Wie im Fall der DNA wird die Ausrichtung durch ein lokales Koordinatensystem beschrieben.  $\vec{u}_i$  zeigt in Richtung des geometrischen Mittelpunkts der Ankopplungspunkte der DNA-Segmente,  $\vec{v}_i$  ist parallel zur Achse des Nukleosoms (Abbildung 3.3).

Für die Kopplung der DNA und dem Nukleosom wurden zwei verschiedene Modelle benutzt (Abbildung 3.4):

<sup>2</sup>Genaugenommen handelt es sich hier um das *Chromosom*.

**Modell 1** Kopplung an zwei distinkten Punkten des Nukleosoms im Abstand der einlaufenden und auslaufenden DNA

**Modell 2** Kopplung an zwei distinkten Punkten im Abstand der einlaufenden und auslaufenden DNA in definiertem Abstand vom Nukleosom (am Nukleosomenstamm)

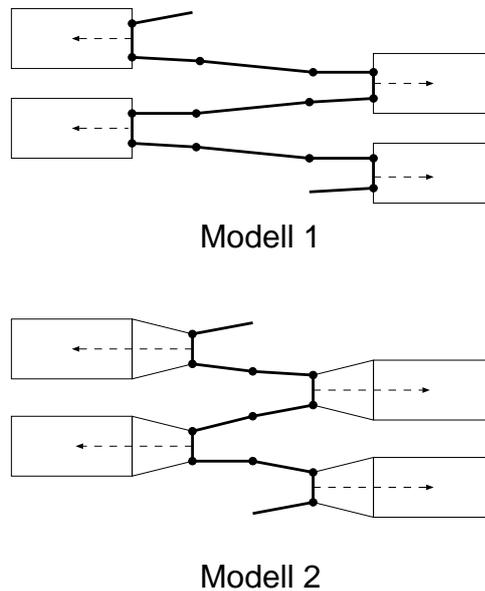


**Abbildung 3.4:** Modelle der Kopplung von DNA und Nukleosom. Kopplung an 1) distinkten Punkten des Nukleosoms, 2) distinkten Punkten außerhalb des Nukleosoms (am Nukleosomenstamm).

Die Grundeinheiten DNA und Nukleosomen werden miteinander zu einer Zick-Zack-förmigen Kette verbunden (Abbildung 3.5).

## 3.2 Wechselwirkungen

DNA besitzt Biege-, Torsions- und Streckelastizität. Aufgrund der Phosphatgruppen ist die DNA elektrostatisch geladen. Die elektrostatische Wechselwirkung wird durch Ionen im Lösungsmittel Wasser abgeschirmt. Die Nukleosomen sind elastisch an die Linker DNA angekoppelt. Die Nukleosomen können sich nicht durchdringen und können sich aufgrund ihrer elektrostatischen Ladung anziehen. In diesem Abschnitt werden die Potentiale beschrieben, die diese Eigenschaften der DNA-Segmente und der Nukleosomen beschreiben.



**Abbildung 3.5:** Skizze der geometrischen Struktur des Modells der Chromatinfiber. Die verschiedenen Arten der Kopplung von Nukleosomen und DNA sind wie in Abbildung 3.4 bezeichnet. Die Segmente sind fett eingezeichnet. Die Punkte zeigen die Segmentgrenzen an. Die Rechtecke symbolisieren die Nukleosomen, die gestrichelten Linien zeigen zum Schwerpunkt der Nukleosomen.

### 3.2.1 Elastische Energien

Gekoppelte Segmente, DNA und Nukleosomen, bewirken elastische Kräfte bei Streckung, Biegung und Torsion. Die Kräfte werden durch die harmonischen Potentiale  $U_s$ ,  $U_b$  und  $U_t$  beschrieben. Ihre Stärke hängt von der Länge der Segmente und von dem entsprechenden Parameter der Wechselwirkungsstärke  $\alpha_Y^{(X)}$ , ab.  $X$  steht für den Typ der Wechselwirkung,  $s$  – Streckung,  $b$  – Biegung,  $t$  – Torsion, und  $Y$  für den Typ der beteiligten Segmente,  $DNA$  falls es sich um DNA-Segmente handelt,  $Nuc$  falls eines der Segmente ein Nukleosom ist.

Die Größen  $\alpha_i$ ,  $\beta_i$ ,  $\gamma_i$ ,  $b_i$  und  $B_i$  sind wie in Abschnitt 3.1 definiert. Potential der Streckmomente:

$$U^{(s)}(b_i) = \frac{\alpha_Y^{(s)}}{b_i^0} (b_i - b_i^0)^2 \quad (3.2)$$

$b_i^0$  bezeichnet den Gleichgewichtswert von  $b_i$ .

Potential der Biegemomente eines Segmentes ohne intrinsische Krümmung:

$$U^{(b)}(\beta_i) = \frac{\alpha_Y^{(b)}}{b_i^0} \beta_i^2 \quad (3.3)$$

Potential der Biegemomente eines Segmentes mit intrinsischer Krümmung:

$$U^{(b)}(\theta_i) = \frac{\alpha_Y^{(b)}}{b_i^0} \theta_i^2 \quad (3.4)$$

mit  $\cos(\theta_i) = \vec{B}_i \vec{u}_{i+1}$

Potential der Torsionsmomente:

$$U^{(t)}(\alpha_i + \gamma_i) = \frac{\alpha_Y^{(t)}}{b_i^0} (\alpha_i + \gamma_i - \tau_i)^2 \quad (3.5)$$

$\tau_i$  gibt die intrinsische Torsion zwischen benachbarten Segmenten an.

### 3.2.2 Elektrostatische Energie der DNA

DNA ist aufgrund von Phosphatgruppen negativ geladen (Kapitel 1.1). Die Coulomb-Kraft wird durch Ionen, die im umgebenden Wasser gelöst sind, abgeschirmt. Die Wechselwirkung wird durch die Poisson-Boltzmann-Gleichung beschrieben. Eine analytische Lösung dieser Gleichung ist nur in speziellen Fällen möglich, die hier aber nicht anwendbar sind. Eine numerische Lösung ist aufgrund der Komplexität bisher nicht möglich [Hol93]. In älteren Arbeiten über die Simulation von superhelikaler DNA wird deswegen näherungsweise ein stark vereinfachtes zylinderförmiges Potential vom Lennard-Jones-Typ benutzt, bei dem der Durchmesser des Zylinders in Abhängigkeit von der Salzkonzentration des Lösungsmittels gewählt wird [VC95]. Eine neuere, ausführliche Untersuchung zeigte hingegen die Notwendigkeit, die elektrostatische Wechselwirkung detaillierter zu beschreiben [DGS97].

In der vorliegenden Arbeit wird die elektrostatische Energie gemäß dem Vorschlag von Konstantin Klenin berechnet [KML98]. Die Gültigkeit dieser Näherung wurde bereits in einer Vielzahl von Simulationen von DNA nachgewiesen [WMSL98, KML98, MRKL98, HSM<sup>+</sup>97, HBK<sup>+</sup>98]. Die DNA wird durch Linienladungen beschrieben. Statt der Poisson-Boltzmann-Gleichung wird die Debye-Hückel-Näherung benutzt. Die Wechselwirkungsenergie wird numerisch berechnet.

Zunächst werden geeignete Parameter eingeführt, die die relative Position der Linienladungen beschreiben. Die Mittelpunktsvektoren  $\vec{r}_i^*$ ,  $\vec{r}_j^*$  der Segmente  $i$  und  $j$  werden wie folgt definiert:

$$\vec{r}_i^* = \vec{r}_i + \frac{1}{2} b_i \vec{u}_i. \quad (3.6)$$

Die relative Orientierung der zwei Segmente läßt sich durch die vier Parameter  $\varrho_{ij}$ ,  $\gamma_{ij}$ ,  $\gamma_{ji}$ ,  $\sigma_{ij}$  beschreiben:

$$\varrho_{ij} = |\vec{\varrho}_{ij}| \quad , \quad \vec{\varrho}_{ij} = \frac{1}{b_0}(\vec{r}_j^* - \vec{r}_i^*) \quad (3.7)$$

$$\gamma_{ij} = \frac{\vec{u}_i \vec{\varrho}_{ij}}{|\vec{\varrho}_{ij}|} \quad (3.8)$$

$$\gamma_{ji} = -\frac{\vec{u}_j \vec{\varrho}_{ij}}{|\vec{\varrho}_{ij}|} \quad (3.9)$$

$$\sigma_{ij} = \frac{[\vec{u}_i \times \vec{\varrho}_{ij}][\vec{u}_j \times \vec{\varrho}_{ji}]}{||[\vec{u}_i \times \vec{\varrho}_{ij}]||[\vec{u}_j \times \vec{\varrho}_{ji}]||} \quad (3.10)$$

$\varrho_{ij}$  ist ein Maß für den Abstand der Segmente,  $\gamma_{ij}$  und  $\gamma_{ji}$  parametrisieren den Winkel zwischen den Segmenten und dem Abstandsvektor,  $\sigma_{ij}$  ist ein Parameter für den Winkel zwischen den Segmenten.

Die elektrostatische Potential einer Ladung  $q$  im Abstand  $r$  ist gemäß der Debye–Hückel–Näherung [Jac83]

$$\Phi(r) = \frac{q}{4\pi\epsilon\epsilon_0} \cdot \frac{e^{-\kappa r}}{r}. \quad (3.11)$$

$\epsilon_0$  bezeichnet die elektrische Feldkonstante und  $\epsilon$  die Dielektrizitätskonstante des Lösungsmittels.  $\kappa$  ist die inverse Debye–Länge. Für einwertige Ionen wie  $\text{Na}^+$  gilt:

$$\kappa^2 = \frac{2e_0\rho N_A}{\epsilon_0\epsilon k_B T} \quad (3.12)$$

Hierbei ist  $e_0$  die Elementarladung,  $\rho$  die Molarität der Ionenstärke des Lösungsmittels,  $N_A$  die Avogadro–Zahl. Typische Werte für die Debye–Länge  $1/\kappa$  sind in Tabelle 3.1 angegeben.

Zur Berechnung der Wechselwirkungsenergie von zwei Linienladungen der Länge  $b_i$  und  $b_j$  wird Gleichung 3.11 integriert:

$$U_{ij}^{(e)}(\gamma_{ji}, \gamma_{ij}, \sigma_{ij}, \varrho_{ij}) = \alpha^{(e)} k_b T \int_0^{b_i} d\lambda_i \int_0^{b_j} d\lambda_j \frac{e^{-\kappa|\vec{r}_{ij}|}}{|\vec{r}_{ij}|}$$

mit  $\vec{r}_{ij} = \vec{r}_{ij}(\lambda_i, \lambda_j; \gamma_{ji}, \gamma_{ij}, \sigma_{ij}, \varrho_{ij})$  (3.13)

Die zwei Linien werden durch die Parameter  $\lambda_i$  und  $\lambda_j$  parametrisiert. Der Abstandsvektor  $\vec{r}_{ij}$  zweier Punkte mit den Parametern  $\lambda_i$  und  $\lambda_j$  auf den Linien, deren relative Orientierung durch die Größen  $\varrho_{ij}$ ,  $\gamma_{ij}$ ,  $\gamma_{ji}$  und  $\sigma_{ij}$  gegeben

Molarität [mol/l]	Debye-Länge [nm]
0,01	3,04
0,07	1,15
0,10	0,96
0,70	0,36
1,00	0,30

**Tabelle 3.1:** Debye Längen von Wasser ( $\varepsilon = 80$ ) bei verschiedenen Konzentrationen von einwertigen Ionen

ist, berechnet sich wie folgt:

$$\vec{r}_{ij}(\lambda_i, \lambda_j; \gamma_{ji}, \gamma_{ij}, \sigma_{ij}, \varrho_{ij}) = \begin{pmatrix} 0 \\ 0 \\ b_0 \varrho_{ij} \end{pmatrix} - \lambda_i \begin{pmatrix} \sqrt{1 - \gamma_{ij}^2} \sigma_{ij} \\ \sqrt{1 - \gamma_{ij}^2} \sqrt{1 - \sigma_{ij}^2} \\ \gamma_{ij} \end{pmatrix} + \lambda_j \begin{pmatrix} \sqrt{1 - \gamma_{ji}^2} \\ 0 \\ -\gamma_{ji} \end{pmatrix}. \quad (3.14)$$

Unter Benutzung dieses Zusammenhangs läßt sich Gleichung 3.13 numerisch integrieren.

In den Parameter  $\alpha^{(e)}$  gehen die effektive lineare Ladungsdichte  $\nu^*$  und die Dielektrizitätskonstante  $\varepsilon$  der Lösung wie folgt ein:

$$\alpha^{(e)} = \frac{\nu^{*2}}{4\pi\varepsilon\varepsilon_0} \quad (3.15)$$

Die lineare Ladungsdichte der DNA läßt sich angeben als  $\nu = -2e/\Delta$  mit  $\Delta = 0,34$  nm Ganghöhe pro Basenpaar (Kapitel 1.1.2). Sie wird renormiert, um den Effekt von an der DNA immobilisierten Gegenionen zu berücksichtigen und Fehler durch Benutzung der Debye-Hückel-Näherung anstatt der Poisson-Boltzmann-Gleichung zu korrigieren [Sti77]. Die Stern-Schicht von immobilisierten Gegenionen reduziert die lineare Ladungsdichte um einen Faktor 0,73 bei NaCl-Konzentrationen zwischen 1 und 500 mM [SS77]. Die reduzierte Ladungsdichte ist damit  $\nu^\dagger = 0,73 \nu$ . Die Debye-Hückel-Näherung stellt eine Linearisierung der Poisson-Boltzmann-Gleichung dar und ist nur für kleine Potentiale  $\Phi \ll k_B T/e$  gültig. Die effektive Ladungsdichte  $\nu^*$  wurde so gewählt, daß die bekannte Lösung der Debye-Hückel-Gleichung einer Linienladung mit Ladungsdichte  $\nu^*$  mit der Lösung der Poisson-Boltzmann-Gleichung für einen Zylinder des DNA-Radius und der reduzierten Ladungsdichte  $\nu^\dagger$  in den Regionen mit  $\Phi \ll k_B T/e$  übereinstimmt.

Um die so genäherte Wechselwirkungsenergie in Simulationenrechnungen zu verwenden, müßte in jedem Schritt der Simulation für alle auftretenden Paare von DNA-Segmenten Integral 3.13 numerisch ausgewertet werden. Diese numerische Auswertung wäre sehr zeitaufwendig. Aus diesem Grund wird eine große Anzahl von möglichen relativen Orientierungen vorausberechnet und in einer Tabelle gespeichert. Bei der Durchführung von Simulationsrechnungen werden dann die Tabellenwerte benutzt, um daraus interpolierte Werte zu erhalten.

Seien die tabellierten Werte mit  $E(\varrho^a, \gamma^b, \gamma^c, \sigma^d)$  bezeichnet.  $\varrho^a$  sei aus dem Intervall  $[\varrho_{min}, \varrho_{max}]$ .  $\gamma^b, \gamma^c, \sigma^d$  sind aus dem Intervall  $[-1, 1]$ . Es sei  $a = 1 \dots N_\varrho, b = 1 \dots N_\gamma, c = 1 \dots N_\gamma, d = 1 \dots N_\sigma$ .

Man wählt  $\varrho_{max}$  so, daß die entsprechende elektrostatische Energie bei diesem Abstand einen so großen Wert annimmt ( $\varrho_{max} > 10k_B T$ ), daß eine Konfiguration mit diesem Abstand in der Simulation unwahrscheinlich ist.  $\varrho_{min}$  wird so gewählt, daß für Entfernungen  $> \varrho_{min}$  die Energie im Vergleich zu den anderen im System auftretenden Energien vernachlässigt werden kann ( $< 0.01k_B T$ ). Die Anzahl der tabellierten Werte  $N_\varrho$  wird so gewählt, daß  $(\varrho_{max} - \varrho_{min}) / (N_\varrho - 1)$  kleiner ist als die Debye-Länge.  $N_\gamma$  und  $N_\sigma$  werden so gewählt, daß bei Schritten von  $b, c$  oder  $d$  um 1 die Versetzung der Enden der Segmente kleiner ist als die Debye-Länge.

Zur Berechnung der elektrostatischen Energie eines Segment-Paares werden Werte für  $a, b, c$ , und  $d$  so berechnet, so daß  $\varrho^a < \varrho_{ij} < \varrho^{a+1}, \gamma^b < \gamma_{ij} < \gamma^{b+1}, \gamma^c < \gamma_{ji} < \gamma^{c+1}, \sigma^d < \sigma_{ij} < \sigma^{d+1}$ . Zwischen den entsprechenden Werten von  $E$  wird linear interpoliert.

Die so berechnete Energie ist an den Stützstellen der Tabellierung nicht differenzierbar. Dies stellt bei Monte Carlo-Rechnungen kein Problem dar, da für die Durchführung des Algorithmus eine Berechnung der Kraft und damit eine Berechnung von Ableitungen des Potentials nicht notwendig ist. Wenn man die Dynamik des Systems z.B. mit Brownscher Dynamik-Protokollen berechnen wollte, müssten Kräfte zwischen den Segmenten bestimmt werden. In diesem Fall müßte man zur Interpolation beispielsweise Splines benutzen.

### 3.2.3 Internukleosomale Kräfte

Die Berechnung der Wechselwirkung zwischen Nukleosomen ist im Vergleich zu DNA deutlich komplexer, da Nukleosomen positive und negative Ladungen tragen. Weiterhin befindet sich ein beträchtlicher Teil der positiven Ladungen auf den flexiblen N-terminalen Histon-Schwänzen. Deren Lage läßt sich in Röntgenstreuexperimenten nicht bestimmen, da sie anscheinend keine feste Position besitzen [LMR<sup>+</sup>97].

Zur Modellierung wird ein Potential gewählt, das die attraktiven und repulsiven Wechselwirkungen global beschreibt. Die Parameter eines solchen Potentials sind Wechselwirkungsstärke und Geometrie der wechselwirkenden Teilchen. Das Nukleosom ist näherungsweise ein flacher Zylinder, dessen Durchmesser doppelt so groß ist wie seine Dicke. Die Beschreibung durch ein kugelförmiges Potential würde nicht diese Asymmetrie widerspiegeln. Deswegen wird in der vorliegenden Arbeit das Gay–Berne–Potential benutzt, das die wechselwirkenden Teilchen durch Ellipsoide beschreibt [GB81].

Das Gay–Berne–Potential beruht auf dem Lennard–Jones–Potential zweier Teilchen im Abstand  $r$  (siehe z.B. [AT87])

$$V = 4\epsilon \left[ \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^6 \right]. \quad (3.16)$$

Die Potentialtiefe  $\epsilon$  und der Abstand des Nulldurchgangs des Potentials  $\sigma$  werden modifiziert, damit das ursprünglich kugelsymmetrische Potential asymmetrisch wird [Cor48]. Berne und Pechukas benutzen die Überlappung zweier Rotations–Ellipsoide um Ausdrücke für  $\epsilon$  und  $\sigma$  zu erhalten [BP72]. Die Anisotropie wird im wesentlichen durch eine Skalierung von  $\sigma$  je nach relativer Ausrichtung der Teilchen zueinander erreicht. Die Anwendbarkeit dieser Ausdrücke in Computer–Simulationen von anisotropen Flüssigkeiten wurde in der Folgezeit von Kushick und Berne nachgewiesen [KB73, KB75]. Das Gay–Berne–Potential beinhaltet weitere Modifikationen [GB81]:

$$V = 4\epsilon \left[ \left( \frac{1}{r - \sigma + 1} \right)^{12} - \left( \frac{1}{r - \sigma + 1} \right)^6 \right] \quad (3.17)$$

Das Potential wird verschoben statt skaliert. Dadurch ist die Potential–Breite unabhängig von der Orientierung der Wechselwirkungspartner. Weiterhin werden zusätzliche Parameter eingeführt, um unterschiedliche Potentialtiefen für laterale und longitudinale Orientierungen zu ermöglichen. Mit geringfügigen Modifikationen von Kabadi besitzt das Gay–Berne–Potential die folgende Form [Kab86a, Kab86b]:

$$V(\hat{u}_1, \hat{u}_2, \hat{r}) = 4\epsilon(\hat{u}_1, \hat{u}_2, \hat{r}) \left( \left( \frac{\sigma_0}{r - \sigma(\hat{u}_1, \hat{u}_2, \hat{r}) + \sigma_0} \right)^{12} - \left( \frac{\sigma_0}{r - \sigma(\hat{u}_1, \hat{u}_2, \hat{r}) + \sigma_0} \right)^6 \right) \quad (3.18)$$

$r$  ist der Abstand der Mittelpunkte der Teilchen und  $\hat{r}$  dessen normierte Richtung.  $\hat{u}_1$  und  $\hat{u}_2$  sind normierte Vektoren, die in die Richtung der Rotations–symmetrieachse der Teilchen zeigen.  $\sigma_0$  gibt den Abstand des Nulldurchgangs

bei lateraler Orientierung an.  $\sigma$  berechnet sich allgemein wie folgt:

$$\sigma(\hat{u}_1, \hat{u}_2, \hat{r}) = \sigma_0 \left( 1 - \frac{1}{2} \chi \left\{ \frac{(\hat{r}\hat{u}_1 + \hat{r}\hat{u}_2)^2}{1 + \chi(\hat{u}_1\hat{u}_2)} \right\} + \left\{ \frac{(\hat{r}\hat{u}_1 - \hat{r}\hat{u}_2)^2}{1 - \chi(\hat{u}_1\hat{u}_2)} \right\} \right)^{-1/2} \quad (3.19)$$

Der Parameter der Anisotropie  $\chi$  ist durch die Wert von  $\sigma$  bei paralleler und senkrechter Orientierung der Teilchen zueinander definiert:

$$\chi = (\sigma_{\parallel}^2 - \sigma_{\perp}^2) / (\sigma_{\parallel}^2 + \sigma_{\perp}^2) \quad (3.20)$$

Alternativ kann die Anisotropie auch durch den Faktor  $\kappa = \sigma_{\parallel}/\sigma_{\perp}$  beschrieben werden. Man erhält damit

$$\chi = \frac{\kappa^2 - 1}{\kappa^2 + 1} \quad (3.21)$$

Die Potentialtiefe  $\epsilon$  berechnet sich wie folgt:

$$\epsilon(\hat{u}_1, \hat{u}_2, \hat{r}) = \epsilon^{\nu}(\hat{u}_1, \hat{u}_2) \epsilon'^{\mu}(\hat{u}_1, \hat{u}_2, \hat{r}) \quad (3.22)$$

$$\epsilon(\hat{u}_1, \hat{u}_2) = \epsilon_0 [1 - \chi^2 (\hat{u}_1\hat{u}_2)^2]^{-1/2} \quad (3.23)$$

$$\epsilon'(\hat{u}_1, \hat{u}_2, \hat{r}) = 1 - \frac{1}{2} \chi' \left[ \frac{(\hat{r}\hat{u}_1 + \hat{r}\hat{u}_2)^2}{1 + \chi'(\hat{u}_1\hat{u}_2)} + \frac{(\hat{r}\hat{u}_1 - \hat{r}\hat{u}_2)^2}{1 - \chi'(\hat{u}_1\hat{u}_2)} \right] \quad (3.24)$$

Man wählt üblicherweise  $\nu = 1$  und  $\mu = 2$  [GB81]. Andere Möglichkeiten der Wahl von  $\nu$  und  $\mu$  wurden von Luckhurst und Simmonds im Fall von prolaten Ellipsoiden diskutiert [LS93].  $\chi'$  ist ein zusätzlicher Anisotropie-Faktor, mit dem sich die Potentialtiefe für laterale ( $s$  — „side-to-side“) und longitudinale ( $e$  — „end-to-end“) Orientierungen regulieren läßt.

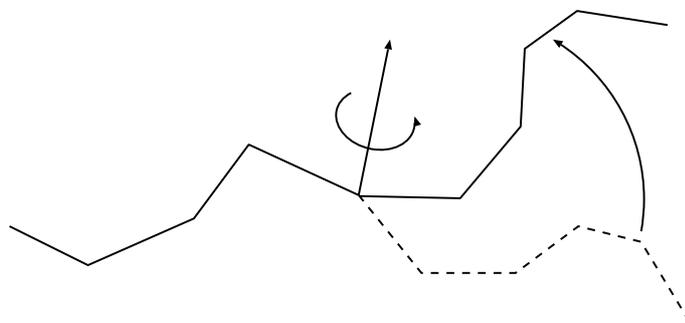
$$\chi' = (\epsilon_s^{1/\mu} - \epsilon_e^{1/\mu}) / (\epsilon_s^{1/\mu} + \epsilon_e^{1/\mu}) \quad (3.25)$$

### 3.3 Monte Carlo–Schritte

In diesem Abschnitt werden die in dieser Arbeit gewählten Monte Carlo–Schritte beschrieben. Bei dem Metropolis–Algorithmus handelt es sich um ein dynamisches Verfahren, d.h. die erzeugten Konfigurationen sind miteinander korreliert (Kapitel 2.3). Um möglichst viele Konfigurationen in kurzer Zeit erzeugen zu können, müssen die Schritte so gewählt werden, daß die Korrelationen zwischen den einzelnen Schritten möglichst gering sind. Dies wird dadurch erreicht, daß nicht nur ein Schrittyp gewählt wird, sondern

zwei [Sok96a]. Es wurden Schritte gewählt, die sich in anderen Projekten als effizient erwiesen haben, der Pivot-Schritt und der Rotations-Schritt.

Bei einem Pivotschritt wird ein Segment der Kette zufällig ausgewählt. Es erfolgt eine Drehung um eine zufällige Achse durch den Endpunkt dieses Segments für alle in der Kette nachfolgenden Elemente (Abbildung 3.6). Der Drehwinkel wird zufällig aus dem Intervall  $[-\delta, \delta]$  ausgewählt. Die Effektivität dieses Schritts wurde zuerst für Gittermodelle gezeigt [MS88] und später auf kontinuierliche Systeme verallgemeinert [FH76]. Es wurde gezeigt, daß eine Wahl von  $\delta$  günstig ist, mit der eine Akzeptanzrate von ungefähr 30% erreicht wird.

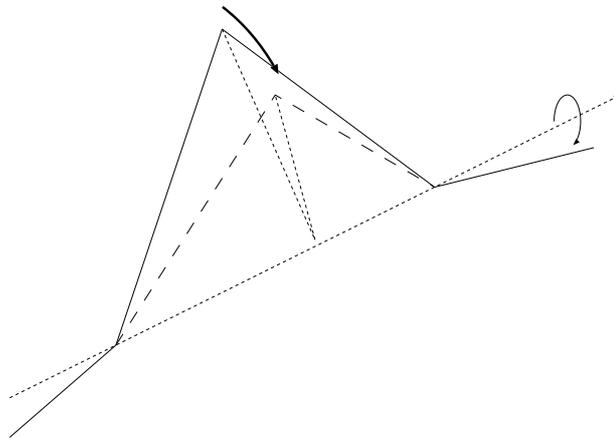


**Abbildung 3.6:** Skizze eines Pivot-Schritts

Bei einem Rotationsschritt wird ein zufällig ausgewähltes Element der Kette um die Verbindung zwischen dem vorherigem und dem nachfolgendem Element gedreht (Abbildung 3.7). Der Drehwinkel wird zufällig aus dem Intervall  $[-\phi, \phi]$  ausgewählt. Handelt es sich um das erste bzw. letzte Element der Kette, so wird der neue Zustand durch eine zufällige Drehung des ersten bzw. letzten Segments um eine zufällige Achse gebildet [AT93, Kapitel 1]. Dieser Schritt basiert auf der Verdier-Stockmayer Methode für Gitterpolymere [VS62] und wurde von Baumgärtner und Binder [BB79] für Polymersysteme verallgemeinert. Es wurde gezeigt, daß es günstig ist,  $\phi$  so zu wählen, daß eine Akzeptanzrate von ungefähr 50% erzielt wird.

### 3.4 Parametrisierung der Wechselwirkungspotentiale

Die Wechselwirkungspotentiale werden mit Hilfe von Ergebnissen experimenteller Untersuchungen parametrisiert. Eine Übersicht über die Parameter gibt



**Abbildung 3.7:** Skizze eines Rotations-Schritts

Tabelle 3.2. Die Systeme werden bei einer Temperatur von 20°C berechnet, da die Experimente, deren Ergebnisse reproduziert werden sollen, in der Regel bei dieser Temperatur durchgeführt werden.

Streckmodul DNA	665 kJ/(nm·mol)
Biegemodul DNA	120 kJ/(nm·mol)
Torsionsmodul DNA	157 kJ/(nm·mol)
elektrostatischer Radius DNA	1,2 nm
Streckmodul Nukleosom	665 kJ/(nm·mol)
Torsionsmodul Nukleosom	785 kJ/(nm·mol)
Nukleosom: $\sigma_0$	10,3 nm
$\chi$	-0,506
$\chi'$	0,383
Temperatur	20°C

**Tabelle 3.2:** In den Simulationen benutzte Parameter

### 3.4.1 Parametrisierung der elastischen Wechselwirkungen

Die Parameter für Biegung und Torsion der DNA wurden so gewählt, wie in [CL94] diskutiert. Der Biegemodul entspricht dem Konsenswert der Persistenzlänge  $\tilde{l}_b = 50$  nm, das Torsionsmodul einer Persistenzlänge  $\tilde{l}_t = 65$  nm. Für den Streckmodul wurde der von Smith et al. mit Laser-Pinzetten bestimmte Wert benutzt [SCB96].

Die Elastizität von Chromosomen wurde bisher nicht experimentell bestimmt. Das Chromosom wird deswegen vereinfacht als steif angesehen. Unendlich steife Segmente sind numerisch nur schwer handhabbar. Deswegen wird für den Torsionsmodul willkürlich das fünffache des Wertes der DNA und für das Streckmodul der gleiche Wert wie der von DNA gewählt.

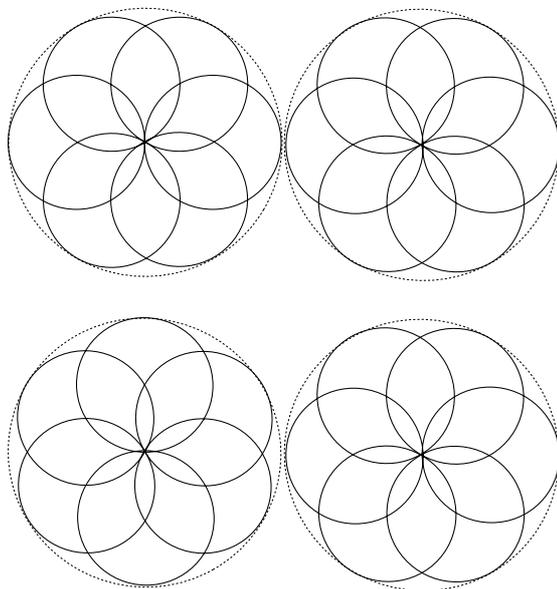
Weiterhin wird angenommen, daß die DNA an den Enden am Nukleosom fest eingespannt ist. Dies ist für Torsion experimentell nachgewiesen [MC85]. Falls das Nukleosom azetyliert ist, gilt dies nicht mehr [KB98]. Es ist zu vermuten, daß auch andere post-translationalen Modifikationen der Histone ähnliche Effekte haben. Da im folgenden nur nicht modifizierte Nukleosomen betrachtet werden sollen, hat dies aber keine Auswirkung.

### 3.4.2 Parametrisierung des Gay–Berne–Potentials

Eine direkte Bestimmung der Stärke der internukleosomalen Wechselwirkung z.B. experimentell mit Hilfe eines Rasterkraft-Mikroskops oder theoretisch mittels der Modellierung zweier Nukleosomen mit Molekulardynamik-Techniken, steht noch aus. Deswegen wird in diesem Abschnitt beschrieben, wie die Werte der Parameter des Gay–Berne–Potentials abgeschätzt werden können.

Der Anisotropie-Faktor der Potentialtiefe  $\chi'$  wird wie bei prolaten Ellipsoiden abgeschätzt [GB81]. Die Scheibe wird durch sechs Lennard–Jones Kugeln des Radius 5,5 nm wie in Abbildung 3.8 genähert und der Unterschied zwischen dem Minimum bei longitudinaler und lateraler Orientierung der Scheiben zueinander berechnet. Es gibt verschiedene Möglichkeiten der radialen Orientierung der Scheiben zueinander (Abbildung 3.8). Bei beiden Orientierungen erhält man allerdings fast identische Werte. Man bestimmt so  $\kappa' = \epsilon_s/\epsilon_e = 1/5$ , was nach Gleichung 3.25 einem Wert von  $\chi' = 0,383$  entspricht. Eine systematische Diskussion des Einflusses des Parameters  $\kappa'$  bei prolaten Ellipsoiden findet sich in [MRBA96].

Die Parameter der Geometrie der Nukleosomen und der Stärke der internukleosomalen Wechselwirkung können aus dem Vergleich von experimentellen Daten von Flüssigkristallen von Nukleosomen und theoretischen Berech-



**Abbildung 3.8:** Approximation einer Scheibe durch Kugeln. Oben und unten sind zwei verschiedene Möglichkeiten der Orientierung zueinander dargestellt.

nungen abgeschätzt werden.

Leforestier und Livolant zeigten, daß Nukleosomen–Core–Partikel bei geeigneten Konzentrationen in eine Flüssigkristallphase übergehen [LL97]. Untersuchungen mit Licht– und Elektronenmikroskopie ergeben, daß es sich um eine columnar hexagonale Phase handelt (siehe Kapitel 2.2, Abbildung 2.1). Der Abstand  $a$  zwischen zwei Säulen beträgt  $(11,55 \pm 1)$  nm, der Abstand  $s$  zwischen zwei Nukleosomen in einer Säule  $(7,16 \pm 0,65)$  nm. Es wurden bisher keine Flüssigkristalle aus Chromatosomen (mit dem Linker–Histon H1) beobachtet, da Chromatosomen eine starke Tendenz zur Aggregation haben.

Aus diesen Werten kann man  $\sigma_0$  und den Anisotropiefaktor  $\kappa$  abschätzen. Es wird angenommen, daß die gemessenen Abstände der Position den Minima des Potentials bei der entsprechenden Orientierung entsprechen. Das Minimum von Gleichung 3.18 bei lateraler Orientierung liegt bei  $r_{min}^{lat} = 2^{1/6} \sigma_0$ . Mit dem experimentellen Wert  $r_{min}^{lat} = 11,55$  nm errechnet man  $\sigma_0 = 10,3$  nm. Bei longitudinaler Orientierung berechnet man für die Lage des Minimums

$$r_{min}^{long} = \sigma_0 \left( 2^{1/6} - 1 + \left( 1 - 2 \frac{\chi}{1 + \chi} \right) \right). \quad (3.26)$$

Mit dem experimentellen Wert  $r_{min}^{long}=5,5$  nm und dem oben berechneten Wert für  $\sigma_0$  erhält man  $\chi=-0,506$ . Dies entspricht nach Gleichung 3.21  $\kappa=0,573$ .

Computer-Simulationen eines Gay-Berne-Discogens wurden von Emerson et al. durchgeführt [ELW94]. Es wurde ein System mit  $N = 256$  Gay-Berne-Teilchen ( $\kappa=0,345$ ,  $\kappa'=0,25$ ,  $\nu=1$ ,  $\mu=2$ ) bei einer skalierten Dichte  $\rho^* = N\sigma_0^3/V = 2,5$  mit Molekulardynamik simuliert. Man beobachtet einen Sprung in den Ausrichtungsparametern bei skalierten Temperaturen  $T^* = k_B T/\epsilon_0 = 4$  und  $T^* = 11$ , was einem Phasenübergang von der columnar hexagonale Phase in die diskotisch nematische Phase und dem Übergang in die ungeordnete Phase entspricht. In der diskotisch columnare Phase wird ein Abstand zwischen den Säulen von  $0,9\sigma_0$  und zwischen den Scheiben ein Abstand von  $0,5\sigma_0$  beobachtet, wobei  $\sigma_0$  der Durchmesser und  $0,345\sigma_0$  die Dicke der Teilchen ist. Bei Temperaturen unter  $T^* = 4$  geht die hexagonale columnare Phase immer mehr in eine hexagonal rektangulare Phase über.

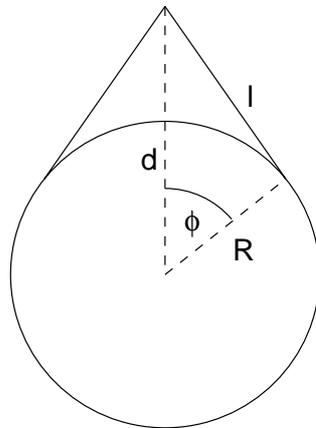
Die aus den Computersimulationen gewonnenen Werte können auf Flüssigkristalle aus Nukleosom-Core-Partikel übertragen werden und zur Eichung der Stärke des Gay-Berne-Potentials benutzt werden. Die in dieser Arbeit gewählten Werte für  $\kappa'$ ,  $\nu$  und  $\mu$  sind die gleichen wie in der Simulation;  $\kappa$  ist etwas kleiner. Flüssigkristalle liegen beim Experiment bei einem ähnlichen Wert der Dichte  $\rho^*=3,5$  vor wie bei der Simulation. Da in der Simulation der Phasenübergang zur columnaren Phase bei  $T^* = 4$  geschieht und Nukleosomen-Core-Partikel bei etwa gleichen Bedingungen columnare Phasen ausbilden, kann  $\epsilon_0 = \frac{1}{4}k_B T$  gewählt werden.

### 3.4.3 Länge der Verbindungs-DNA

Die Gesamtlänge der DNA-Segmente zwischen zwei Nukleosomen des Modells ist je nach der verwendeten Kopplung zwischen DNA und Nukleosom unterschiedlich (Abbildung 3.4). Der genaue Verlauf der DNA außerhalb des Kernteilchens ist nicht bekannt. Aus der Geometrie des Core-Partikels und des Stammmotifs kann aber die Länge der Verbindungs-DNA abgeschätzt werden.

Im Nukleosomen-Core-Partikel sind 146 bp DNA in 1,65 Windungen um die Histone gewunden. Zwei Windungen entsprechen 177 bp. Die Länge der Verbindung ist also bei direkter Kopplung die Repeat-Länge abzüglich 177 bp.

Der Stamm des Nukleosoms endet etwa 8 nm von der Achse des Nukleosoms entfernt. Wenn man annimmt, daß die DNA vom Ende des Kontakts mit dem Kernteilchen direkt zu diesem Ende läuft, kann man diese Länge wie in Abbildung 3.9 skizziert berechnen. Man erhält eine Länge von 7,36 nm entsprechend 22 bp. Die Gesamtzahl der Basenpaare der DNA im Core-Partikel



**Abbildung 3.9:** Skizze zur Berechnung der Länge der DNA im Stamm.  $R$  ist der Nucleosomen-Radius,  $d$  der Abstand zwischen Nucleosomenachse und Ende des Stamms,  $l$  die gesuchte Länge der DNA, die nicht an den Nucleosomen-Kern gebunden ist.

und dem Stamm ist  $146 \text{ bp} + 2 \times 22 \text{ bp} = 190 \text{ bp}$ . Die Länge der Verbindungs-DNA ist also die Repeat-Länge abzüglich  $190 \text{ bp}$ .

Eine Aufstellung der Länge der Verbindungs-DNA für verschiedene Zelltypen findet sich in Tabelle 3.3. Bei HeLa-Zellen wäre die Verbindungslänge bei der Annahme eines Nucleosomenstamms  $2 \text{ bp}$ , bei COS-7-Zellen gar  $0 \text{ bp}$ . Dies ist sehr kurz. Man muß deswegen annehmen, daß Nucleosomen dieser Zellen keinen Stamm besitzen. Die meisten Simulationen mit Ankopplung am Stamm wurden mit einer Verbindungslänge von  $11 \text{ bp}$  ( $3,8 \text{ nm}$ ) durchgeführt, um ein durchschnittliches System zu beschreiben.

Organismus	Repeatlänge [bp]	Verbindungslänge [bp]	
		kein Stamm	Stamm
COS-7	190	13	0
HeLa	192	15	2
Rattenleber	200	23	10
Hühnererythrozyten	212	35	22
Seeigel-Sperma	220	43	30

**Tabelle 3.3:** Länge der Linker-DNA und der Länge der Verbindungs-DNA für verschiedene Zelltypen. Der erste Wert der Verbindungslänge bezieht sich auf die Modelle mit direkter Ankopplung der DNA an das Kernteilchen, der zweite Wert auf die Ankopplung an den Nukleosomen-Stamm. Werte der Repeat-Länge aus [Hol89]

# Kapitel 4

## Aufbau und Verifikation des Rechen- und Auswerteprogrammpackets

Das vorgestellte physikalische Modell zur Beschreibung von Polynukleosomen läßt sich mit analytischen Methoden nicht berechnen. Stattdessen mußte ein Computerprogramm entwickelt werden, mit dem das Modell simuliert werden kann.

Die Umsetzung des physikalischen Modell in eine Computerprogramm ist aufwendig: Insgesamt wirken im System fünf verschiedene Wechselwirkungen, von denen drei lokal (Biegung, Torsion und Streckung) und zwei nicht lokal wirken (Gay-Berne-Wechselwirkung und Elektrostatik). Weiterhin sollte das Programm so ausgelegt sein, daß Elemente mit weiterführende Eigenschaften eingefügt und auch andere Simulationsprotokolle wie Brownsche Dynamik implementiert werden können.

Das Projekt wurde in objektorientierter Weise entwickelt, um die leichte Erweiterbarkeit zu gewährleisten. Es wurden Objektdiagramme in OMT<sup>1</sup> erstellt und anschließend in der Programmiersprache C++ umgesetzt<sup>2</sup>. Dabei kamen verschiedenen Entwurfsmuster wie Iteratoren, Erbauer oder Brücken zum Einsatz [GHJV96].

Um genügend statistisch unabhängige Konfigurationen zu erzeugen, sind Rechenzeiten von 20 Tagen die Regel. Um dies zu beschleunigen, wurde das

---

<sup>1</sup>OMT, *Object Modeling Technique*, definiert Diagramme, mit denen objektorientierte Entwürfe graphisch dargestellt werden können [Rum91]. Sie ist ein Vorläufer der UML (Kapitel B.1)

<sup>2</sup>Die Vorgehensweise, das Programm unter Verwendung eines CASE-Tools in UML zu entwerfen und anschließend automatisch in Quellcode umzusetzen zu lassen, ist erst seit kurzem möglich.

Programm parallelisiert, so daß mehrere Prozessoren gleichzeitig kooperativ an einem System rechnen können.

## 4.1 Überblick objektorientierte Programmierung

In der objektorientierten Programmierung werden Elemente herkömmlicher Programmiersprachen verwendet. So ist beispielsweise C++ von der Sprache C abgeleitet oder Oberon von Modula und Pascal. In allen objektorientierten Sprachen können Variablen und Konstanten, wenn-dann Abfragen (if-then-else) und bedingte Schleifen (while-do) ausgedrückt werden. Der wesentliche Fortschritt objektorientierter Programmiersprachen besteht in der Möglichkeit, Klassen zu definieren. Eine Klasse besteht aus Daten und Funktionen<sup>3</sup>. Sowohl Daten als auch Funktionen können entweder *öffentlich* oder *privat* sein. Funktionen, die nicht zu derselben Klasse gehören, können nur *öffentliche* Funktionen aufrufen und auf *öffentliche* Variablen zugreifen. *Private* Funktionen oder Daten können nur innerhalb der Klasse verwendet werden. Damit ist es möglich, Daten zu kapseln und den Zugriff auf diese Daten zu kontrollieren.

Man muß zwischen dem Begriff der Klasse, der lediglich ein Namen für die Art und den Aufbau von Datenstrukturen ist, und dem Begriff des Objekts unterscheiden, der eine konkrete Instanzierung der Klasse darstellt. Im allgemeinen kann es viele Objekte einer Klasse geben<sup>4</sup>. Häufig wird anstelle des Begriffs *Objekt* auch der Begriff *Instanz* verwendet.

Klassen können über den Mechanismus der Vererbung Eigenschaften von anderen Klassen übernehmen. Dadurch ist es möglich, Mechanismen einmal zu programmieren und für verschiedene Anwendungen einzusetzen, ohne sie neu programmieren zu müssen<sup>5</sup>.

Die Sprache C++ stellt die Möglichkeit von parametrisierten Klassen (Template-Klassen) zur Verfügung. Bei der Definition von parametrisierten Klassen kann die Art der benutzten Klassen offengelassen werden. Sie werden durch Statthalter beschrieben<sup>6</sup>. Es handelt sich also um eine Art Makro.

---

<sup>3</sup>Häufig wird anstatt von Funktion auch von Methode gesprochen, um zwischen herkömmlichen Funktionen und Funktionen, die zu Klassen gehören, zu unterscheiden.

<sup>4</sup>Beispiel: Die Klasse *VW-Käfer* besitzt viele Instanzen – die real gebauten Autos.

<sup>5</sup>Beispiel: Die Klasse *Auto* besitzt die Variable *muß\_zum\_TÜV*, die alle 24 Monate den Wert *wahr* annimmt. Die abgeleitete Klasse *VW-Käfer* besitzt durch den Mechanismus der Vererbung diese Variable ebenfalls automatisch.

<sup>6</sup>Beispiel: Die Funktion  $max < T > (x, y)$  vergleicht die Werte von  $x$  und  $y$  vom Typ  $T$  und liefert den größeren als Resultat. Dies kann unter der Voraussetzung, daß

Die benutzte Standard Template Library (STL) besteht ausschließlich aus parametrisierten Klassen [MS96].

## 4.2 Programmstruktur

In diesem Kapitel wird nur ein kurzer Überblick über die Struktur des Programmes gegeben. Eine vollständige Darstellung findet sich im Anhang B.2. Zur Darstellung wurde die Unified Modeling Language (UML) gewählt [Bur97, BRJ99]. Im Anhang B.1 befindet sich ein kurzer Überblick über UML.

Insgesamt kann man die Klassen und Funktionen des Programmes in sechs Bereiche aufteilen: Datenstruktur, Energieberechnung, Monte Carlo-Schritte, Simulations-Kontrolle, Auswertungsprogramme und Hilfsbibliotheken.

In der Datenstruktur werden die Daten gehalten, die die Chromatinfiber repräsentieren. Objekte, deren Klasse von der Klasse *Bead* abgeleitet worden sind, beschreiben die Segmente der Struktur<sup>7</sup>. Sie halten die Eigenschaften, wie den Typ des Beads<sup>8</sup> und eine Identifikationsnummer, grundlegende Größen, wie den Ortsvektor des Segments und das lokale Koordinatensystem, vorberechnete Größen, wie den Massenschwerpunkt, und eine Kopie der veränderten Werte, um sie im Fall der Ablehnung eines Monte Carlo-Schritts wieder herstellen zu können. Diese Beads werden in einer doppelt verketteten Liste gehalten [CLR90], die um die Möglichkeit erweitert wurde an jedem Element zusätzlich auch Seitenketten einhängen zu können<sup>9</sup>. Entsprechende Iteratoren [GHJV96] ermöglichen einen einfachen Zugriff auf alle Elemente. Diese verkettete Liste mit Seitenketten (*Branched Linked List* — BLL) und die Iteratoren wurde als Template-Klasse implementiert, um sie auch für andere Projekte nutzen zu können.

Die Energieberechnung erfolgt pro Wechselwirkungstyp in zwei Klassen. Die jeweilige Energie-Klasse benutzt einen Iterator, um alle Segmente bzw. Paare von Segmente zu erhalten und berechnet mit Hilfe der entsprechenden lokalen Energie-Klassen den Wertes der Energie für dieses Segment oder Segmentpaar.

---

ein Vergleichsoperator  $>$  für Elemente des Typs  $T$  definiert ist, unabhängig vom Typ  $T$  programmiert werden.

<sup>7</sup>Ein *Bead* ist eigentlich ein Perle einer Perlenkette, während die Klasse *Bead* ein Segment beschreibt, also eigentlich *Segment* heißen sollte.

<sup>8</sup>Im ANSI Standard von C++ ist vorgesehen, daß sich Objekte selber identifizieren können. Diese Eigenschaft wurde nicht benutzt, da nicht alle verwendeten Compiler diese Eigenschaft unterstützen.

<sup>9</sup>Die Eigenschaft, Seitenelemente einhängen zu können, wurde im vorliegenden Modell nicht benutzt.

Monte Carlo-Schritte benutzen den Metropolis-Algorithmus um neue Konfigurationen zu erzeugen. Dieser Algorithmus ist unabhängig vom Typ des Schritts. Der Metropolis-Algorithmus ist deswegen in einer Basis-Klasse implementiert. Rotations- und Pivot-Schritte sind von dieser Klasse abgeleitet und implementieren lediglich die Funktion, die die Segmente versetzt.

Der Ablauf des Programmes wird von einer übergeordneten Klasse kontrolliert, dem *Controller*. Der der davon abgeleitete *SimController* bestimmt den Ablauf von Simulationen, der *AnaController* den von Auswertungsprogrammen.

Programme, die zur Auswertung der berechneten Daten benutzt werden (Kapitel 5) verwenden Controller, die von der Klasse *SimController* abgeleitet worden sind. In den Auswertungsprogrammen muß deshalb lediglich der Algorithmus programmiert werden. Das Einlesen der Daten u.s.w. wird automatisch erledigt.

Funktionen und Objekte die auch in anderen Programme eingesetzt werden könnten wie die Vektor-Klasse, die Matrix-Klasse und die Fileformat-Routinen, sind in einer Hilfsbibliothek zusammengefaßt. Sie kann leicht in anderen Programmen verwendet werden.

### 4.3 Einheitensystem

Computer besitzen endliche Rechengenauigkeit. Dadurch treten Rundungsfehler auf. Insbesondere bei der Subtraktion fehlerbehafteter Zahlen gleicher Größenordnung tritt das Problem auf, daß die Differenz eine reine Zufallszahl sein kann. Man bezeichnet dieses Problem mit numerischer Auslöschung. Weiterhin kann durch Bereichsüberschreitung des Zahlenbereiches z.B. bei Division durch kleine Zahlen Stellenverlust auftreten. Aus diesen Gründen ist es notwendig, die Einheiten so zu wählen, daß alle relevanten Größen wie Längen, Zeiten und Energien von der Größenordnung um 1 sind.

Im Rahmen des GROMACS-Projektes wurde ein für die Molekulardynamik angepaßtes Einheitensystem entworfen [SBA<sup>+</sup>96]. Es ist in Tabelle 4.1 dargestellt. Die numerischen Werte der in die Simulation eingehenden Naturkonstanten sind in Tabelle 4.2 zusammengefaßt. Der elektrische Konversionsfaktor  $f = 1/(4\pi\epsilon_0)$  gibt die Stärke der elektrischen Wechselwirkung an. Damit ist beispielsweise im Abstand  $r$  einer Punktladung  $q$  das Potential  $\Phi = f \cdot q/r$ .

Größe	Symbol	Einheit
Länge	r	nm = $10^{-9}$ m
Masse	m	u (atomare Masseneinheit) $1,66054 \cdot 10^{-27}$ kg
Zeit	t	ps = $10^{-12}$ s
Ladung	q	e = $1,6201 \cdot 10^{-19}$ C (Proton-Ladung)
Temperatur	T	K
Energie	V	kJ/mol
Elektrisches Feld	E	kJ/(mol nm e) = $1,036427 \cdot 10^7$ V/m

**Tabelle 4.1:** Elementare und abgeleitete Einheiten im GROMACS-Einheitensystem

Name	Symbol	Wert
Avogadrozahl	$N_A$	$6,0221 \cdot 10^{23}$ mol $^{-1}$
Allgemeine Gaskonstante	$R$	$8,3145 \cdot 10^{-3}$ kJ/(mol K)
Boltzmann Konstante	$k_B$	$8,3145 \cdot 10^{-3}$ kJ/(mol K)
Elektrischer Konversionsfaktor	f	138,935 kJ/(mol) nm/e $^2$

**Tabelle 4.2:** Konstanten im GROMACS-Einheitensystem [SBA<sup>+</sup>96]

## 4.4 Zufallszahlengenerator

Zur Durchführung der Pivot- und der Rotationsschritte sowie des Metropolis-Algorithmus werden Zufallszahlen benötigt. Da ein Computer keine echten Zufallszahlen erzeugen kann, werden Zufallszahlengeneratoren eingesetzt, die Pseudozufallszahlen erzeugen.

Zur Erzeugung der Zufallszahlen wurde der Algorithmus R250 von Kirkpatrick und Stoll verwendet [KS81]. Er besitzt eine lange Periode, d.h. es dauert sehr lange bis sich die Abfolge der Zufallszahlen wiederholt. Außerdem ist die Korrelationslänge der erzeugten Zahlen sehr groß. Weiterhin wird zwischen zwei Monte Carlo-Schritten eine Zufallszahl gezogen und verworfen, um mögliche Korrelationen zu zerstören.

Der mögliche Einfluß des Zufallszahlengenerators auf die Ergebnisse wurde durch vergleichende Rechnungen mit dem UNIX-Zufallszahlengenerator

*rand()* und dem einfachen Zufallszahlengenerator R16807<sup>10</sup> [Knu81] überprüft. Hierzu wurde exemplarisch der mittlere End-zu-End-Abstand sowie seine Standardabweichung von simulierten Polynukleosomen mit 25 Nukleosomen berechnet (Tabelle 4.3). Die Ergebnisse stimmen innerhalb der Standardabweichung überein.

Zufallszahlengenerator	End-zu-End-Abstand (nm)	
	Mittelwert	Standardabweichung
<i>rand()</i>	57,1	5,0
R16807	56,7	5,5
R250	57,8	5,1

**Tabelle 4.3:** Mittelwert und Standardabweichung des End-zu-End-Abstands von simulierten Ketten von 25 Nukleosomen unter Verwendung verschiedener Zufallszahlengeneratoren.

## 4.5 Parallelisierung

Simulationsrechnungen von Polynukleosomen mit 100 Nukleosomen benötigen typischerweise etwa 20 Tage Rechenzeit (Prozessor Pentium II 450), um eine genügend große Anzahl von statistisch unabhängigen Konfigurationen zu erzeugen. Um die Rechenzeit zu reduzieren, wurde das Simulationsprogramm so verändert, daß mehrere Prozessoren gleichzeitig kooperativ das Programm ausführen können.

Die Parallelisierung des Programmcodes erfolgte für Shared Memory-Architekturen unter Verwendung von POSIX-Threads [Tan95, Kapitel 12]. Die Daten werden bei Maschinen dieses Typs in einem gemeinsamen Speicher gehalten, auf den alle Prozessoren gleichberechtigt zugreifen können. Dies ist beispielsweise bei Mehrprozessor-Workstations wie Linux Dual-Prozessor-PCs oder Shared Memory-Großrechnern wie der HP V2250 technisch möglich.

Parallelisiert wurden nur die Teile des Programms, die den größten Anteil der Rechenzeit verbrauchen, um den Implementationsaufwand möglichst gering zu halten. Dazu wurde die Ausführungszeit für die verschiedenen Programmteile mit einem *Profiling*-Werkzeug<sup>11</sup> bei einem Probelauf mit 100

<sup>10</sup> Anfangszahl ist eine beliebige Zahl  $ibm_0$ . Eine neue Zufallszahl wird erzeugt nach der Regel  $ibm_{i+1} = 16807 \times ibm_i$ . So einfach ist das.

<sup>11</sup> Benutzt wurde das zum GNU-Compiler *gcc* gehörige Profiling-Programm *gprof*.

Nukleosomen ermittelt. Der Hauptanteil der Laufzeit des Programms fällt mit ca. 92% auf die Berechnung der nicht-lokalen Wechselwirkungen, der Elektrostatik und des Gay-Berne-Potentials. Der Anteil der Laufzeit für die elastischen Wechselwirkungen und die Monte Carlo-Schritte beträgt jeweils ca. 4%. Für eine nicht zu große Prozessoranzahl ist es also ausreichend, nur die nicht-lokalen Wechselwirkungen zu parallelisieren.

Zur Parallelisierung wird die Menge der Segmente in soviel Blöcke aufgeteilt, wie Prozessoren zur Verfügung stehen. Jeder Block enthält möglichst die gleiche Anzahl an Segmenten. Das Programm startet seriell und führt die gleichen Schritte aus, wie im rein seriellen Fall. Wenn die Methoden der Energieberechnung für Elektrostatik und Gay-Berne-Potential aufgerufen werden, erzeugen sie so viele Threads, wie Prozessoren zur Verfügung stehen. In jedem Thread wird dann jeweils die Wechselwirkung von Segmenten eines Blocks mit allen anderen Segmenten berechnet.<sup>12</sup> Jeder Thread benutzt dabei einen Prozessor. Danach werden die Threads wieder gelöscht.

Die Parallelisierung wurde überprüft, indem die Energie von ausgewählten Konfigurationen unter Verwendung mehrerer Prozessoren berechnet und mit dem Ergebnis der seriellen Rechnung verglichen wurde. Weiterhin wurden einige Simulationen mit identischer Initialisierung des Zufallszahlengenerators sowohl parallel als auch seriell durchgeführt. Es ergaben sich identische Trajektorien.

Die prinzipiell erreichbare Beschleunigung eines Computerprogramms durch Verwendung mehrerer Prozessoren wird durch Amdahls Gesetz [Amd67, Sch98] beschrieben. Sei  $\nu$  der Anteil der Laufzeit der parallelisierbaren Teile eines Programms und  $N$  die Anzahl der Prozessoren. Dann ist die Laufzeit des Programms  $T_N$ :

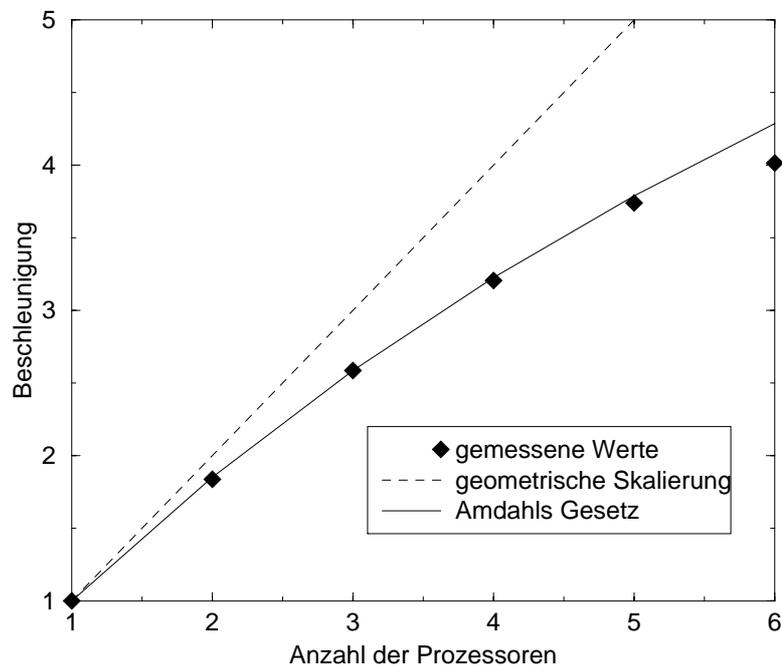
$$T_N = T_1 \left( 1 - \nu + \frac{\nu}{N} \right) \quad (4.1)$$

Den Quotienten  $T_1/T_N$  bezeichnet man als *Beschleunigung*.

Der Anteil des parallelisierten Programmteils an der Laufzeit beträgt 92%. Man erwartet also bei Verwendung von 4 Prozessoren eine Beschleunigung um den Faktor 3,2. Die auf einer HP V2250 erzielten Werte in Abhängigkeit von der Anzahl der Prozessoren im Vergleich mit der theoretischen Kurve sind in Abbildung 4.1 dargestellt. Abweichungen von Amdahls Gesetz ergeben sich erst bei einer Anzahl von mehr als vier Prozessoren.

---

<sup>12</sup>Ein spezieller Algorithmus verhindert, daß Energien doppelt berechnet werden.



**Abbildung 4.1:** Beschleunigung des parallelisierten Programms mit der Anzahl der Prozessoren. Simuliert wurden 100 Nukleosomen auf einem HP V2250-Großrechner. Es wurden 5000 Schritte berechnet. Ein Prozessor benötigt dafür 1096 s. Zum Vergleich ist die geometrische Skalierung eingetragen (Beschleunigung = Anzahl der Prozessoren) sowie die Werte, die sich nach Amdahls Gesetz bei  $\nu=92\%$  ergeben.

## 4.6 Verifikation

Verifikation<sup>13</sup> ist der „Vergleich des Produktes einer Entwicklungsphase mit seinen Vorgaben (z.B. Baustein gegen seine Spezifikation)“ [GI-96]. Bei der formalen Verifikation wird der Programm-Code zum Zeitpunkt des Übersetzens in ein ausführbares Programm — zur *Compile-Zeit* — auf syntaktische Korrektheit und während der Ausführung des Programms — zur *Laufzeit* — auf formale Konsistenz überprüft. Bei der inhaltlichen Verifikation wird zur Laufzeit die logische und inhaltliche Konsistenz und ausgewählte Resultate auf Übereinstimmung mit physikalischen Prinzipien überprüft (Tabelle 4.4).

<sup>13</sup>Im Unterschied zur Verifikation ist Validierung die „Prüfung der Anwendbarkeit einer Problemlösung (eines Produkts bzw. Systems) in seiner Umwelt“ [GI-96].

Zeitpunkt der Verifikation	formal	inhaltlich
Compilierung	X	
Ausführung	X	X
Resultate		X

**Tabelle 4.4:** Zeitpunkte der formalen und inhaltlichen Verifikation beim Prozeß des Programmtests.

### 4.6.1 Formale Verifikation zur Compile-Zeit

Zur formalen Verifikation wurden verschiedene C++-Compiler auf verschiedenen Plattformen benutzt<sup>14</sup>. Es wurde sichergestellt, daß die Compiler alle Warnungen ausgeben. Der Code wurde so lange verbessert, bis auf allen Plattformen keine Fehlermeldungen oder Warnungen mehr ausgegeben wurden. Zur tiefergehenden Analyse zur Compile- und Laufzeit wurde Purify von Rational und Insure++ von ParaSoft eingesetzt [Par98]. Dies sind Programme, die automatisch viele Programmier- und Laufzeit-Fehler einschließlich Algorithmen-Anomalien oder Speicherzuweisungsfehlern erkennen.

### 4.6.2 Überprüfung zur Laufzeit

Zur Überprüfung des Laufzeitverhaltens der Klassen wurden Programme erstellt, mit denen die Funktionstüchtigkeit der Klassen einzeln ausgetestet wurde. Weiterhin wurde das Laufzeitverhalten der Klassen während der Simulationsrechnungen überprüft. Dafür stehen verschiedene, sich ergänzende Methoden bereit. Diese können einzeln zur Übersetzungszeit aktiviert oder deaktiviert werden. Diese Überprüfungen verbrauchen bei der Durchführung von Simulationsrechnungen einen erheblichen Anteil der Rechenzeit. Deswegen wurden diese Überprüfungsmechanismen nach Abschluß von Testphasen abgeschaltet.

Formale Fehler sind beispielweise Argumente von `arcsin`, die nicht im Intervall  $[-1, 1]$  liegen, oder ein Zugriff auf Zeiger, die nicht auf gültige Speicheradressen zeigen. Diese Fehler werden mit der C-Funktion `assert` geprüft.

Die inhaltliche Konsistenz wird in der Regel über Klassen-Funktionen mit Namen `isLegal()` sichergestellt. Bei Aufruf dieser Funktion überprüfen die Objekte ihre innere Konsistenz. Beispielsweise prüft die Klasse `Bead`, ob die das lokale Koordinatensystem aufspannenden Vektoren normiert und

<sup>14</sup>Compiler: gcc 2.7.2, gcc 2.8.1, egcs 1.1, egcs1.2, aCC, SunPro, AIX-CC. Plattformen: Linux, HP-UX 11.0, Solaris, AIX

zueinander orthogonal sind. Die Klassen der Monte Carlo-Schritte besitzen zusätzlich Mechanismen, die die Invarianz bestimmter Eigenschaften überprüfen können. So sollten beispielsweise bei einem Pivot-Schritt die Winkel zwischen Segmenten, an denen nicht gedreht wurde, unverändert bleiben.

### 4.6.3 Inhaltliche Verifikation anhand von Resultaten

Es wurden verschiedene idealisierte Systeme simuliert, deren Eigenschaften analytisch berechnet werden können, die Gaußsche Kette und die Porod-Kratky-Kette (Kapitel 2.1).

Es wurden für Ketten von 100 Segmenten der Länge  $10 \text{ nm}$   $5,1 \times 10^6$  Monte Carlo-Schritte durchgeführt. Für die Berechnung des Erwartungswerts des Quadrats des End-zu-End-Abstandes  $\langle \vec{R}^2 \rangle$  wurden jeweils die ersten  $10^5$  Schritte verworfen.

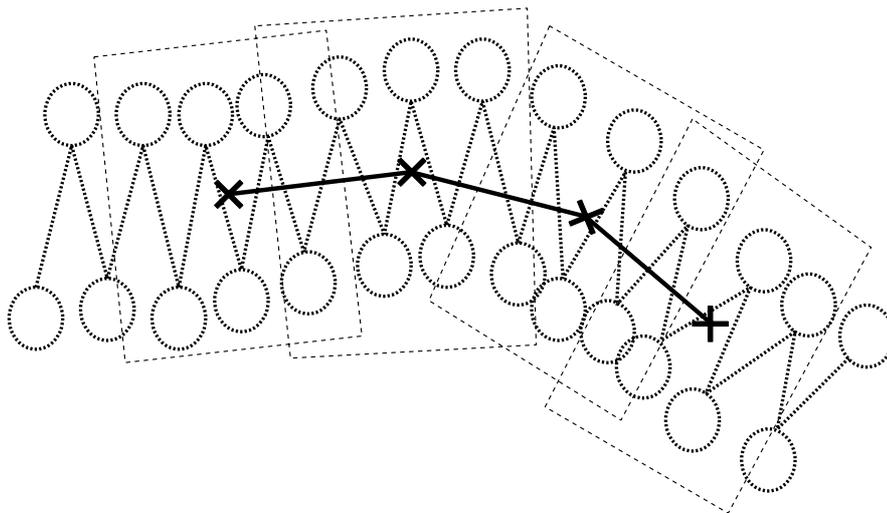
Bei der Simulation der Gaußschen Kette wurden die Parameter der Stärke der Wechselwirkungspotentiale auf 0 gesetzt, der Wert der Streckelastizität auf den von DNA. Der Mittelwert und Standardabweichung des Quadrats des End-zu-End-Abstand der simulierten Kette betragen  $\langle \vec{R}^2 \rangle = (10412 \pm 8616) \text{ nm}^2$ . Dies ist mit dem Wert der analytischen Rechnung gemäß Gleichung 2.6 von  $10000 \text{ nm}^2$  verträglich.

Bei der Simulation der Porod-Kratky-Kette wurde zusätzlich der Parameter der Biegeelastizität auf den Wert von DNA gesetzt. Mittelwert und Standardabweichung des Quadrats des End-zu-End-Abstands der simulierten Kette betragen  $\langle \vec{R}^2 \rangle = (96951 \pm 72587) \text{ nm}^2$ . Die Länge der Kette entspricht  $10 \times 10$  Kuhn-längen, da die Persistenzlänge von DNA  $50 \text{ nm}$  beträgt. Damit berechnet man mit Gleichung 2.15 den Wert von  $\langle \vec{R}^2 \rangle = 100000 \text{ nm}^2$ . Der Wert der Simulation ist mit diesem Wert verträglich.

# Kapitel 5

## Algorithmen zur Auswertung

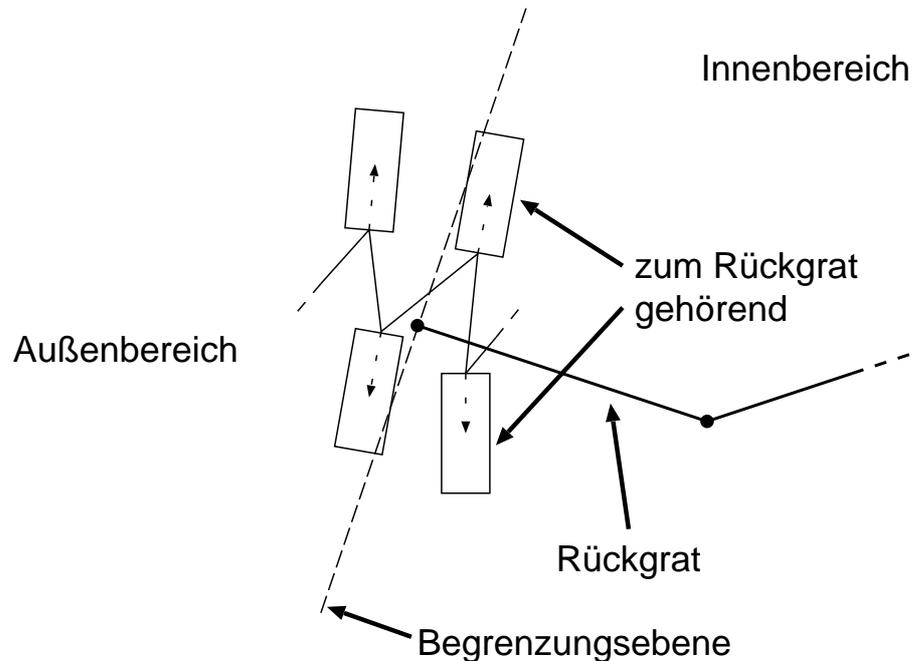
### 5.1 Konturlänge und Massenbelegungsdichte



**Abbildung 5.1:** Schematische Skizze zur Berechnung des Rückgrats der Fiber

Die Konturlänge  $L_{comp}$  einer simulierten Fiber wird als Länge eines approximierenden Rückgrats definiert (Abbildung 5.1). Dazu wird der Schwerpunkt von Beads aus einem Fenster mit jeweils  $N_{mittel}$  Elementen berechnet. Dieses Fenster wird über die Kette mit einer Schrittweite von  $N_{schritt}$  bewegt. An jedem Ende der Fiber werden  $N_{ign}$  Beads nicht berücksichtigt, um Verfälschungen durch Endeffekte zu vermeiden. Die Kette der berechneten Schwerpunkte stellt das Rückgrat der Fiber dar, das sich i.a. nicht bis zu

den Enden der Fiber erstreckt. Die benutzten Parameter sind in Tabelle 5.1 aufgelistet.



**Abbildung 5.2:** Skizze zur Berechnung der Anzahl der Nucleosomen, die zum Rückgrat der Fiber gehören.

Die Massenbelegungsichte  $\mu$  wird als das Verhältnis der (effektiven) Anzahl der Nucleosomen auf dem Rückgrat und der entsprechenden Konturlänge  $L_{comp}$  definiert:  $\mu = N_{nuk}/L_{comp}$ . Zur Bestimmung der Massenbelegungsichte werden nur die Nucleosomen herangezogen, die unmittelbar zu diesem Rückgrat gehören: An den Enden werden nur die Nucleosomen berücksichtigt, die innerhalb eines Halbraumes liegen, dessen Begrenzungsebene senkrecht zum jeweils letzten Segments des Rückgrats orientiert ist und in Richtung des Rückgrats weist (Abbildung 5.2). Die Anzahl der zum Rückgrat gehörenden Nucleosomen  $N_{nuk}$  ergibt sich dann aus der Summe der Anzahl der Nucleosomen, die an beiden Fiberenden zum Rückgrat gehören und der Anzahl der Nucleosomen zwischen den Enden.

Bei Polynucleosomen mit 100 Nucleosomen erhält man sowohl für künstlich erzeugte Strukturen als auch für simulierte Fibern einen durchschnittlichen Wert von  $N_{nuk}=80$ , solange die Fiber nicht sehr entartet ist. Bei Fibern mit 200 Nucleosomen erhält man einen Wert von 180.

Parameter	Wert
$N_{ign}$	6
$N_{mittel}$	40
$N_{schritt}$	8

**Tabelle 5.1:** Benutzte Parameter des Algorithmus zur Berechnung der Konturlänge.

## 5.2 Durchmesser der Fiber

Mit Hilfe des in Abschnitt 5.1 beschriebenen Rückgrats kann auch der Durchmesser der Fiber abgeschätzt werden. Dazu wird für jedes zum Rückgrat gehörige Nukleosom der minimale Abstand zum Rückgrat bestimmt. Von diesen minimalen Abständen wird sowohl der Mittelwert als auch das Maximum berechnet. Da die Nukleosomen etwa senkrecht zur Fiberachse orientiert sind, wird zu diesen Abständen der Nukleosomen–Radius addiert, um den entsprechenden Durchmesser der Fiber zu erhalten.

## 5.3 Neigung der Nukleosomen und der DNA zur Fiberachse

Ebenfalls mit Hilfe des in Abschnitt 5.1 beschriebenen Rückgrats kann der mittlere Neigungswinkel der Nukleosomen und der Linker–DNA zur Fiberachse berechnet werden. Für jedes zum Rückgrat gehörige Nukleosom und die diese verbindenden DNA werden die Winkel zum nächstliegenden Segment des Rückgrats berechnet. Der mittlere Neigungswinkel ist als der arithmetische Mittelwert der Winkel für alle Nukleosomen bzw. DNA–Segmente definiert.

## 5.4 Persistenzlänge

Die Persistenzlänge  $\tilde{l}_b$  der Fiber kann mit Hilfe von Gleichung 2.14 oder für lange Fibern mit Gleichung 2.15 bestimmt werden. Dazu muß der Erwartungswert des Quadrats des End–zu–End–Abstands  $\langle \vec{R}^2 \rangle$  und die Konturlänge  $L$  bekannt sein. Bei Fibern, die in der Größenordnung der Persistenzlänge liegen, ist dieses Verfahren allerdings nicht handhabbar, da hier

bereits kleine Fehler in  $L$  und  $\langle \vec{R}^2 \rangle$  zu großen Fehlern in  $\tilde{l}_b$  führen. Für diese Art von Fibern ist es besser, Gleichung 2.13 zu benutzen. Damit kann die Persistenzlänge aus dem Erwartungswert des Produkts der Richtungsvektoren an den Enden und der Konturlänge berechnet werden.

Die Konturlänge wird wie in Abschnitt 5.1 beschrieben berechnet. Zur Ermittlung der Richtungsvektoren an den Enden wird eine interpolierende Raumkurve durch die Punkte des Rückgrats gelegt. Dazu werden die Funktionen der Koordinaten der Punkte des Rückgrats von der Konturlänge durch natürliche interpolierende kubische Splines mit Hilfe eines Algorithmus aus den *Numerical Recipes* interpoliert [PTVF92]. Mit Hilfe dieses Algorithmus können auch aus der Ableitung der Splines die Richtungsvektoren an den Enden bestimmt werden.

Interpolierende kubische Splines sind differenzierbare Funktionen, die zwischen  $N$  vorgegebenen Werten  $y_i$  an den Stützstellen  $x_i$  interpolieren. Die resultierende Funktion  $f(x)$  ist zweifach differenzierbar; die zweite Ableitung  $f''$  ist stetig. Man wählt  $f''$  als lineare Funktion in den Intervallen  $[x_i, x_{i+1}]$ . Die zweiten Ableitungen an den Punkten  $x_i$  erhält man aus der Stetigkeitsbedingung für  $f''$  und den Randbedingungen  $f''(x_1)$  und  $f''(x_N)$ . Wählt man die Randbedingung  $f''(x_1) = f''(x_N) = 0$ , erhält man natürliche Splines. Mit den gegebenen Punkten und den so berechneten zweiten Ableitungen kann man die Funktion  $f$  an den Stellen zwischen Stützstellen berechnen.

## 5.5 Visualisierung

Zur Überprüfung der Ergebnisse von Auswerteprogrammen auf Plausibilität sollten die Konformationen räumlich dargestellt werden. Ziel war die dreidimensionale Darstellung der verallgemeinerten Koordinaten der Segmente, die das Simulationsprogramm als Ergebnisdaten liefert.

In der vorliegenden Arbeit wurde hierzu ein neues Programm entwickelt. Dieses erzeugt Daten im sogenannten RenderMan-Format [Pix89]. Das RenderMan-Format ist eine Beschreibungssprache für dreidimensionale Szenen und ist vergleichbar mit PostScript für zweidimensionale Objekte. Die erzeugten Daten („RenderMan Interface ByteStream“ (RIB)) enthalten Informationen über Bildgröße, Lichtquellen, Sichtwinkel und -Richtung, Positionen, Farben und Oberflächenstrukturen der darzustellenden Objekte usw. Jedes Segment wird dabei durch einen Zylinder dargestellt, dessen Größe und Farbe von der Art (DNA oder Nukleosom) abhängen. Die RenderMan-Daten dienen dann als Grundlage zur Visualisierung. Mit geeigneten Programmen wie den Blue Moon Rendering Tools [Gri96] oder RenderMan von Pixar können GL-Graphiken auf dem Bildschirm erzeugt werden oder hochwertige Graphiken erstellt werden („Raytracing“).

# Kapitel 6

## Ergebnisse

In diesem Kapitel werden die Ergebnisse von Monte Carlo-Simulationen von Polynukleosomen vorgestellt. Die Kopplung zwischen der Verbindungs-DNA und den Nukleosomen kann am Core-Partikel oder am Nukleosomen-Stamm erfolgen (Abbildung 3.4). Die Mehrzahl der Simulationen wurden bei Kopplung der Linker-DNA am Nukleosomen-Stamm durchgeführt.

### 6.1 Ankopplung der Verbindungs-DNA am Nukleosomen-Stamm

In diesem Abschnitt werden Polynukleosomen untersucht, bei denen die Verbindungs-DNA am Nukleosomen-Stamm angekoppelt ist (Abbildung 3.4, Modell 2). Der Abstand zwischen den Ankopplungspunkten beträgt 3,1 nm; diese liegen 8 nm von der Achse des Nukleosoms entfernt.

Soweit nicht anders angegeben wurden bei der Simulation die folgenden Parameter benutzt:

- Anzahl der Nukleosomen: 100
- Länge der Verbindungs-DNA 11 bp entsprechend 3,8 nm
- Stärke der internukleosomalen Wechselwirkung  $\epsilon_0 = k_B T/4$
- Salzkonzentration des Lösungsmittels 100 mM NaCl
- Temperatur: 20°C

Die Parameter für die Monte Carlo-Schritte wurden so gewählt, daß die Rotations-Schritte eine Akzeptanzrate von etwa 50% und die Pivot-Schritte eine Akzeptanzrate von etwa 30% erreichten. Der maximale Winkel bei einem Rotationsschritt beträgt typischerweise 1,5°, bei einem Pivot-Schritt 2°.

### 6.1.1 Korrelation der simulierten Daten

Wenn man den Mittelwert der Observablen  $A$  und einen Fehlerbalken dafür berechnen will, so bestimmt man durch Aufsummation  $\langle A \rangle$  und die Standardabweichung  $\sigma = (\langle A^2 \rangle - \langle A \rangle^2)^{1/2}$ . Der Fehler des Mittelwerts bei  $N$  unabhängigen Messungen ist  $\sigma/\sqrt{N}$ .

Bei dem zur Simulation verwendeten Metropolis-Algorithmus handelt es sich um einen dynamischen Algorithmus, d.h. die erzeugten Konfigurationen sind miteinander korreliert (Kapitel 2.3). Die Anzahl der unabhängigen Messungen  $N$  ist dann nicht die Anzahl der Monte Carlo-Schritte, sondern die Anzahl der erzeugten unabhängigen Konfigurationen. Dazu muß bestimmt werden, über wieviele Simulationsschritte die erzeugte Konfiguration mit der Ausgangskonfiguration korreliert ist. Um dies zu beurteilen, wird die Korrelationszeit<sup>1</sup> von Eigenschaften des Systems wie z.B. der Energie verwendet. Zur Bestimmung der Korrelationszeit in Bezug auf die Größe  $A$  wird die Autokorrelationsfunktion  $C_{AA}$  herangezogen:

$$C_{AA}(j) = \frac{\langle A(k)A(k+j) \rangle - \langle A \rangle^2}{\langle A^2 \rangle - \langle A \rangle^2} \quad (6.1)$$

$j$  und  $k$  sind Indizes, die die berechneten Konfigurationen durchnummerieren.  $C_{AA}$  fällt typischerweise exponentiell mit der Zeitkonstante  $\tau_{exp,A}$  ab. Diese Zeitkonstante kann bestimmt werden, indem die Funktion  $e^{-j/\tau_{exp,A}}$  nach der Methode der kleinsten Quadrate an die berechnete Autokorrelationsfunktion angepasst wird. Alternativ kann die integrierte Korrelationszeit  $\tau_{int,A}$  berechnet werden:

$$\tau_{int,A} = \frac{1}{2} + \sum_{j=1}^{\infty} C_{AA}(j). \quad (6.2)$$

Man kann zeigen, daß im allgemeinen gilt  $\tau_{exp,A} \approx \tau_{int,A}$ , wenn  $C_{AA}(j) \sim e^{-j/\tau}$  und  $\tau \gg 1$  [Sok96b]. Nach ungefähr  $2\tau_{int,A}$  bzw.  $2\tau_{exp,A}$  Monte Carlo-Schritten sind Ausgangs- und Endkonfiguration statistisch unabhängig [AT87, Sok96a, BH88]. Zur Berechnung von Fehlern und Schwankungen werden nur Konfigurationen verwendet, zwischen denen mehr Simulationsschritte durchgeführt worden sind, als die in Bezug auf verschiedene Observablen maximale Korrelationslänge.

Es ist möglich, daß die Korrelation einer Observable nicht nur durch eine sondern mehrere Korrelationszeiten beschrieben werden muß. Bei zwei Korrelationszeiten fällt die Autokorrelationsfunktion ab wie  $C_{AA}(j) = e^{-j/\tau_1} e^{-j/\tau_2}$ . Ist  $\tau_1 \ll \tau_2$ , so ist  $\tau_2$  die Bestimmung durch eine Anpassung nach der Methode der kleinsten Quadrate ungenau. Stattdessen kann man  $\tau_2$  bestimmen, indem

<sup>1</sup>Es handelt sich hier nicht um eine reale Zeit sondern um eine Anzahl von Schritten.

man bei der Berechnung der Autokorrelationsfunktion nur Konfigurationen auswertet, zwischen denen mehr als  $\tau_1$  Schritte liegen.

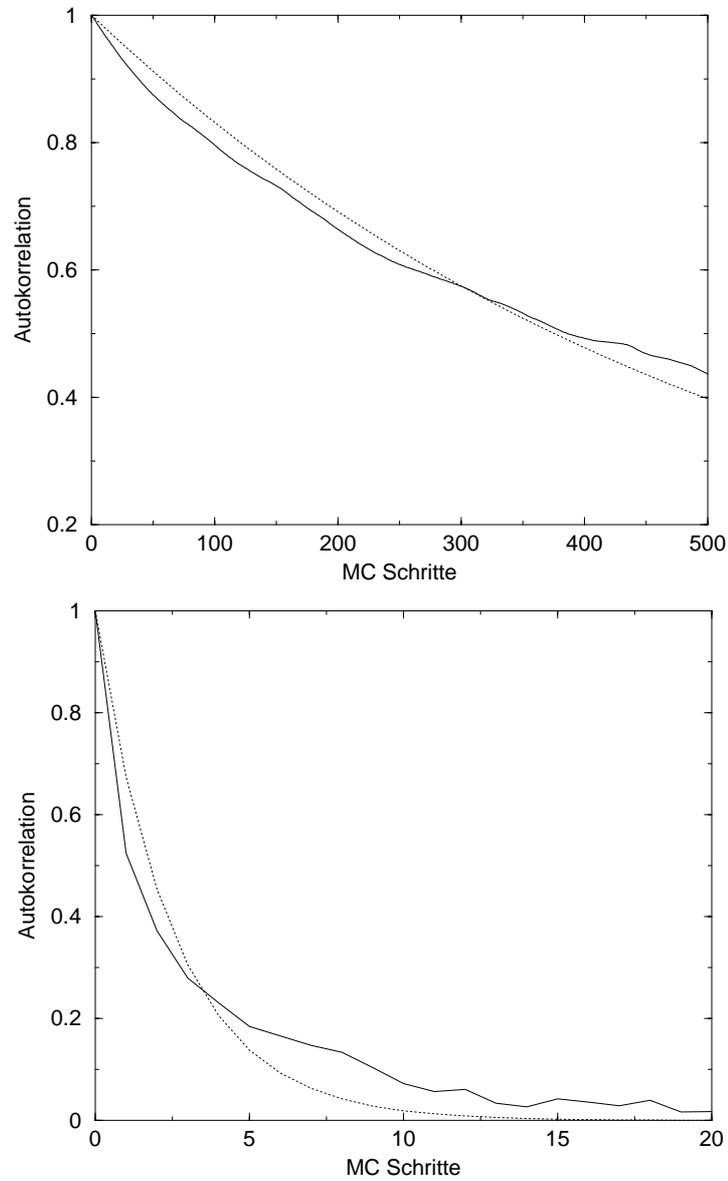
In Abbildung 6.1 ist exemplarisch die Autokorrelationsfunktion der Energie eines Polynukleosoms dargestellt. Um eine möglicherweise existente zweite, größere Korrelationszeit zu finden, wurde ebenfalls die Autokorrelationsfunktion berechnet, bei der nur jede tausendste Konfiguration ausgewertet wird. Tabelle 6.1 gibt einen zusammenfassenden Überblick über die Korrelationszeiten der Energie, des End-zu-End-Abstandes und der Massenbelegungsichte. Die Daten der Tabelle zeigen, daß nach etwa 5000 Monte Carlo-Schritten eine statistisch nicht korrelierte, neue Konfiguration erzeugt worden ist. Stichprobenartige Überprüfungen der Korrelationslängen bei anderen Simulationen ergaben keine wesentlichen Abweichungen von diesem Wert.

	Korrelationszeiten (Schritte)			
	kurz		lang	
	$\tau_{exp}$	$\tau_{int}$	$\tau_{exp}$	$\tau_{int}$
Energie	459	541	3200	2500
End-zu-End-Abstand	559	751	1700	600
Massenbelegungsichte	818	2440	2700	1570

**Tabelle 6.1:** Korrelationszeiten in Anzahl der Monte Carlo-Schritte von verschiedenen Eigenschaften aus einer Simulation mit  $\alpha=26^\circ$ ,  $\beta=110^\circ$ . Es wird die integrierte Korrelationslänge  $\tau_{int}$  und die durch einen Fit einer Exponentialfunktion an die numerisch berechnete Korrelationslänge  $\tau_{exp}$  angegeben. Bei „kurz“ wurde jede Konfiguration ausgewertet, bei „lang“ nur jede tausenste.

### 6.1.2 Anzahl der Simulationschritte

In den Simulationen wurden  $2,5 \times 10^6$  Monte Carlo-Schritte durchgeführt, soweit nicht anders angegeben. Als Ausgangsstruktur dient eine kompakte Struktur, in der die Summe der Energien der elastischen Wechselwirkungen 0 ist. Die ersten  $0,5 \times 10^6$  Schritte entsprechend 100 unabhängigen Konfigurationen wurden nicht zur Auswertung benutzt, da sie zur Relaxierung aus der willkürlichen Anfangsstruktur dienen. Es wird durch mehrere Kriterien



**Abbildung 6.1:** Autokorrelationsfunktion der Energie einer Simulation mit  $\alpha=26^\circ$ ,  $\beta=110^\circ$ . Der oberen Autokorrelation liegen 10000 Monte Carlo-Schritte (MC Schritte) zugrunde, der unteren  $3 \times 10^6$  Schritte bei Verwendung jedes 1000 Schrittes. Gestrichelt ist die mit der Methode der kleinsten Quadrate angepaßten Funktion  $e^{-\text{Schritte}/\text{Korrelationszeit}}$  dargestellt. Die so berechnete Korrelationszeit beträgt oben 541 Schritte und unten 2500 Schritte.

sichergestellt, daß diese Anzahl von Schritten zur Relaxierung genügt: Laut Sokal ist es ausreichend, die 20-fache Anzahl der Korrelationslänge für die Relaxierung zu verwerfen [Sok96a, Sok96b]. Der gewählte Wert von  $0,5 \times 10^6$  Schritte entsprechend 100 Korrelationslängen ist deutlich höher. Weiterhin verändern sich visuell beurteilt die Eigenschaften wie Energie oder Massenbelegungsichte nach wesentlich weniger Simulationsschritten nicht mehr. Außerdem unterscheiden sich die Werte für die berechneten Mittelwerte nicht von Werten aus Simulationen, bei denen als Ausgangskonfiguration eine gestreckte Konfiguration gewählt wurde (Abschnitt 6.1.5).

Bei Systemen, die keine stabile Fiber bilden oder die visuell beurteilt noch nicht relaxiert schienen, wurden  $5,5 \times 10^6$  Schritte berechnet und die anfänglichen  $10^6$  Schritte verworfen.

Die Mittelwerte werden also aus 400 statistisch unabhängigen Konfigurationen berechnet. Ein typischer Workstation-Prozessor (Intel Pentium II 450 MHz) benötigt für die Berechnung dieser Anzahl von Monte Carlo-Schritten etwa zwei Wochen.

### 6.1.3 Öffnungswinkel Linker-DNA

In der Simulation wird der Gleichgewichtswinkel  $\alpha_{ggw}$  zwischen einlaufender und auslaufender Linker-DNA vorgegeben (Abbildung 1.12). Der Begriff *Gleichgewicht* bezieht sich hier auf die elastischen Wechselwirkungen. Im Modell wird ein davon abweichender Wert ermittelt, da die elektrostatische Abstoßung zwischen den DNA-Segmenten eine Öffnung dieses Winkels bewirkt. Folglich muß  $\alpha_{ggw}$  so gewählt werden, daß der tatsächliche Winkel im Modell  $\alpha_{eff}$  mit dem experimentell bekannten Wert  $\alpha_{exp}$  übereinstimmt. Zur Anpassung von  $\alpha_{ggw}$  wurden Simulationen mit einem Winkel zwischen aufeinanderfolgenden Nukleosomen von  $\beta=80^\circ$  und verschiedenen Öffnungswinkeln  $\alpha_{ggw}$  durchgeführt. Abbildung 6.2 zeigt die Abhängigkeit des tatsächlichen Öffnungswinkels  $\alpha_{eff}$  vom vorgegebenen Öffnungswinkel  $\alpha_{ggw}$ . Bei Winkeln  $\alpha_{ggw}$  zwischen  $24^\circ$  und  $28^\circ$  entspricht der tatsächliche Winkel  $\alpha_{eff}$  den experimentellen Werten von  $35^\circ$  bis  $40^\circ$ . Für die weiteren Rechnungen wird ein Winkel von  $\alpha_{ggw}=26^\circ$  gewählt. Im folgenden wird mit  $\alpha$  immer  $\alpha_{ggw}$  bezeichnet.

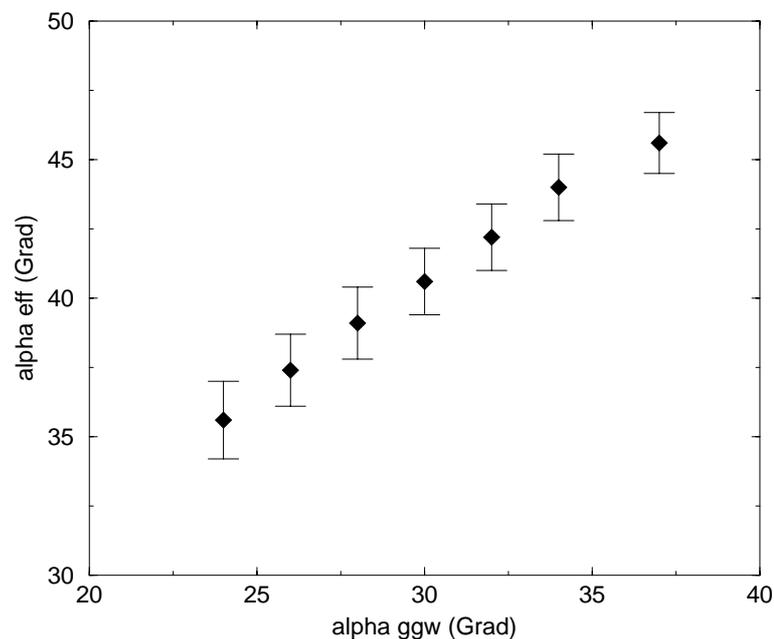
### 6.1.4 Verdrehungswinkel zwischen Nukleosomen

Der Verdrehungswinkel  $\beta$  zwischen zwei aufeinanderfolgenden Nukleosomen ist bisher experimentell nicht bestimmt worden. Er wird so angepaßt, daß die Fiber die experimentell bekannte Massenbelegungsichte besitzt.

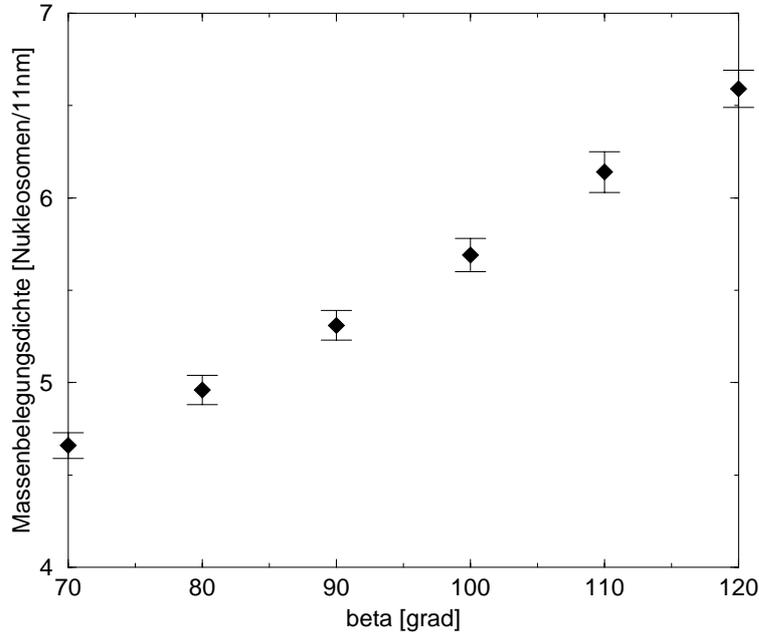
Abbildung 6.3 zeigt die Massenbelegungsdichte eines simulierten Polynukleosoms mit  $\alpha=26^\circ$  in Abhängigkeit von  $\beta$ . Übereinstimmung mit dem experimentellen Wert von 6 Nukleosomen/11 nm wird bei  $\beta=110^\circ$  erzielt. Soweit nicht anders angegeben wird dieser Wert  $\beta=110^\circ$  bei den weiteren Simulationen benutzt.

Die Eigenschaften dieser Fiber sind mit anderen experimentellen Ergebnissen im Einklang: Der mittlere Durchmesser dieser Fiber beträgt  $(31,8\pm 0,1)$  nm und der mittlere maximale Durchmesser  $(35,0\pm 0,6)$  nm, was mit den experimentellen Werten verträglich ist. Der mittlere Winkel zwischen der Nukleosomenachse und der Fiberachse ist  $50,7^\circ\pm 0,7^\circ$ , der mittlere Winkel zwischen der DNA und der Fiberachse ist  $92,5^\circ\pm 0,6^\circ$ . Die Standardabweichung beträgt  $11,5^\circ$  bzw.  $10,2^\circ$ . Beide Werte stimmen mit den experimentellen Werten überein. Die Visualisierung einer Konfiguration zeigt Abbildung 6.4.

Weiterhin wurde untersucht, wie sich die Eigenschaften der Fiber verändern, wenn  $\beta$  längs der Kette nicht konstant ist. Dazu wurde  $\beta$  gemäß einer



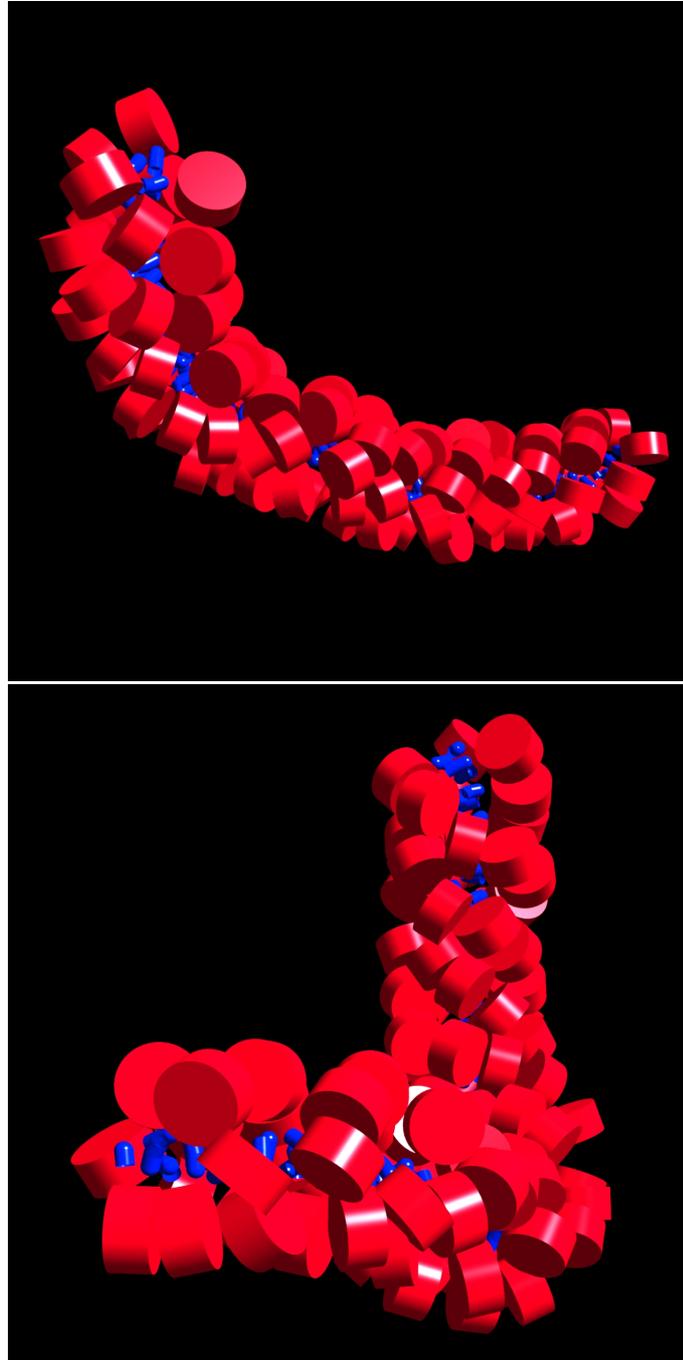
**Abbildung 6.2:** Tatsächlicher Öffnungswinkel  $\alpha_{eff}$  in Abhängigkeit vom vorgegebenen Öffnungswinkel  $\alpha_{ggw}$ .  $\beta$  ist  $80^\circ$ . Die Fehlerbalken entsprechen der Standardabweichung.



**Abbildung 6.3:** Massenbelegungsdichte eines simulierten Polynukleosoms bei  $\alpha=26^\circ$  in Abhängigkeit vom Verdrehungswinkel  $\beta$  zwischen zwei aufeinanderfolgenden Nukleosomen.

Verteilungsbreite	mittlerer Radius [nm]	Massenbelegungsdichte [Nukleosomen/11 nm]
$0^\circ$	$31,8 \pm 0,1$	$6.14 \pm 0.11$
$20^\circ$	$(31,8 \pm 0,0) \pm 0,1$	$(6,15 \pm 0,01) \pm 0,10$
$40^\circ$	$(31,4 \pm 0,2) \pm 0,2$	$(6,07 \pm 0,05) \pm 0,10$

**Tabelle 6.2:** Mittlerer Radius und Massenbelegungsdichte von Polynukleosomen mit  $\alpha=26^\circ$  und  $\beta$  normalverteilt mit Schwerpunkt bei  $110^\circ$  und verschiedenen Verteilungsbreiten. Der erste angegebene Fehler ist der Fehler der Mittelwerte, der zweite angegebene Fehler die Standardabweichung.



**Abbildung 6.4:** Visualisierung zweier Konfigurationen einer simulierten Chromatinfiber mit  $\alpha=26^\circ$  und  $\beta=110^\circ$ . Die roten Zylinder stellen die Nukleosomen dar, die blauen Zylinder die DNA.

Normalverteilung zufällig gewählt<sup>2</sup>. Der Schwerpunkt der Verteilung ist  $110^\circ$ . Es wurden für verschiedene Verteilungsbreiten jeweils drei verschiedene Konfigurationen berechnet. In Tabelle 6.2 sind die Mittelwerte der Radien und der Massenbelegungsdichten für verschiedene Verteilungsbreiten wiedergegeben. Im Rahmen der Fehlergrenzen stimmen diese Werte überein.

### 6.1.5 Relaxierung aus gestrecktem Zustand

Es muß sichergestellt werden, daß sich das System nach der Relaxierung im Gleichgewicht und nicht in einem metastabilen Zustand befindet. Dazu wird eine genügend große Anzahl der am Anfang erzeugten Konfigurationen verworfen (siehe den vorangehenden Abschnitt). Zum Nachweis eines stabilen Gleichgewichtszustands werden Simulationen mit unterschiedlichen Anfangskonfigurationen durchgeführt und die Resultate miteinander verglichen.

Es wurden drei Simulationen durchgeführt, bei denen nicht mit einer kompakten, sondern mit einer gestreckten Konformation begonnen wurde. Bei diesen Konformationen liegen alle DNA-Segmente und die Verbindungslinien der Koplungspunkte der DNA am Nukleosom auf einer Geraden (Abbildung 6.6, Bild 1). Die Nukleosomenachsen sind dann parallel zu dieser Geraden orientiert. Die Torsion zwischen den Elementen wird so gewählt, daß aufeinanderfolgende Nukleosomen gegenüber liegen.

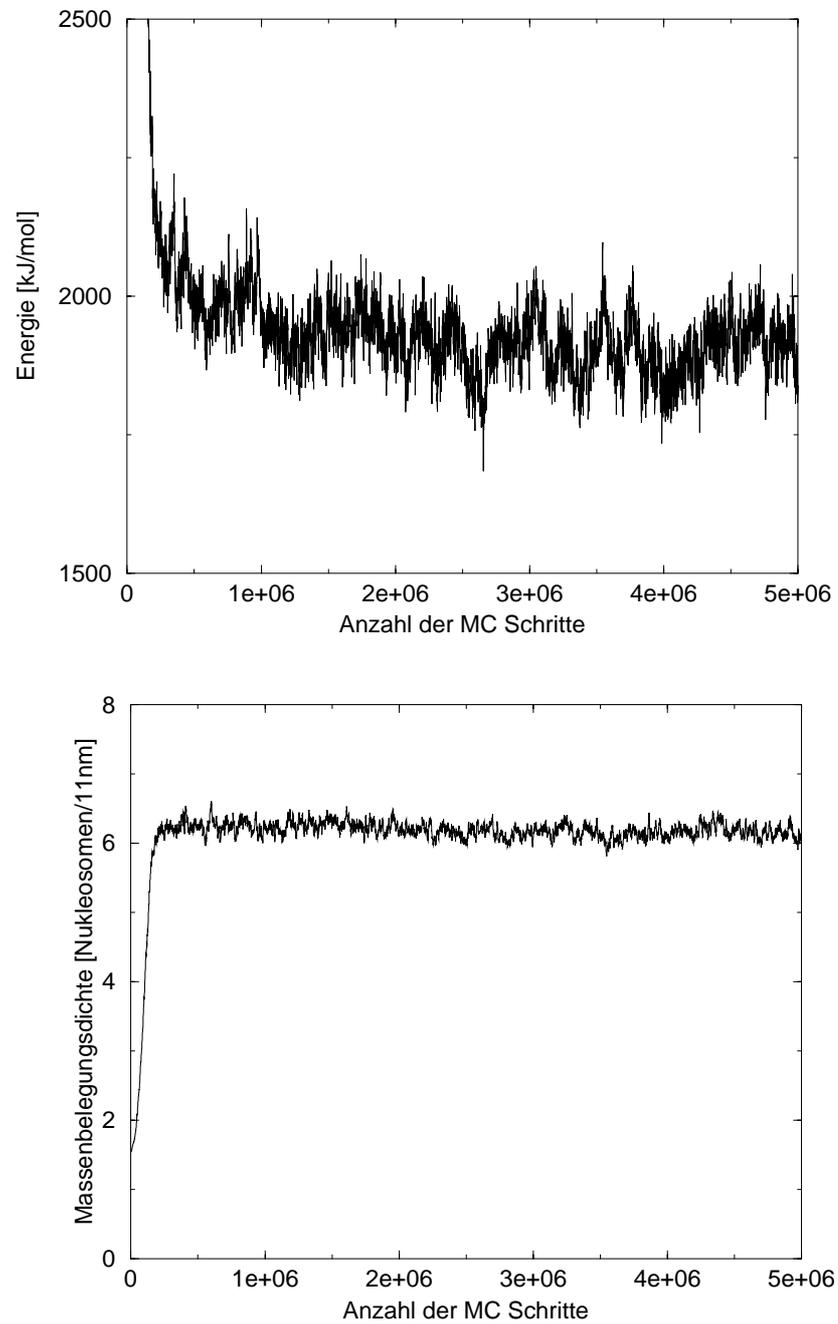
Für diese Rechnungen wurde ein kleinerer Wert für den maximalen Winkel der Pivot-Schritte benutzt ( $1^\circ$ ), da bei Vorversuchen festgestellt wurde, daß sich das System beim Equilibrieren sonst leicht in einen metastabilen Zustand begibt und ungeordnete Strukturen erzeugt werden. Die maximale Korrelationszeit für dieses System tritt bei der Massenbelegungsdichte auf und wurde auf 5000 Schritte bestimmt. Es wurden jeweils  $5 \times 10^6$  Schritte gerechnet.

Nach etwa  $3 \times 10^6$  Schritten ist das System im Gleichgewicht, wie die Entwicklung der Energie und der Massenbelegungsdichte erkennen läßt (Abbildung 6.5). Die Bildabfolge in Abbildung 6.6 gibt einen Eindruck von der Dynamik einer der Rechnungen. Man erkennt auf dem ersten Bild die gestreckte Fiber und im letzten Bild eine kompakte Fiber. Tabelle 6.3 zeigt die Mittelwerte des End-zu-End-Abstands und der Massenbelegungsdichte der

<sup>2</sup>Eine Zufallsgröße heißt normalverteilt, wenn sie eine Dichte folgender Gestalt hat:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2} \frac{(x-a)^2}{\sigma^2}}$$

$a$  ist der Schwerpunkt der Verteilung,  $\sigma$  die Standardabweichung und  $2\sigma$  die Verteilungsbreite.



**Abbildung 6.5:** Entwicklung der Energie und der Massenbelegungsdichte eines simulierten Polynukleosoms im Verlauf einer Simulation, bei der mit einer gestreckten Konfiguration begonnen wurde.

relaxierten Konfigurationen bei unterschiedlicher Startkonfiguration. Diese Werte sind im Rahmen der Schwankung unabhängig von der Startkonfiguration.

Start- konfiguration	End-zu-End-Abstand [nm]	Massenbelegungsichte [Nukleosomen/11 nm]
kompakt	152±13	6,14±0,11
gestreckt (1)	149±16	6,16±0,10
gestreckt (2)	154±13	6,13±0,10
gestreckt (3)	151±13	6,15±0,10

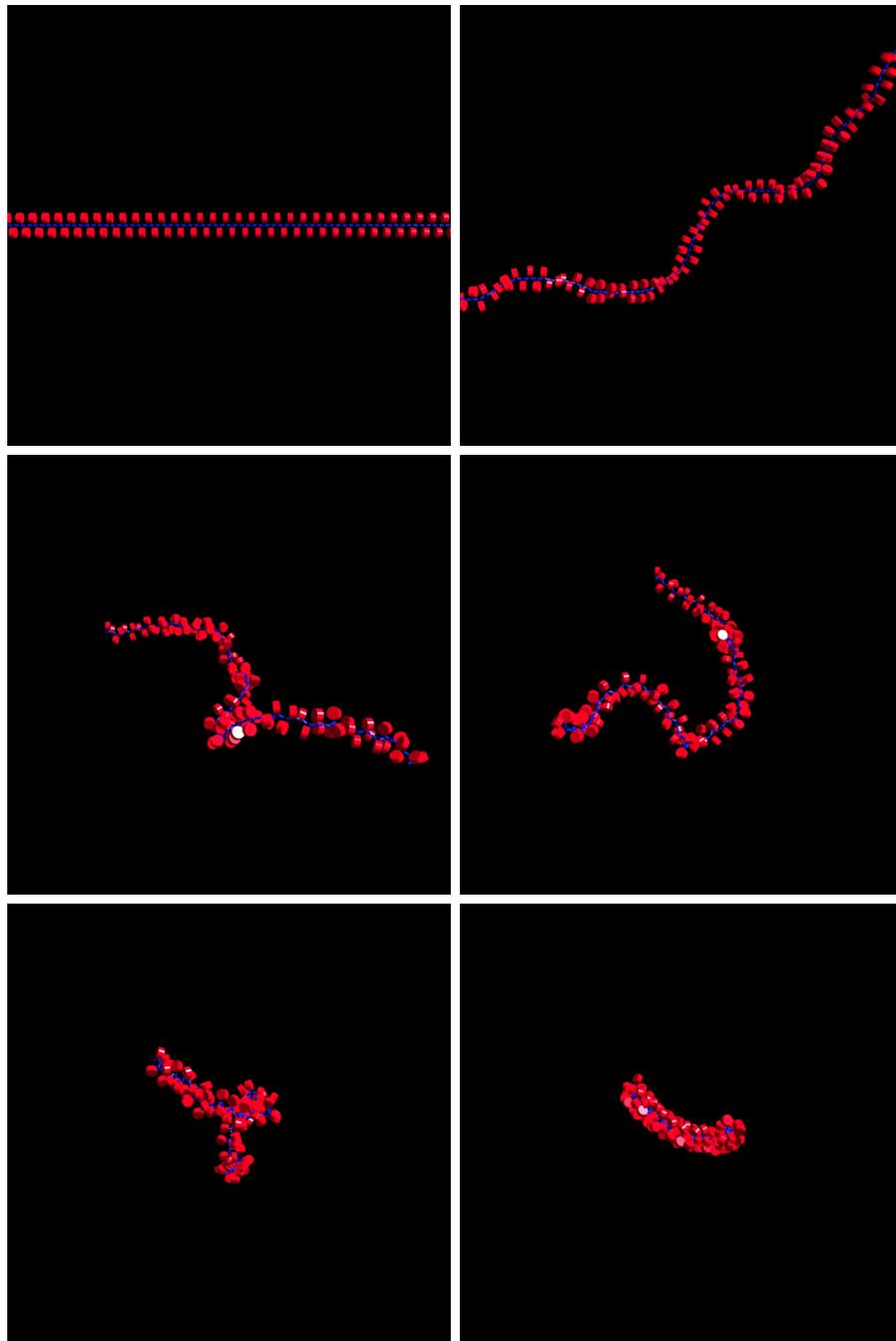
**Tabelle 6.3:** Mittelwert und Standardabweichung des End-zu-End-Abstands und der Massenbelegungsichte in Abhängigkeit von der Startkonfigurationen.

### 6.1.6 Einfluß der Stärke der internukleosomalen Wechselwirkung

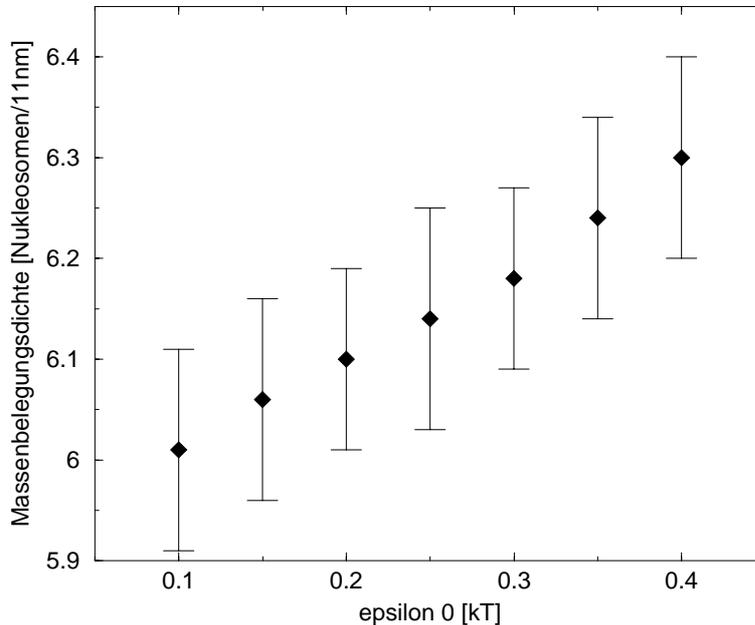
Die Stärke der internukleosomalen Wechselwirkung wird durch den Parameter  $\epsilon_0$  des Gay-Berne-Potentials skaliert. Der Wert dieses Parameters kann aus experimentellen Ergebnissen nur eingegrenzt werden (Kapitel 3.4.2). Um den möglichen Fehler durch die Wahl von  $\epsilon_0$  abzuschätzen, wurde der Einfluß dieses Parameters durch Simulationen von Polynukleosomen bei verschiedenen Werten von  $\epsilon_0$  untersucht. In Abbildung 6.7 ist die Massenbelegungsichte in Abhängigkeit von  $\epsilon_0$  abgebildet. Die Werte liegen in einem Bereich von 6,0 bis 6,3 Nukleosomen pro 11 nm. Die Massenbelegungsichte nimmt mit wachsender Stärke der internukleosomalen Wechselwirkung zu. Oberhalb  $0,4 k_B T$  bildet sich keine stabile Fiber aus. Sie kollabiert ähnlich wie in Abbildung 6.16.

### 6.1.7 Einfluß der Salzkonzentration

Die Salzkonzentration besitzt einen wesentlichen Einfluß auf die Konformation der Chromatinfiber (Kapitel 1.2.2). In dem vorgestellten Modell ändert sich mit der Salzkonzentration die elektrostatische Wechselwirkung der DNA. Um die Auswirkung dieser Änderung quantitativ zu studieren, wurden Simu-



**Abbildung 6.6:** Kompaktierung eines Polynukleosoms ausgehend von einer gestreckten Konfiguration. Links oben die Startkonfiguration, rechts unten die Endkonfiguration. Zwischen den ersten 5 Bildern liegen etwa  $1,5 \times 10^4$  Monte Carlo-Schritte, zwischen den letzten beiden  $10^5$  Schritte.

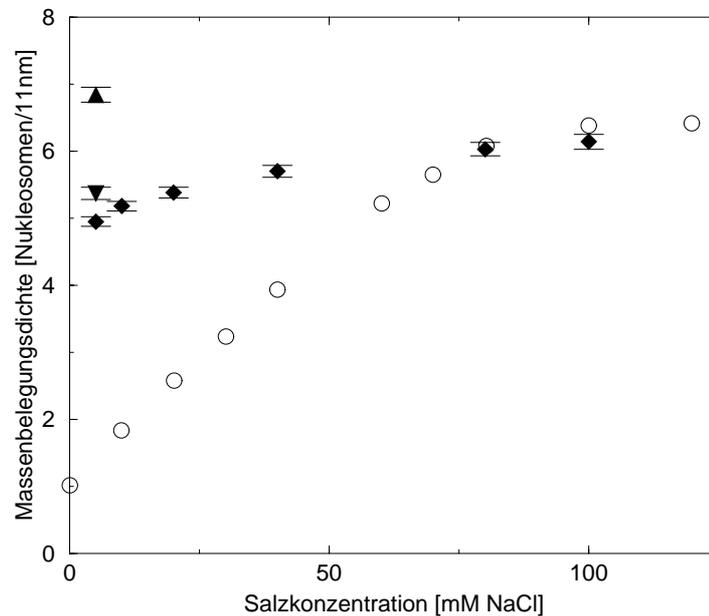


**Abbildung 6.7:** Mittelwert und Standardabweichung der Massenbelegungsichte eines simulierten Polynucleosoms in Abhängigkeit vom Parameter der Stärke der internucleosomalen Wechselwirkung  $\epsilon_0$ .

lationen von Polynucleosomen bei verschiedener Salzkonzentration durchgeführt.

Es wurden Polynucleosomen mit den Parametern  $\alpha=26^\circ$ ,  $\beta=110^\circ$  und einer Linker-DNA-Länge von 3,9 nm simuliert. In Abbildung 6.8 sind die Massenbelegungsdichten der simulierten Fibern in Abhängigkeit von der Salzkonzentration zusammen mit experimentellen Werten (Abbildung 1.9) dargestellt. Bei Salzkonzentrationen unter 70 mM NaCl werden die experimentellen Werte nicht reproduziert.

Die Abhängigkeit des effektiven Winkels zwischen einlaufender und auslaufender Linker-DNA von der Salzkonzentration wird in Abbildung 6.9 dargestellt. Die Werte bei niedrigen Ionenstärken stimmen nicht mit den experimentellen Werten überein (Kapitel 1.2.1). Deswegen wurden auch Simulationen bei einer Salzkonzentration 5 mM NaCl mit  $\alpha=42^\circ$  und  $\alpha=75^\circ$  durchgeführt. Die Ergebnisse sind ebenfalls in den Abbildungen 6.8 und 6.9 abgebildet. Die resultierenden Fibern sind noch kompakter. Die Massenbelegungsdichten stimmen deutlich schlechter mit den experimentellen Ergebnissen überein.

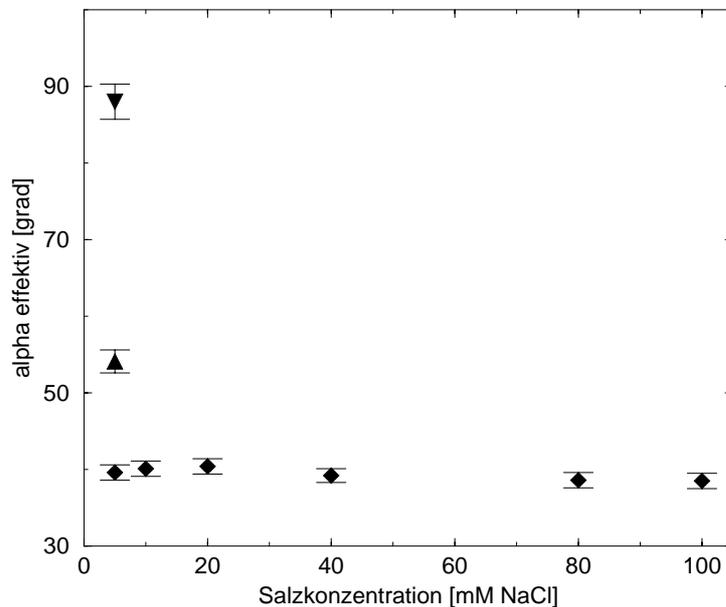


**Abbildung 6.8:** Massenbelegungsichte eines simulierten Polynukleosoms in Abhängigkeit von der Salzkonzentration. Dargestellt sind die experimentellen Werte (o) und die Werte aus der Simulation bei  $\alpha=26^\circ$  (◆),  $42^\circ$  (▲) und  $75^\circ$  (▼).

### 6.1.8 Persistenzlänge

Die Persistenzlänge der Fibern kann wie in Kapitel 5.4 beschrieben mit Hilfe von Splines bestimmt werden. Für die Konfiguration  $\alpha=26^\circ$  und  $\beta=110^\circ$  ergibt sich ein Wert von 265 nm, der im oberen Bereich der experimentellen Werte liegt (Kapitel 1.2.2). Die Standardabweichung dieses Wertes ist in der Größenordnung des Wertes selber. Eine Übersicht der Persistenzlängen bei verschiedenen Werten von  $\beta$  und  $\epsilon_{nuc}$  gibt Abbildung 6.10. Die Persistenzlänge wird mit wachsendem  $\beta$  größer. Die Stärke der internukleosomalen Wechselwirkung spielt dagegen keine Rolle. Die Schwankungen liegen im Bereich der Ungenauigkeit der Rechnung. Aus diesen Ergebnissen kann man den Fehler des Mittelwerts auf 20 nm abschätzen.

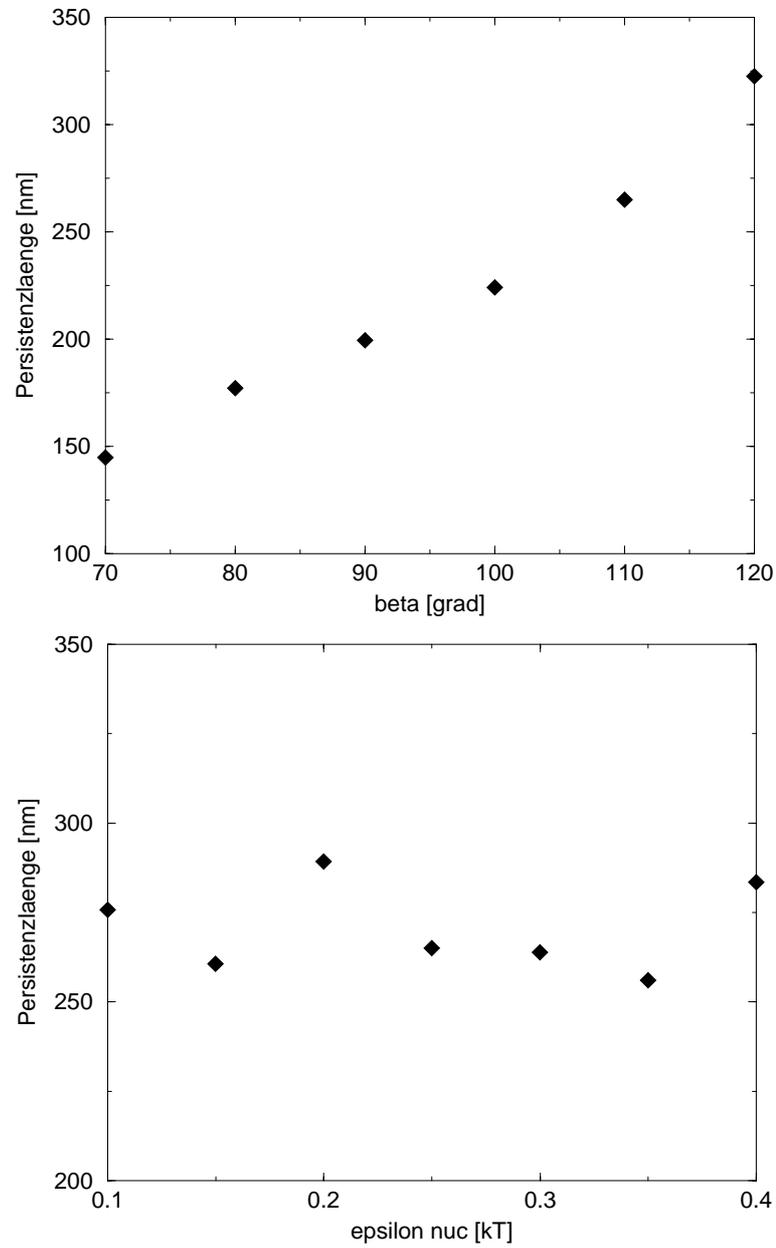
Ist  $\beta$  nicht in der gesamten Fiber konstant, sondern wird gemäß einer Gaußverteilung mit dem Schwerpunkt bei  $110^\circ$  zufällig gewählt, ergeben sich geringere Werte der Persistenzlänge. Die Persistenzlängen für die verschiedenen Konfigurationen sind in Tabelle 6.4 angegeben.



**Abbildung 6.9:** Winkel zwischen einlaufender und auslaufender Linker-DNA eines simulierten Polynukleosoms in Abhängigkeit von der Salzkonzentration. Dargestellt sind die Werte aus der Simulation bei  $\alpha=26^\circ$  ( $\blacklozenge$ ),  $42^\circ$  ( $\blacktriangle$ ) und  $75^\circ$  ( $\blacktriangledown$ ).

Breite der Verteilung	Persistenzlänge (nm)
$0^\circ$	265
$20^\circ$	270 268 260
$40^\circ$	205 182 244

**Tabelle 6.4:** Persistenzlänge von simulierten Polynukleosomen mit  $\alpha=26^\circ$  und  $\beta$  gaußverteilt mit Schwerpunkt bei  $110^\circ$  bei verschiedenen Verteilungsbreiten. Für Verteilungsbreiten von  $20^\circ$  und  $40^\circ$  wurden jeweils drei verschiedene Konfigurationen simuliert.

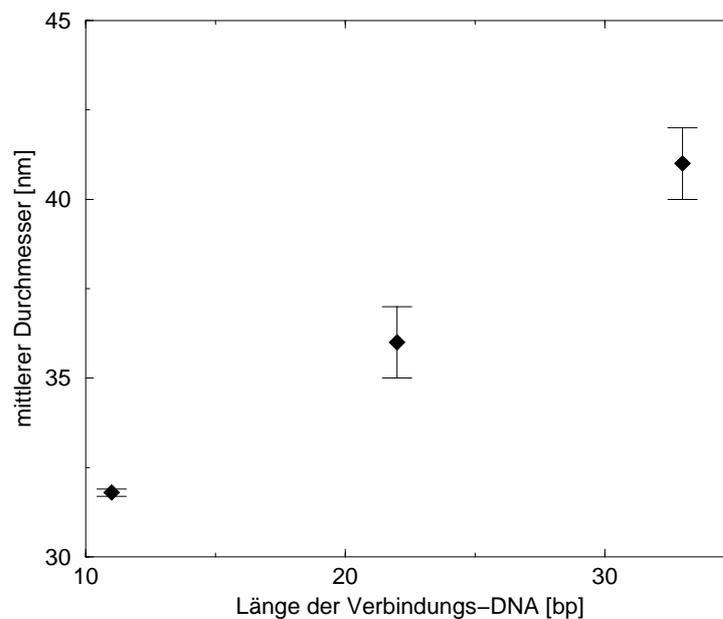


**Abbildung 6.10:** Persistenzlänge einer Fiber mit  $\alpha=26^\circ$  in Abhängigkeit von  $\beta$  (oben) und dem Parameter der Stärke der internukleosomalen Wechselwirkung  $\epsilon_{nuc}$  (unten).

### 6.1.9 Internukleosomaler Abstand

Die mittlere Abstand zwischen direkt verbundenen Nukleosomen der Fiber mit  $\alpha=26^\circ$  und  $\beta=110^\circ$  bei einer Salzkonzentration von 100 mM NaCl berechnet sich zu  $19,33\pm 0,05$  nm. Dieser liegt über den experimentellen Werten von 14 nm für Dinukleosomen. Um zu überprüfen, ob dieser Wert durch Bedingungen der Packung in die Fiber bedingt wird, wurden Dinukleosomen mit den gleichen Parametern simuliert. Es wurden  $10^6$  Monte Carlo-Schritte durchgeführt. Aus diesen Daten berechnet man für den internukleosomalen Abstand  $19,5\pm 0,4$  nm. Diese Zahl liegt ebenfalls über dem experimentell bestimmten Wert.

### 6.1.10 Einfluß der Länge der Linker-DNA

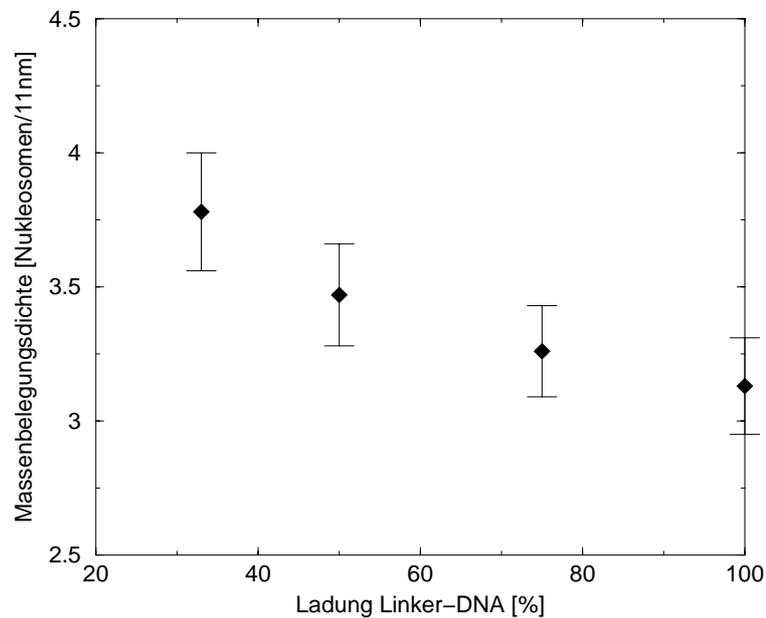


**Abbildung 6.11:** Mittlerer Durchmesser simulierter Polynukleosomen in Abhängigkeit von der Länge der Verbindungs-DNA

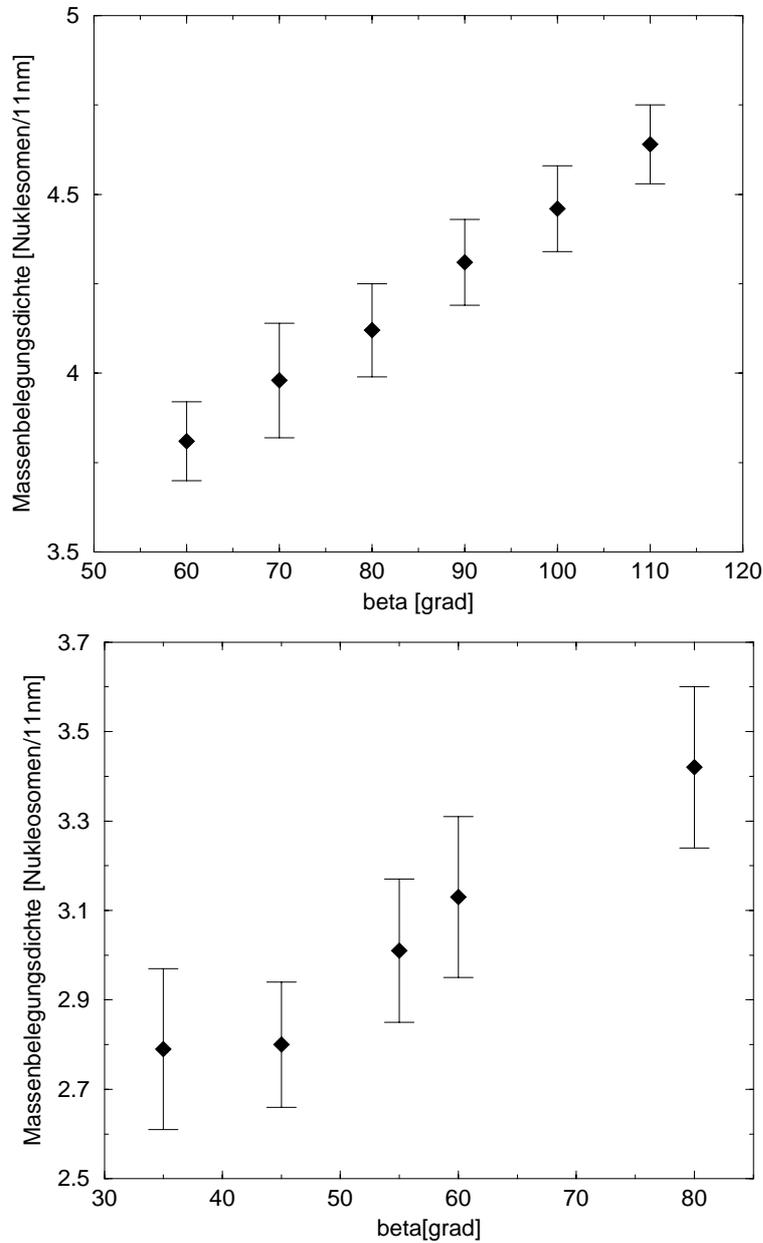
Um den Einfluß der Länge der Linker-DNA zu untersuchen, wurden Systeme mit längerer Verbindungs-DNA berechnet. In Abbildung 6.13 ist die Abhängigkeit der Massenbelegungsdichte vom Verdrehungswinkel  $\beta$  zwischen Nukleosomen bei simulierten Fibern mit 100 Nukleosomen,  $\alpha=26^\circ$  und einer Verbindungslänge von 22 bp und 33 bp dargestellt. Bei einer Verbin-

dungslänge von 33 bp bildet sich für Werte von  $\beta$  oberhalb  $80^\circ$  keine stabile Fiber. Der experimentelle Wert von 6 Nukleosomen pro 11 nm wird nicht erreicht. Der mittlere Durchmesser der Fibern sind in Abbildung 6.11 dargestellt. Konfigurationen dieser Fibern sind in Abbildung 6.14 visualisiert.

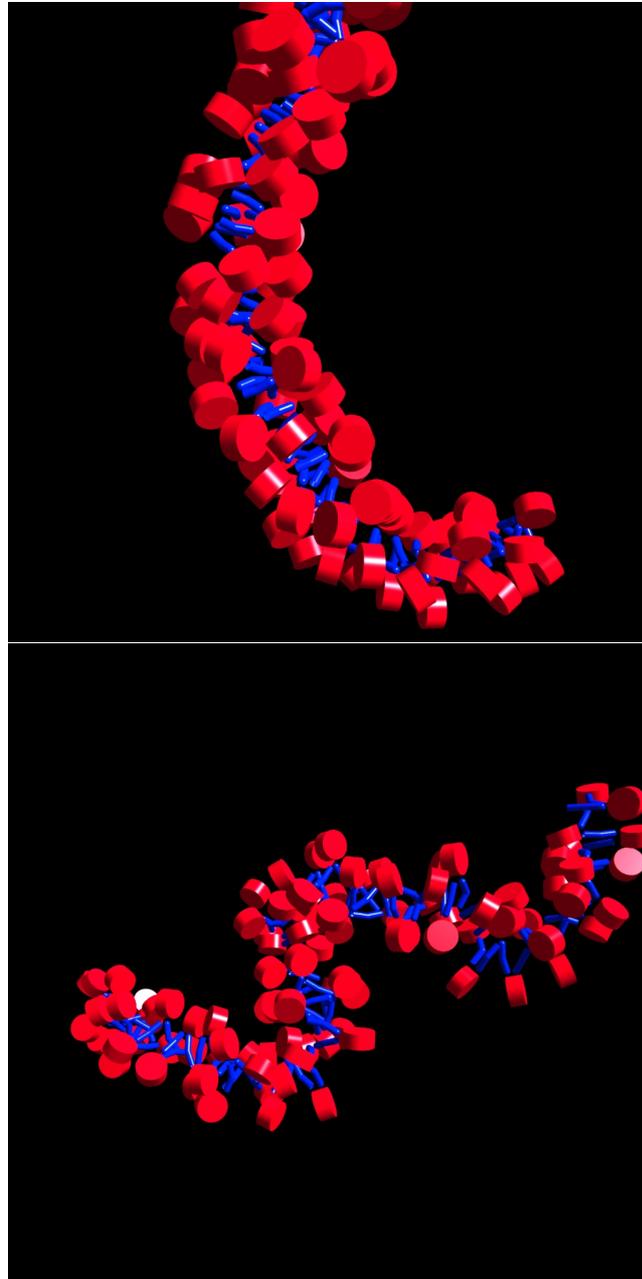
Zur Untersuchung des Einflusses der Ladungsdichte der Linker-DNA wurden Polynukleosomen mit einer Verbindungslänge von 33 bp und  $\beta=60^\circ$  mit reduzierter Ladungsdichte simuliert. Bis zu einer Reduktion auf 33% bildet sich eine stabile Fiber. Abbildung 6.12 zeigt die Abhängigkeit der Massenbelegungsichte von der Reduktion. Auch mit einer maximalen Reduktion auf 33% wird der experimentellen Wert nicht reproduziert.



**Abbildung 6.12:** Massenbelegungsichte eines simulierten Polynukleosoms mit einer Länge Linker-DNA von 33 bp bei  $\alpha=26^\circ$  und  $\beta=60^\circ$  in Abhängigkeit von der Ladungsdichte der Linker-DNA.



**Abbildung 6.13:** Massenbelegungsdichte eines simulierten Polynukleosoms mit einer Länge Linker-DNA von 22 bp (oben) und 33 bp (unten) bei  $\alpha=26^\circ$  in Abhängigkeit vom Verdrehungswinkel  $\beta$  zwischen Nukleosomen. Die DNA ist an den Nukleosomen-Stamm gekoppelt.

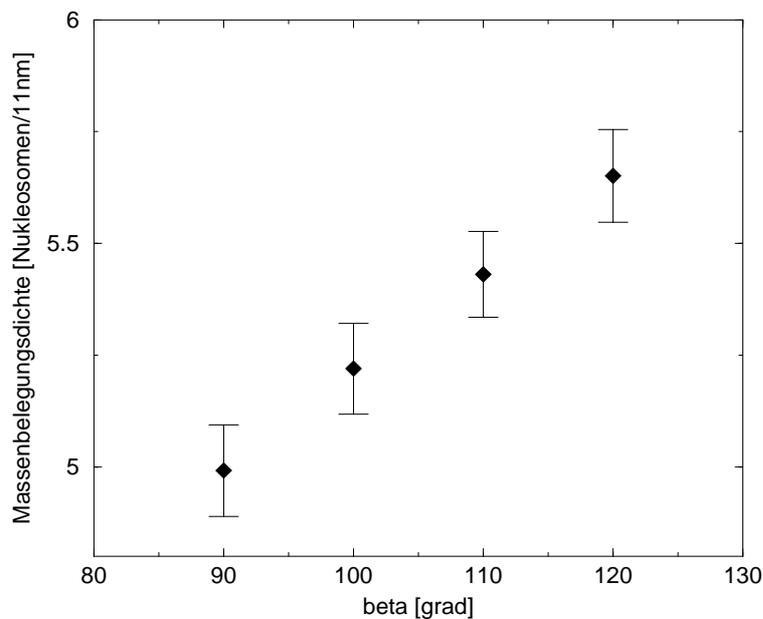


**Abbildung 6.14:** Visualisierungen von simulierten Chromatinfibern mit Verbindungs-DNA der Länge 22 bp und 33 bp. Die DNA ist an den Nukleosomen-Stamm gekoppelt. Die roten Zylinder entsprechen den Nukleosomen, die blauen Zylinder der Linker-DNA.

## 6.2 Ankopplung der Verbindungs-DNA am Core-Zylinder

In diesem Abschnitt werden Polynukleosomen untersucht, bei denen die Verbindungs-DNA in zwei Punkten im Abstand von 3,1 nm an dem Zylinder des Nukleosomen-Core-Partikels angekoppelt ist (Abbildung 3.4, Modell 1).

Die Verbindungs-DNA wird durch drei Segmente modelliert, von denen die direkt an das Nukleosom gekoppelten Segmente keine Ladung tragen. Entsprechend der Neutralisierung durch das Linker-Histon besitzen die ungeladenen Segmente eine Länge von 5 bp. Es wurden Systeme mit 100 Nukleosomen mit Verbindungs-DNA der Längen 15 bp und 40 bp berechnet. Die Salzkonzentration des Lösungsmittels beträgt 100 mM NaCl. Die Simulation wurde bei einer Temperatur von 20°C durchgeführt.

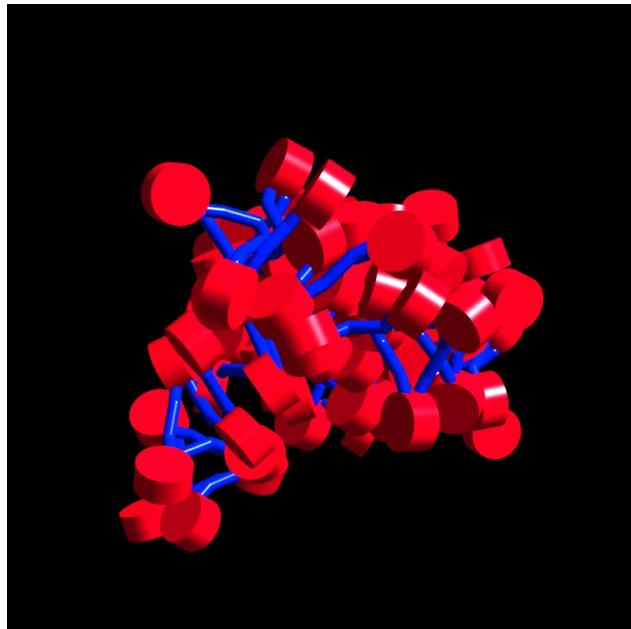


**Abbildung 6.15:** Massenbelegungsichte eines simulierten Polynukleosoms mit einer Verbindungs-DNA der Länge 15 bp bei  $\alpha=30^\circ$  in Abhängigkeit von  $\beta$ . Die DNA ist an den Nukleosomen-Core gekoppelt.

Bei Polynukleosomen mit einer Verbindungs-DNA der Länge 15 bp wurde ein Wert von  $\alpha=30^\circ$  zwischen ein- und auslaufender Linker-DNA gewählt. Der effektive Winkel beträgt dann etwa  $36^\circ$ . Die Abhängigkeit der Massenbelegungsichte der Fiber von  $\beta$  ist in Abbildung 6.15 dargestellt. Bei  $\beta=120^\circ$  entspricht der Wert der Massenbelegungsichte in etwa dem experimentel-

len Wert. Der mittlere Durchmesser der Fiber ist  $(28,0 \pm 0,1)$  nm, der mittlere maximale Radius  $(31,2 \pm 1,7)$  nm. Die Persistenzlänge ist 280 nm. Der mittlere Winkel zwischen Nukleosomenachse und Fiberachse ist  $55,2^\circ \pm 0,7^\circ$ , zwischen der DNA und Fiberachse  $85,6^\circ \pm 0,8^\circ$ .

Bei Polynukleosomen mit Verbindungs-DNA der Länge 40 bp wurde ein Wert von  $\alpha=37^\circ$ , der Winkel  $\beta$  wurde aus dem Bereich von  $25^\circ$  bis  $40^\circ$  gewählt. Der Parameter  $\epsilon_0$  wurde von  $k_B T/4$  bis  $2k_B T$  variiert. Trotz der großen Variation in den Parametern ist es nicht gelungen, eine stabile Fiber zu erzeugen. Abbildung 6.16 zeigt eine der Konfigurationen.



**Abbildung 6.16:** Visualisierung einer simulierten Chromatinfiber mit einer Verbindungs-DNA der Länge 40 bp. Die DNA ist an den Nukleosomen-Core gekoppelt. Die roten Zylinder entsprechen den Nukleosomen, die blauen Zylinder der Linker-DNA. Es bildet sich keine stabile Fiber.

# Kapitel 7

## Diskussion und Zusammenfassung

Das Verständnis der Struktur der Chromatinfiber ist ein zentrales Problem der molekularen Biochemie, da sie für die Kontrolle der Genaktivität von fundamentaler Bedeutung ist. Bisherige Modelle beschreiben den Aufbau der Fiber durch eine geometrische Anordnung ihrer Bestandteile, den Nukleosomen und der sie verbindenden DNA. Geometrische Modelle vernachlässigen aber den Einfluß von thermischen Fluktuationen, der sog. Brownschen Bewegung. Sie sind nicht in der Lage, die Auswirkung von Modifikationen der Bestandteile der Fiber oder Veränderungen des umgebenden Mediums zu berücksichtigen. In dieser Arbeit wurde ein neuartiges physikalisches Modell entwickelt, das die Eigenschaften der Bestandteile der Fiber und den Einfluß der Temperatur explizit und detailliert modelliert. Damit war es möglich, Aufbau und Eigenschaften der Fiber quantitativ zu erklären. Die Leistungsfähigkeit des vorgestellten Modells wurde durch Vergleich mit experimentell ermittelten Eigenschaften nachgewiesen. Darüberhinaus ist es nun in einfacher Weise möglich, in systematischen Studien den Einfluß sowohl von äußeren Parametern als auch von Modifikationen der Bestandteile auf die Fiber zu untersuchen.

### **Vergleich mit bisherigen Modellen von Chromatin**

Das Zick-Zack-Modell von Woodcock et al. [WGHW93] beschreibt Chromatin durch die geometrische Lage der Nukleosomen und der sie verbindenden DNA (Kapitel 1.2.3). Dieses Modell wird durch den Winkel zwischen der an das Nukleosom angekoppelten DNA  $\alpha$ , den Verdrehungswinkel zwischen aufeinanderfolgenden Nukleosomen  $\beta$  und die Länge der Linker-DNA parametrisiert (Abbildung 1.12). Darauf aufbauend wurde von Ehrlich et al.

ein Computermodell der Chromatinfiber vorgeschlagen [Ehr96, EMCL97]. Die DNA wurde hierbei durch elastische, elektrisch geladene Zylinder beschrieben, die Nukleosomen vereinfacht durch sphärische, repulsive Lennard-Jones-Potentiale. Weiterhin wurde angenommen, daß die Nukleosomen steif an die DNA gebunden sind. Die Konfigurationen wurden mit Hilfe eines Brownsche Dynamik-Protokolls erzeugt, mit dem zeitabhängige Trajektorien berechnet wurden. Mit diesem Simulationsverfahren konnten experimentell bestimmte Sedimentationskoeffizienten reproduziert werden. Gestreckte Ausgangskonfigurationen von Nukleosomenketten relaxierten in fiberartige Gebilde. Es zeigten sich jedoch drei wesentliche Schwächen:

- Es war nicht möglich, die Massenbelegungsdichte zu reproduzieren.
- Simulationen von Ketten mit mehr als 25 Nukleosomen sind sehr aufwendig in Bezug auf Rechenzeit, da das verwendete Brownsche Dynamik-Protokoll von einer Zeitkomplexität von  $N^3$  ist, mit  $N$  als Anzahl der Segmente.
- Veränderungen des Modells an sich sind durch die Struktur des Computerprogramms nur mit erheblichem Zeitaufwand möglich. Ebenso sind Portierungen auf andere Rechnerplattformen sehr aufwendig.

Im hier entwickelten Modell wurde zunächst auf die Zeitabhängigkeit verzichtet und ein deutlich schnellerer Monte Carlo-Algorithmus zur Berechnung von repräsentativen Gleichgewichtsensembeln implementiert. Damit war es möglich, Struktur und Wechselwirkungseigenschaften der Chromatinfiber realistischer zu beschreiben. Die Nukleosomen werden als Ellipsoide modelliert. Ihre Wechselwirkung wird durch das Gay-Berne-Potential beschrieben. Nukleosomen und DNA sind elastisch gekoppelt.

Zur Berechnung des Modells wurde ein objektorientiertes Computer-Programm entwickelt. Es wurde nachgewiesen, daß es prinzipiell auf allen UNIX-Plattformen lauffähig ist<sup>1</sup>. Durch Parallelisierung des Programms konnte die Rechenzeit bei Verwendung von Shared Memory-Rechnern wesentlich verkürzt werden.

Das Modell von Ehrlich benötigt lediglich zur Relaxierung eines Systems mit 25 Nukleosomen etwa zwei Wochen CPU-Zeit eines heutigen Workstationprozessors. Die Relaxierung eines Systems mit 100 Nukleosomen würde also etwa 2 Jahre erfordern, während das hier verwendete Programm nur 3 Tage benötigt. In zwei Wochen können darüberhinaus etwa 400 statistisch

---

<sup>1</sup>Voraussetzung ist die Übereinstimmung mit den wesentlichen POSIX-Standards und ein genügend leistungsfähiger C++-Compiler.

---

unabhängige Konfigurationen eines Systems von 100 Nukleosomen erzeugt werden<sup>2</sup>. Durch die Parallelisierung ist es damit möglich, auf einer gewöhnlichen Dualprozessor-Workstation die Eigenschaften einer langen Fiber innerhalb einer Woche zu berechnen.

Die Struktur des Programms ist so beschaffen, daß ein Brownsche Dynamik-Algorithmus leicht eingefügt werden kann.

### **Parametrisierung des Modells**

Die Wechselwirkungsparameter des Modells wurden aus Ergebnissen experimenteller Untersuchungen abgeleitet. Für DNA wurden die Wechselwirkungsparameter wie in herkömmlichen Simulationsrechnungen superhelikaler DNA gewählt (Tabelle 3.2). Sie sind direkt in einer Vielzahl von Experimenten bestimmt worden. Die Wahl der Parameter der internukleosomalen Wechselwirkung erfolgte durch Vergleich von experimentellen Daten von Flüssigkristallen aus Nukleosomen-Core-Partikeln mit Ergebnissen von Simulationsrechnungen von Gay-Berne-Teilchen. Sowohl bei Simulationsrechnungen als auch im Experiment ist bei geeigneter Konzentration ein Phasenübergang in eine diskotisch columnare Phase zu beobachten. Mit Hilfe der Temperaturabhängigkeit dieses Vorgangs und der Geometrie der columnaren Phase werden die Parameter des Gay-Berne-Potentials kalibriert (Kapitel 3.4.2).

### **Kopplung von DNA und Nukleosomen**

Die Linker-DNA bildet bei Hühnerzellen zusammen mit dem Linker-Histon den Nukleosomenstamm [BHG<sup>+</sup>98]. Besitzt Chromatin jedoch eine kürzere Repeat-Länge als etwa 195 bp, so ist die Existenz eines Nukleosomenstamms unwahrscheinlich, da dann die Länge der DNA zwischen den Stämmen äußerst kurz wäre (Kapitel 3.4.3). In dieser Arbeit wurden Systeme mit und ohne Nukleosomenstamm untersucht.

### **Simulationen mit Nukleosomenstamm**

Bei Simulationen von Chromatin mit Nukleosomenstamm wurden die DNA-Segmente nicht unmittelbar an den Zylinder des Nukleosomen-Cores angekoppelt, sondern an zwei Punkten des Nukleosomenstamms (Kapitel 3.1.2). Der Abstand zwischen den Kopplungspunkten entspricht dem der DNA-Stränge im Nukleosomen-Core-Partikel. Gemäß den Werten aus cryo-elek-

---

<sup>2</sup>In Experimenten werden die Eigenschaften von Polynukleosomen mit typischerweise 25 bis 100 Nukleosomen untersucht.

tronenmikroskopischen Aufnahmen wurde der Abstand der Kopplungspunkte von der Nukleosomenachse gewählt.

In diesem Modell wurde eine detaillierte Anpassung der Winkel zwischen einlaufender und auslaufender Linker-DNA  $\alpha$  und zwischen aufeinanderfolgenden Nukleosomen  $\beta$  (Abbildung 1.12) durch Vergleich mit experimentellen Werten durchgeführt: Der Winkel  $\alpha$  stellt den Gleichgewichtswinkel aufgrund von elastischen Wechselwirkungskräften dar. Im Modell wird jedoch ein davon abweichender Winkel ermittelt und zwar erhöht durch die elektrostatische Wechselwirkung der DNA-Segmente. Der Winkel  $\alpha$  wurde folglich so gewählt, daß die Werte des berechneten und experimentell bestimmten Winkels (im Rahmen der Meßtoleranz) übereinstimmen. Der Winkel  $\beta$  wurde so angepaßt, daß die Massenbelegungsdichte der Fiber mit dem experimentellen Wert übereinstimmt. Bei einer Repeat-Länge von 200 bp erhält man Werte von  $\alpha=28^\circ$  und  $\beta=110^\circ$ .

Berechnungen dieser Fibern, die Polynukleosomen mit den so gewählten Werten von  $\alpha$  und  $\beta$  bilden, reproduzieren (im Rahmen des experimentellen Fehlers) den experimentellen Wert des Durchmessers der Fiber sowie der Neigung der Nukleosomen und der DNA zur Fiberachse.

Ist der Wert von  $\beta$  nicht in der gesamten Fiber konstant, sondern wird längs der Fiber gemäß einer Normalverteilung zufällig gewählt, so ergeben sich Fibern, die den experimentellen Wert der Massenbelegungsdichte und des Radius ebenso reproduzieren.

Der Wert der Persistenzlänge der mit dem hier entwickelten Modell berechneten Chromatinfiber wird mit zunehmender Breite der Verteilung von  $\beta$  geringer. Ist  $\beta$  konstant, erhält man einen Wert von 260 nm. Bei einer Breite von  $40^\circ$  ergibt sich einen Wert von durchschnittlich 210 nm. Diese Werte sind verträglich mit dem grob abgeschätzten Wert von 70 bis 130 nm, der aus experimentellen Daten der Abstandsverteilung von genetischen Markern im Zellkern unter Annahme einfacher Modelle berechnet wurde (Kapitel 1.2.2, Seite 16). Es ist zu erwarten, daß die Persistenzlänge bei größerer Verteilungsbreite von  $\beta$  besser mit den experimentellen Werten übereinstimmt. Die berechneten Persistenzlängen sind mit dem durch Rasterkraftmikroskopie gewonnenen Wert von 30–50 nm nicht vereinbar [HMCL97, Cas94]. Der Grund hierfür ist, daß zur Beobachtung mit Rasterkraftmikroskopie die Chromatinfibern durch elektrostatische Wechselwirkung an eine Glimmer-Oberfläche gebunden werden. Durch diese Kräfte wird jedoch die Struktur gestört, wodurch die Abweichung zu den anderen experimentellen und den hier vorgestellten theoretischen Werten erklärt werden kann.

Die Stärke der internukleosomalen Kräfte ist aus Ergebnissen experimenteller Untersuchungen nur grob abschätzbar. Durch die Simulationen bei ver-

---

schiedenen Wechselwirkungsstärken wurde nachgewiesen, daß dieser Parameter lediglich einen geringen Einfluß auf die Massenbelegungsichte und die Persistenzlänge hat. Dies stellt sicher, daß die Unsicherheit in dieser Größe keine entscheidende Auswirkung auf die Ergebnisse besitzt.

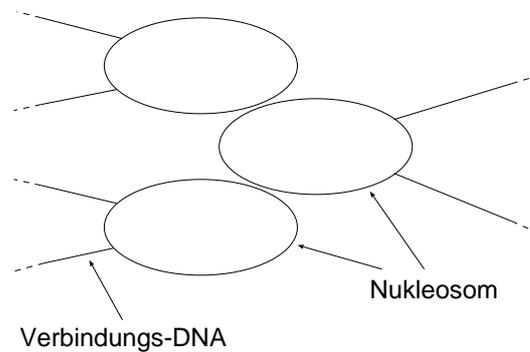
Die bisher diskutierten Polynukleosomen besitzen eine Verbindungslänge von 11 bp und somit einer Repeatlänge von 200 bp, wie sie beispielsweise in Chromatin von Rattenlebern gefunden wird. Die simulierten Polynukleosomen mit größerer Repeat-Länge bilden ebenfalls Fibern. So besitzen die simulierten Fibern mit einer Repeatlänge von 212 bp, wie z.B. in Hühnererythrozyten, Eigenschaften, die mit den experimentellen Werten (im Rahmen des experimentellen Fehlers) vereinbar sind. Bei einer noch größeren Repeat-Länge von 223 bp, wie z.B. in Seeigelsperma, ist der experimentelle Wert der Massenbelegungsichte nicht reproduzierbar. Zellen mit so langer Linker-DNA besitzen weitere Faktoren wie z.B. Spermin. Ohne diese Faktoren ist anscheinend die Kompaktierung der Fiber nicht möglich.

Der experimentelle Wert des internukleosomalen Abstands von Dinukleosomen beträgt bei physiologischer Salzkonzentration 14 nm (Kapitel 1.2.2, Seite 81). Eine Simulation von Dinukleosomen mit einer Verbindungslänge von 11 bp ergibt hingegen  $19,5 \pm 0,4$  nm. Trotzdem stimmen globale Eigenschaften der simulierten Fiber wie die Massenbelegungsichte, der Durchmesser, die Neigung der Nukleosomen und der DNA zur Fiberachse und die Persistenzlänge mit den experimentellen Ergebnissen überein. Hieraus ergibt sich, daß die lokale Eigenschaft des internukleosomalen Abstands keinen wesentlichen Einfluß auf die globalen Eigenschaften der Fiber hat und damit keine entscheidende Auswirkung auf die Ergebnisse des Modells besitzt.

Bei den Simulationen wurde eine physiologische Salzkonzentration von 100 mM NaCl angenommen. Bei in vitro-Experimenten beobachtet man bei abnehmender Salzkonzentration eine geringere Massenbelegungsichte. Dieser Effekt wird qualitativ vom hier entwickelten Modell wiedergegeben. Die Werte der Massenbelegungsichte stimmen quantitativ jedoch nicht überein. Ursache hierfür ist, daß bei geringerer Ionenstärke sich die Konfiguration der Nukleosomen verändert [Hol89] und der Wert von  $\alpha$  zunimmt. Weiterhin kann aufgrund der Konformationsänderung der Nukleosomen eine Veränderung von  $\beta$  erwartet werden. Darüberhinaus ist zu erwarten, daß sich die internukleosomale Wechselwirkung mit der Salzkonzentration verändert. Diese Effekte sind bisher jedoch nicht ausreichend untersucht worden, um sie in den bisherigen Rechnungen berücksichtigen zu können.

### Simulationen ohne Nukleosomenstamm

In hier durchgeführten Simulationen von Chromatin ohne Nukleosomenstamm werden die DNA-Segmente direkt an die Nukleosomen-Core-Partikel gekoppelt. Polynukleosomen, deren Repeat-Länge von 192 bp z.B. der von Chromatin in HeLa-Zellen entspricht, bilden in der Simulation Fibern, deren Massenbelegungsichte, Durchmesser, Neigung von DNA und Nukleosomen zur Fiberachse sowie Persistenzlänge mit den experimentellen Werten vereinbar sind.



**Abbildung 7.1:** Skizze zur Begründung, warum die Fiber bei Kopplung der Verbindungs-DNA an das Core-Partikel und großer Repeat-Länge nicht stabil ist.

Wird im Modell die Länge der Verbindungs-DNA wesentlich länger gewählt, so daß die Repeatlänge beispielsweise der von Chromatin von Hühnererythrozyten entspricht, bilden sich in den Simulationen keine Fibern. Stattdessen erhält man ungeordnete Konfigurationen. Ursache dafür ist, daß Nukleosomen, die auf der Fiber weiter entfernt sind, sich aufgrund der großen Flexibilität der langen Verbindungs-DNA zwischen andere Nukleosomen legen können (Abbildung 7.1). Erst das Vorhandensein des Nukleosomenstamms bei langer Repeatlänge ermöglicht also die Ausbildung einer Fiber.

### Schlußfolgerung

Bisherige Modelle der Chromatinfiber wie das Solenoid-Modell oder das Zick-Zack-Modell sind lediglich in der Lage, qualitative Aussagen über die Struktur der Fiber zu machen. Die Beschreibung bleibt rein geometrisch, ohne auf physikalischen Eigenschaften der Bestandteile einzugehen. Der Erkenntnisgewinn und die Anwendbarkeit dieser Modell bleibt damit gering. Das in dieser Arbeit vorgestellte Modell ist den bisherigen Modellen deut-

lich überlegen, da es experimentell bekannte Eigenschaften wie Durchmesser, Massenbelegungsdichte, Neigung der Nukleosomen und der DNA zur Fiberachse sowie die Persistenzlänge der Fiber quantitativ auf die physikalischen Eigenschaften der Konstituenten zurückzuführen. Es bildet die Basis für weiterführende Arbeiten, indem es die Möglichkeit bietet, den Einfluß beliebiger Faktoren auf die Fiber detailliert zu studieren und Hypothesen über die Mechanismen der Regulation der Packung der Fiber zu überprüfen.



# Ausblick

In dieser Arbeit wurde ein numerisches Computermodell der Chromatinfiber entwickelt. Es wurde nachgewiesen, daß das Modell in der Lage ist, experimentelle Eigenschaften der Fiber zu reproduzieren. Damit wurde gezeigt, daß das Modell den Aufbau der Fiber prinzipiell richtig beschreibt. Dies stellt einen wesentlichen Fortschritt im Verständnis der Chromatinfiber dar.

Die wesentlichen Eigenschaften werden richtig wiedergegeben – das Modell kann in weiten Bereichen jedoch noch verbessert werden. Die Abhängigkeit der Massenbelegungsdichte von der Salzkonzentration wird lediglich qualitativ wiedergegeben. Dies beruht auf bisher unvollständigen Kenntnissen über die Abhängigkeit der Winkel  $\alpha$  und  $\beta$  von der Salzkonzentration. Unbekannt ist ebenfalls die Abhängigkeit der Wechselwirkung der Nukleosomen von der Salzkonzentration. Es sind weitergehende theoretische und experimentelle Untersuchungen dieser Effekte auf molekularer Ebene notwendig, um sie angemessen im Modell berücksichtigen zu können.

Zwei Fragestellungen wurde von dieser Arbeit nicht bearbeitet: die Auswirkungen von Modifikationen der Histone auf die Packung sowie die Zeitbereiche, in denen diese Änderungen stattfinden.

Experimente zeigen, daß Modifikationen der Histone die Packungsdichte der Fiber verändern [Fel96]. Diese Effekte konnten jedoch bisher nicht in die Rechnungen integriert werden, da hierfür weitere Kenntnisse der Details über internukleosomale Wechselwirkungen erforderlich sind. Deswegen ist es notwendig, die Stärke der internukleosomalen Wechselwirkung experimentell zu bestimmen. Dafür kann beispielsweise ein Nukleosom an einer Oberfläche immobilisiert und ein anderes Nukleosom an einer Spitze eines Rasterkraft-Mikroskops befestigt werden. Hiermit könnte die Kraft zwischen den Nukleosomen als Funktion ihrer Entfernung bestimmt werden. Ergänzend hierzu können unter Verwendung der bekannten Struktur des Nukleosomen-Core-Partikels und der globularen Domäne des Linker-Histons theoretische Rechnungen dieses Systems mit Hilfe von Molekulardynamik durchgeführt werden.

Die Zeitbereiche, in denen Umordnungen in der Fiber bei Veränderung beispielsweise der internukleosomalen Wechselwirkung oder des Winkels  $\beta$  stattfinden, sind von Interesse, um ihren Einfluß auf die Kinetik von Regulationsprozessen der Zelle besser zu verstehen. Diese Zeiten könnten mit Hilfe eines Brownsche–Dynamik–Algorithmus anstatt des verwendeten Monte Carlo–Verfahrens berechnet werden [EM78, Fix78]. Das für diese Arbeit erstellte Computerprogramm ist für die Implementation dieses Verfahrens bereits ausgelegt.

# Anhang A

## Momente höherer Ordnung der idealen Zufallskette

In diesem Kapitel wird die Ableitung zu Gleichung 2.7 in Kapitel 2.1.1 angegeben.

Sei  $\vec{R}$  der End-zu-End-Abstand einer idealen Zufallskette mit Segmentlänge  $b_0$  (Gleichung 2.5). Dann gilt für  $\langle \vec{R}^2 \vec{R}^2 \rangle$ :

$$\begin{aligned} \frac{\langle \vec{R}^2 \vec{R}^2 \rangle}{b_0^4} &= \left\langle \left( \sum_{ij} \hat{r}_i \hat{r}_j \right) \left( \sum_{kl} \hat{r}_k \hat{r}_l \right) \right\rangle \\ &= \sum_{ijkl} \langle (\hat{r}_i \hat{r}_j) (\hat{r}_k \hat{r}_l) \rangle \\ &= \sum_{ij} \langle (\hat{r}_i \hat{r}_i) (\hat{r}_j \hat{r}_j) \rangle + \sum_{ij, i \neq j} \langle (\hat{r}_i \hat{r}_j) (\hat{r}_i \hat{r}_j) \rangle + \\ &\quad \sum_{ijkl \text{ sonst}} \langle (\hat{r}_i \hat{r}_j) (\hat{r}_k \hat{r}_l) \rangle \end{aligned} \quad (\text{A.1})$$

$\hat{r}_n$  sind dabei die normierten Verbindungsvektoren. Weiterhin gilt

$$\begin{aligned} \langle (\hat{r}_i \hat{r}_i) (\hat{r}_j \hat{r}_j) \rangle &= 1 \\ \langle (\hat{r}_i \hat{r}_j) (\hat{r}_k \hat{r}_l) \rangle &= 0 \quad ijkl \text{ sonst} \end{aligned} \quad (\text{A.2})$$

Für beliebige  $\hat{r}_i$  und  $\hat{r}_j$  mit  $i \neq j$ , wobei  $\varphi$  der Winkel zwischen den Vektoren ist, gilt:

$$\langle (\hat{r}_i \hat{r}_j)^2 \rangle = \frac{1}{2\pi} \int_0^{2\pi} \cos^2 \varphi d\varphi = \frac{1}{2} \quad (\text{A.3})$$

Einsetzen von Gleichung A.1 und A.2 in A.3 ergibt

$$\begin{aligned}\langle \vec{R}^2 \vec{R}^2 \rangle &= b_0^4 \left( N^2 + \frac{1}{2} N(N-1) \right) \\ &= b_0^4 \left( \frac{3}{2} N^2 - \frac{N}{2} \right)\end{aligned}\quad (\text{A.4})$$

Gleichung A.4 kann mit Gleichung 2.6 auch ausgedrückt werden wie

$$\langle \vec{R}^2 \vec{R}^2 \rangle - \left( \langle \vec{R}^2 \rangle \right)^2 = \frac{1}{2} N(N-1) b_0^4 \quad (\text{A.5})$$

Allgemein gilt [Kle93]

$$\langle R^{2l} \rangle = (-1)^l (2l+1)! \sum_{m_i} \prod_{i=1}^l \frac{1}{m_i!} \left[ \frac{N(2b_0)^{2i} (-)^i B_{2i}}{(2i)! 2i} \right] \quad (\text{A.6})$$

wobei  $B_i$  die Bernoullizahlen  $B_2 = 1/6$ ,  $B_4 = -1/30$ , ... sind und die Summe über  $m_i$  der Einschränkung

$$l = \sum_{i=1}^l i \cdot m_i \quad (\text{A.7})$$

unterliegt.

# Anhang B

## Formale Programmstruktur

### B.1 Übersicht UML

Die Unified Modeling Language (UML) von Grady Booch, Ivar Jacobson und James Rumbaugh ist der Standard der Notation der objektorientierten Modellierung. Sie stellt die Syntax für die Notation als graphische Beschreibungssprache bereit und ist von der Object Modeling Group (OMG)<sup>1</sup> standardisiert. Sie stellt kein Vorgehensmodell bereit. Im folgenden werden die zur Dokumentation des Simulations- und Auswertepackets verwendeten Diagramme beschrieben: Use-Case-Diagramm, Packetdiagramm, Sequenzdiagramm und Klassenstrukturdiagramm. Es existieren weitere Diagrammtypen (z.B. Verteilungsdiagramm, Aktivitätendiagramm, Zustandsdiagramm), auf die aber nicht weiter eingegangen werden soll. Die Darstellung bleibt auf Grund der Kürze zwangsläufig abstrakt. Eine vollständige Darstellung findet sich in den Standards [RU97b, RU97d, RU97a, RU97c] oder in Lehrbüchern [Bur97, BRJ99]. Seit kurzem existieren Werkzeuge, mit denen es möglich ist, direkt aus UML-Graphen Programm-Code zu erzeugen. Dazu gehören beispielsweise Rose von Rational, OTW von OWIS oder Together von Object International.

Zur Übersicht über die im folgenden vorgestellten Notationen sind in den Abbildungen B.1 und B.2 die wichtigsten graphischen Elemente der UML zusammengefasst. Diese Abbildungen sind in leicht modifizierter Form dem Lehrbuch von Booch, Jacobson und Rumbaugh [BRJ99] entnommen.

---

<sup>1</sup>Die OMG ist bekannt durch den CORBA-Standard.

### B.1.1 Use-Case-Diagramm

Use-Case-Diagramme zeigen die Funktionalität eines Systems oder einer Klasse. Ein Use-Case-Diagramm ist ein Graph von Aktoren und Use-Cases, die durch Kommunikation, Teilnahme oder Assoziation verbunden sind.

Use-Case kann am ehesten durch „Anwendungsfall“ übersetzt werden. Er stellt eine zusammenhängende Funktionseinheit dar, die durch Botschaften mit dem System und den Aktoren wechselwirkt. Das Symbol des Use-Case ist eine Ellipse, die den Namen des Use-Case trägt.

Ein Aktor ist eine Rolle eines Objekts, das direkt mit dem System und den Use-Cases interagiert. Ein Objekt kann mehrere Rollen spielen und deswegen durch verschiedene Aktoren dargestellt werden. Sein Symbol ist ein Rechteck mit der Bezeichnung (dem *Stereotyp*) „Aktor“ oder ein Strichmännchen mit dem Namen des Aktors darunter.

Es gibt mehrere Standard-Beziehungen zwischen Use-Cases sowie zwischen Use-Cases und Aktoren:

**kommuniziert** Die Teilnahme eines Aktors an einem Use-Case. Dies ist die einzige Beziehung eines Aktors zu einem Use-Case. Sie wird durch eine durchgezogene Linie dargestellt. Es ist möglich, eine Richtung durch eine einfache Pfeilspitze anzuzeigen.

**erweitert** Ein Use-Case beinhaltet das spezifische Verhalten eines anderen Use-Cases und erweitert es. Die Beziehung wird durch einen Generalisierungspfeil (ein Pfeil mit einem offenen Dreieck als Spitze) dargestellt, der mit dem Stereotyp „erweitert“ beschriftet ist.

**benutzt** Ein Use-Case benutzt einen anderen Use-Case. Insofern ist diese Funktionalität des anderen Use-Case in diesem Use-Case enthalten. Die Beziehung wird deswegen ebenfalls durch einen Generalisierungspfeil dargestellt, der mit dem Stereotyp „benutzt“ beschriftet wird.

### B.1.2 Packetdiagramm

Das Packetdiagramm gibt die Organisation des Systems wieder. Ein Packet stellt eine zusammenhängende Gruppe von Elementen eines Modells dar. Pakete können ineinander verschachtelt sein. Ein Packet kann sowohl andere Pakete als auch normale Elemente enthalten. Ein Element bzw. ein anderes Packet gehört genau einem Packet.

Ein Packet wird als großes Rechteck mit einem kleinen Rechteck an der oberen linken Ecke (dem „Reiter“) gezeichnet. Der Name befindet sich im großen oder kleinen Rechteck. Abhängigkeitsbeziehungen zwischen Packeten werden mit gestrichelten Pfeilen dargestellt.

### B.1.3 Sequenzdiagramm

Ein Sequenzdiagramm stellt eine Interaktion in ihrer zeitlichen Abfolge dar. Es zeigt die Objekte, die an den Interaktionen teilnehmen, anhand ihrer „Lebenslinie“ und die zeitliche Reihenfolge der Botschaften, die sie austauschen.

Ein Sequenzdiagramm besitzt zwei Dimensionen: Die vertikale Dimension repräsentiert die Zeit und die horizontale verschiedene Objekte. Normalerweise verläuft die Zeit von oben nach unten.

Ein Objekt wird als Rechteck dargestellt. Seine „Lebenslinie“ ist eine vertikale, gestrichelte Linie. Die Zerstörung des Objekts wird durch ein großes Kreuz ( $\times$ ) dargestellt.

Die Aktivierung eines Objekts zeigt die Zeitdauer an, in der das Objekt eine Aktion ausführt. Die Aktivierung wird durch ein hohes schmales Rechteck dargestellt, das auf der Lebenslinie des entsprechenden Objekts liegt. Es ist mit dem Namen der entsprechenden Aktion gekennzeichnet.

Ein Botschaft ist eine Kommunikation zwischen Objekten, die Information überträgt. Sie wird durch einen horizontalen, durchgezogenen Pfeil dargestellt. Der Pfeil ist mit dem Namen der Botschaft beschriftet. Ein Objekt kann auch an sich selber eine Botschaft schicken.

### B.1.4 Klassenstrukturdiagramm

Klassendiagramme stellen die statische Struktur eines Modells dar: Klassen und Typen, ihre innere Struktur und ihre Beziehungen zu anderen Klassen. Die zeitliche Entwicklung wird nicht gezeigt. Eng verwandt ist das Objektdiagramm, das den Zusammenhang von Instanzen darstellt. Ein Klassenstrukturdiagramm ist also ein Graph von Elementen, die mit verschiedenen statischen Beziehungen verbunden sind.

Eine Klasse ist eine Beschreibung für einen Satz von Objekten mit ähnlicher Struktur, Verhalten und Beziehungen. Sie wird durch ein durchgezogenes Rechteck dargestellt. Es kann bis zu drei Unterteilungen besitzen, die übereinander angeordnet sind. Die oberste trägt den Klassennamen, die mittlere die Liste der Attribute (Variablen der Klasse) und die unterste die Liste der Operationen (Funktionen der Klasse). Häufig wird nur der Name eingetragen und auf die anderen beiden Teile verzichtet. Attribute und Operationen können eine Sichtbarkeitskennung tragen: + für public, # für protected und – für private.

Ein Template ist die Beschreibung einer Klasse mit einem oder mehreren formalen Parametern. Es ist nicht direkt benutzbar. Ein kleines gestricheltes Rechteck in der oberen rechten Ecke eines Symbols für eine Klasse signalisiert, daß es sich um eine Template-Klasse handelt. Es trägt die Liste der formalen

Parameter. Die Instanzierung eines Templates wird durch einen gestrichelten Pfeil dargestellt.

Die Vererbung von Klassen, auch Generalisierung genannt, ist die Beziehung zwischen einer allgemeineren und einer spezielleren Klasse, die mit der allgemeineren konsistent ist und diese um zusätzliche Eigenschaften erweitert. Sie wird durch einen durchgezogenen Pfeil mit einem offenen Dreieck als Spitze dargestellt, die in Richtung der Basis-Klasse zeigt.

Mit Assoziation bezeichnet man allgemein die Beziehung zwischen Klassen. Sie wird durch einen durchgezogene Linie dargestellt. Richtungen der Assoziation können durch Pfeilspitzen angegeben werden. Die Assoziation kann durch eine Beschriftung des Pfeils beschrieben werden. Aggregation, die enge Zusammengehörigkeit mehrerer Klassen, kann durch eine offene Raute am Anfang des Pfeils ausgedrückt werden. Die Komposition ist eine besonders enge Aggregation, die sich auch auf die Lebenslinie bezieht. Sie wird durch eine ausgefüllte Raute dargestellt. Die Anzahl der assoziierten Instanzen einer Klasse kann durch die Angabe „*minimale Anzahl .. maximale Anzahl*“ an der Pfeilspitze notiert werden.

## B.2 Struktur des Simulations- und Analysepackets

In folgenden werden die wichtigsten Strukturen des Simulations- und Analysepackets in UML-Diagrammen vorgestellt. Um die Darstellung übersichtlich zu halten, wird nicht die komplette Struktur dargestellt, sondern nur die wesentlichen Aspekte.

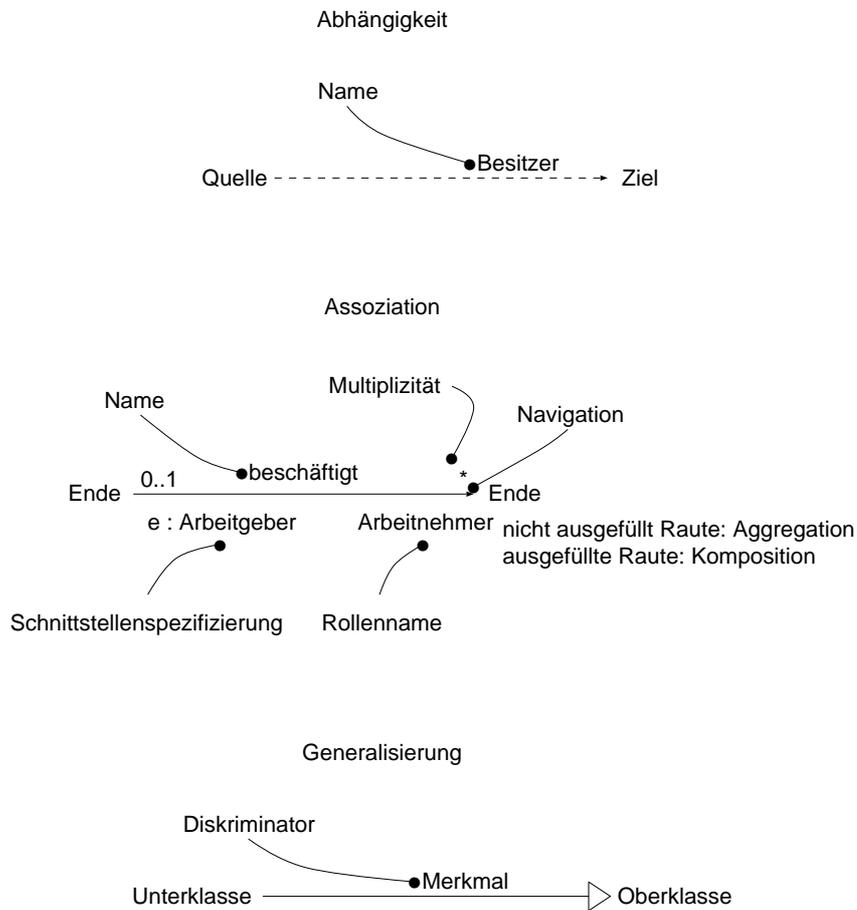
### B.2.1 Use-Case-Diagramm

Das Use-Case-Diagramm (Abbildung B.3) beschreibt die Anwendungsfälle, die die Konfiguration betreffen. Unter Konfiguration versteht man dabei das Objekt, in dem sich die die Konfiguration beschreibenden Parameter befinden. Dazu gehören beispielsweise die Koordinaten der Kette oder die Gesamtenergie. Zu den wichtigsten Anwendungsfällen gehört das Einlesen und Herausschreiben der Konfiguration, die Berechnung der Gesamtenergie und die Durchführung der Monte Carlo-Schritte.

### B.2.2 Packetdiagramm

Das System wurde in drei Teile unterteilt (Abbildung B.4): Das Paket support, in dem sich die Programmelemente (Klassen und Funktionen) befinden,





**Abbildung B.2:** Die wichtigsten graphischen Elemente der UML für die Darstellung von Beziehungen zwischen Elementen.

die auch in anderen Programmen ohne weiteres benutzbar sind oder bereits benutzt worden sind, das Paket `base` in dem sich die Elemente befinden, die zur Simulation benutzt werden, und das Paket `analysis`, in dem sich die Elemente befinden, die zur Analyse der Daten, die bei der Simulation erzeugt wurden, benutzt werden.

Elemente des Pakets `base` benutzen verschiedene Elemente der C++-Standard-Bibliothek `STL` (*Standard Template Library*) wie z.B. verkettete Listen. Elemente des Pakets `analysis` benutzen Klassen der GNU C++ Library wie z.B. die Statistik-Klasse `SmplStat`.

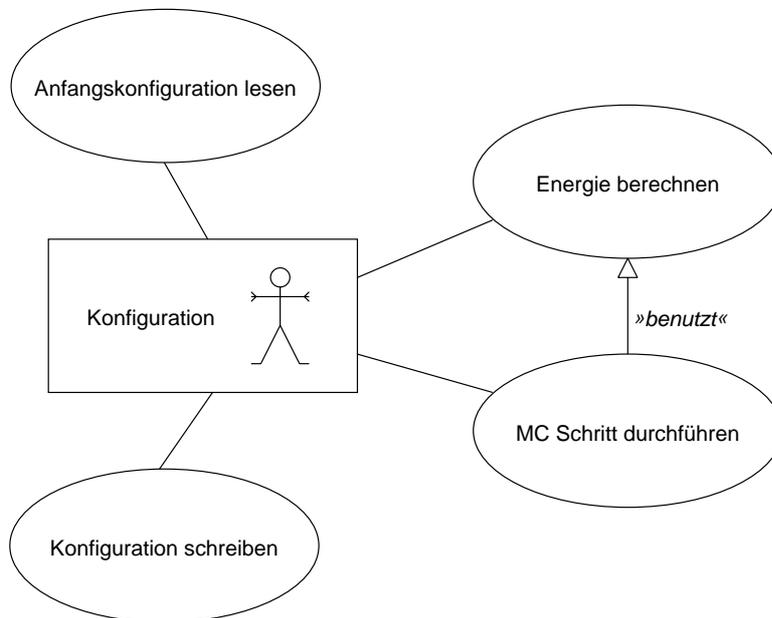


Abbildung B.3: Use-Case-Diagramm des Systems

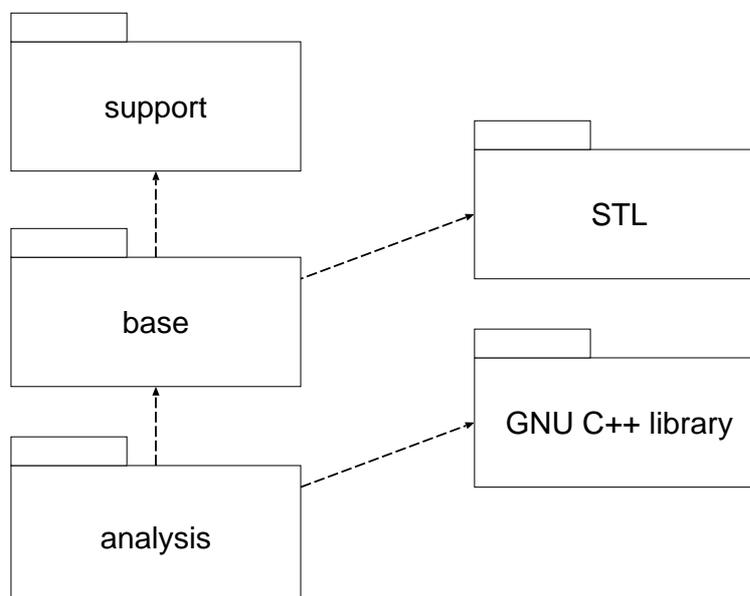


Abbildung B.4: Paketdiagramm des Systems

## B.2.3 Klassendiagramme

### Modul support

Das Modul **support** (Abbildung B.5) besteht aus den Hilfsklassen **Vector3D** (dreidimensionaler Vektor), **Matrix** ( $3 \times 3$ -Matrix), der String-Klasse **Strng** und **XError**, einer Klasse, die zur Ausnahmebehandlung verwendet wird. Die Klassen **RBuiltin**, **R16807** und **R250** enthalten die Implementationen der verschiedenen Zufallszahlengeneratoren (R steht für *random number generator*), die in Kapitel 4.4 beschrieben worden sind. Die Trajektorie-Klassen **TrjIndex**, **TrjEntry** und **TrjEntryIterator** werden zum Zugriff auf Trajektorien (**Trj**) auf der Platte verwendet. Sie stellen eine Weiterentwicklung der von C. Munkel entwickelten Klassen dar. Die Daten werden in einer hierarchischen, doppelt-verketteten Liste gehalten. Die Elemente sind die Instanzen der Klasse **TrjEntry**. Der Zugriff ist besonders performant, da ein ein **Index-File** (**TrjIndex**) direkte Sprünge im Trajektorie-File erlaubt. Der **TrjEntryIterator** stellt einen Zugriff nach dem **Iterator-Muster** bereit.

### Modul base

Im Modul **base** (Abbildung B.6) werden alle Strukturen definiert, die zur Simulation benötigt werden. Alle Daten der Segmente werden in Instanzen von Klassen abgelegt, die von der Klasse **Bead** abgeleitet wurden sind. Die Klasse **Bead** hält die Koordinaten und Verbindungsvektoren. In der Klasse **RealBead** kommt das lokale Koordinatensystem hinzu, das zur Berechnung der Torsion benötigt wird. **DNABead**, **UnchargedDNABead** und **NucleosomeBead** sind die in Simulationen verwendeten Beads. Beads, die sich in Seitenketten befinden sollen, müssen von der Klasse **BranchBead** abgeleitet werden. Diese Eigenschaft wird im vorhanden Programm nicht verwendet.

Die Beads werden in einer doppelt verketteten, hierarchischen Liste gehalten, der *Branched Linked List* **BLL** aus Elementen der Klasse **BLLLink**. Sie ist als Template implementiert, um auch in anderen Projekten verwendbar zu sein. Es existieren drei Iteratoren, die nacheinander Zeiger auf einzelne Beads, auf Paare von Beads und auf Tripel von Beads liefern. Die Kette wird von der Klasse **StandardChainBuilder** aus den Daten der Trajektorie, die mit Hilfe von **TrajectoryIndex** eingelesen werden, aufgebaut. Falls andere Fileformate verwendet werden sollen, kann dies durch eine andere Klasse geschehen, die ebenfalls von der Klasse **ChainBuilder** abgeleitet ist.

Die Kette wird in der Klasse **Configuration** gehalten. Die Energie-Berechnung erfolgt mit der Klasse **TotEnergy**, in der die gesamten Energie-Objekte aggregiert sind. Die Energie-Objekte berechnen unter Verwendung der **BLLI-**

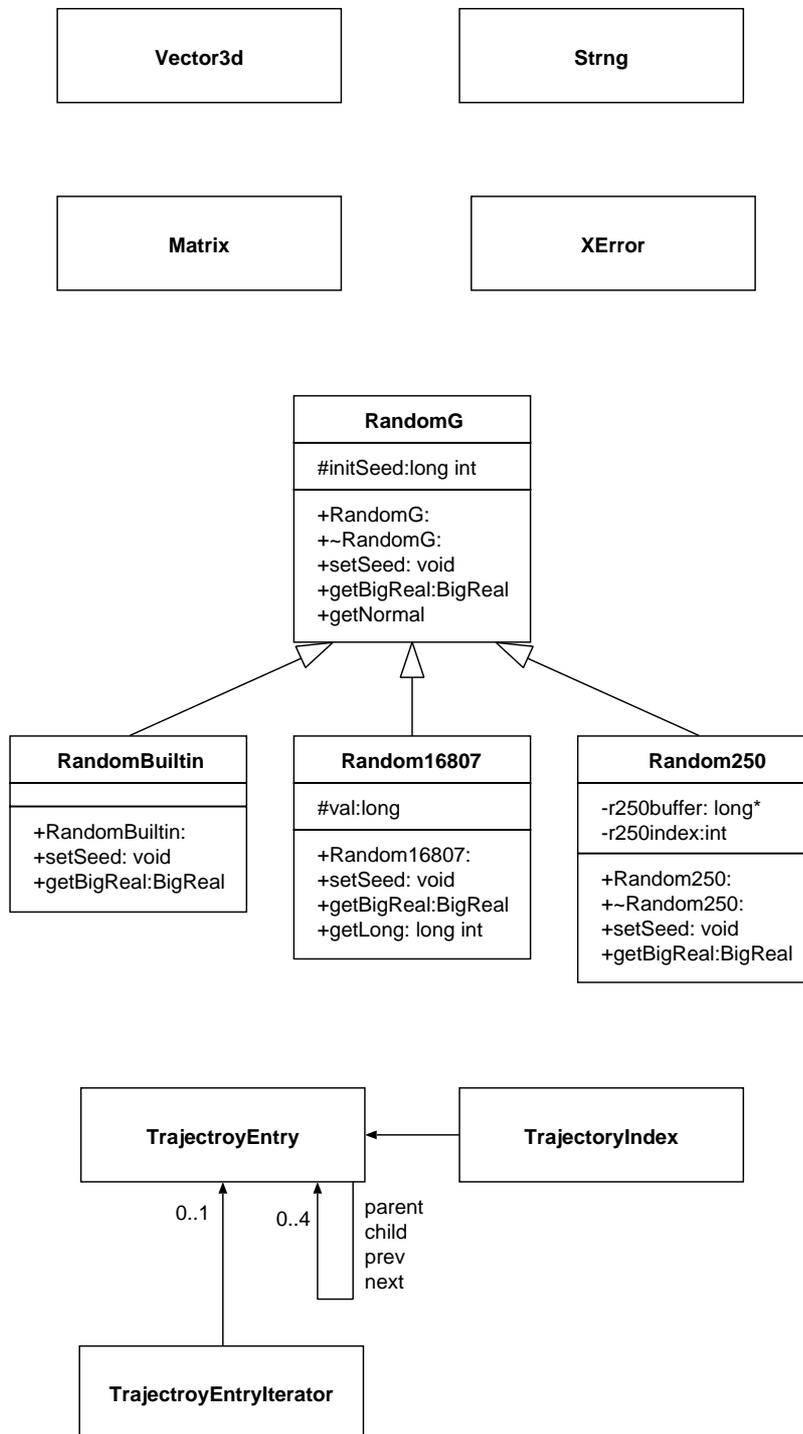


Abbildung B.5: Klassendiagramm des Moduls support



teratoren (Benutzung nicht eingezeichnet) und der lokalen Energie-Klassen<sup>2</sup> (abgeleitet von `LocEnergy`) die totale Energie.

`MCmoves` implementiert den eigentlichen Monte Carlo-Algorithmus (Kapitel 2.3). Die davon abgeleiteten Klassen `MCPivot` und `MCBeadR` implementieren die tatsächlichen Monte Carlo-Schritte, den Pivot-Schritt und den Rotations-Schritt (Kapitel 3.3).

Der Ablauf der Simulation wird vom `SimController` koordiniert. Eine Referenz auf das Objekt der `Parameter`-Klasse muß fast allen Objekten im Konstruktor übergeben werden. Es enthält die gesamten Parameter, die gewählt werden können, wie Temperatur oder die Wechselwirkungsparameter.

Im Klassenstrukturdiagramm des Moduls `base` erkennt man Strukturen des Use-Case-Diagramms wieder, wenn man die Konfiguration mit der Klasse `Configuration` identifiziert, `TotEnergy` mit der Energieberechnung, und `Mcmoves` bzw. die davon abgeleiteten Klassen mit den Monte Carlo-Schritten und das Einlesen der Konfiguration mit `ChainBuilder`. Lediglich das Schreiben der Konfiguration wurde in die Klasse `Configuration` integriert.

## Modul analysis

Das Modul `analysis` (Abbildung B.7) stellt die Klasse `AnaController` bereit, die von der Klasse `Controller` aus dem Modul `base` abgeleitet wurde. Es implementiert die Verarbeitung der Befehlszeile (*Parsen*), das Einlesen des Trajektorien-Files in `Configuration` und eine Schleife über die Konfigurationen der Trajektorien. Die Analyse-Controller (es sind nicht alle abgebildet) implementieren dann je nach Anwendung nur die Analyse einer Einzel-Konfiguration oder die Schleife.

Mit dem Template `autoCorrelatorT` kann die Autokorrelationsfunktion beliebiger Klassen berechnet werden, wie von Zahlen (`Real`) oder Vektoren (`Vector3d`). Dies ist möglich, in dem nicht nur Objekt übergeben wird, sondern auch eine Klasse angegeben wird, die eine Funktion besitzt, die das Produkt von zwei Objekten bildet.

## B.2.4 Sequenzdiagramme

Der Ablauf der verschiedenen Prozesse im Simulationspaket soll nur exemplarisch an den wichtigsten Abläufen skizziert werden.

Der Ablauf des Simulations-Programmes wird vom Simulationskontrolller `SimController` gesteuert. Der Ablauf der Hauptroutine `run()` ist im Sequenzdiagramm in Abbildung B.8 skizziert. Der Simulationskontrolller instanziert

---

<sup>2</sup>„lokal“ ist hier als Gegensatz zu „global“ zu verstehen

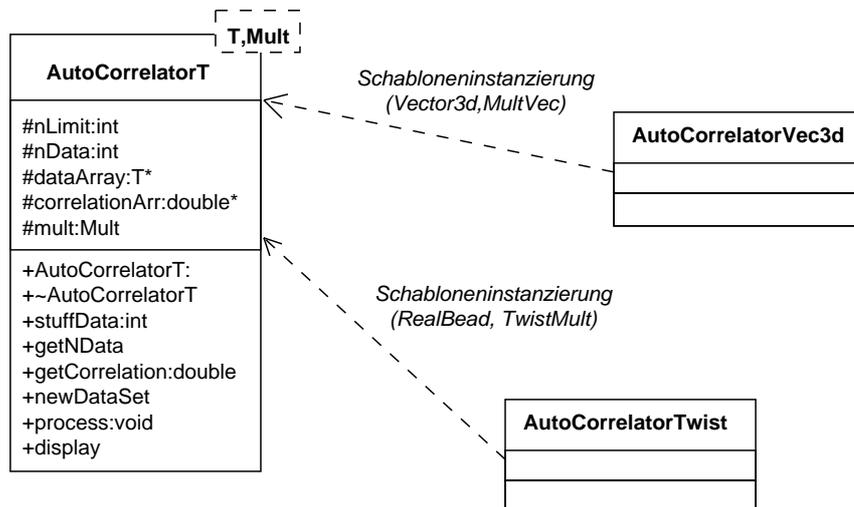
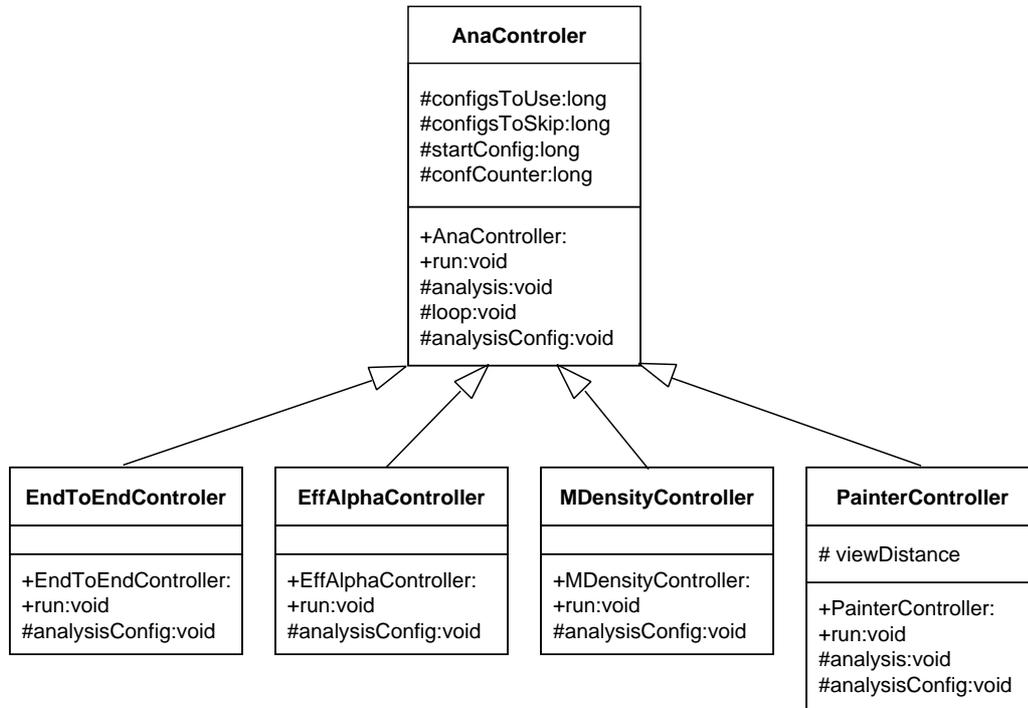


Abbildung B.7: Klassendiagramm des Moduls analysis

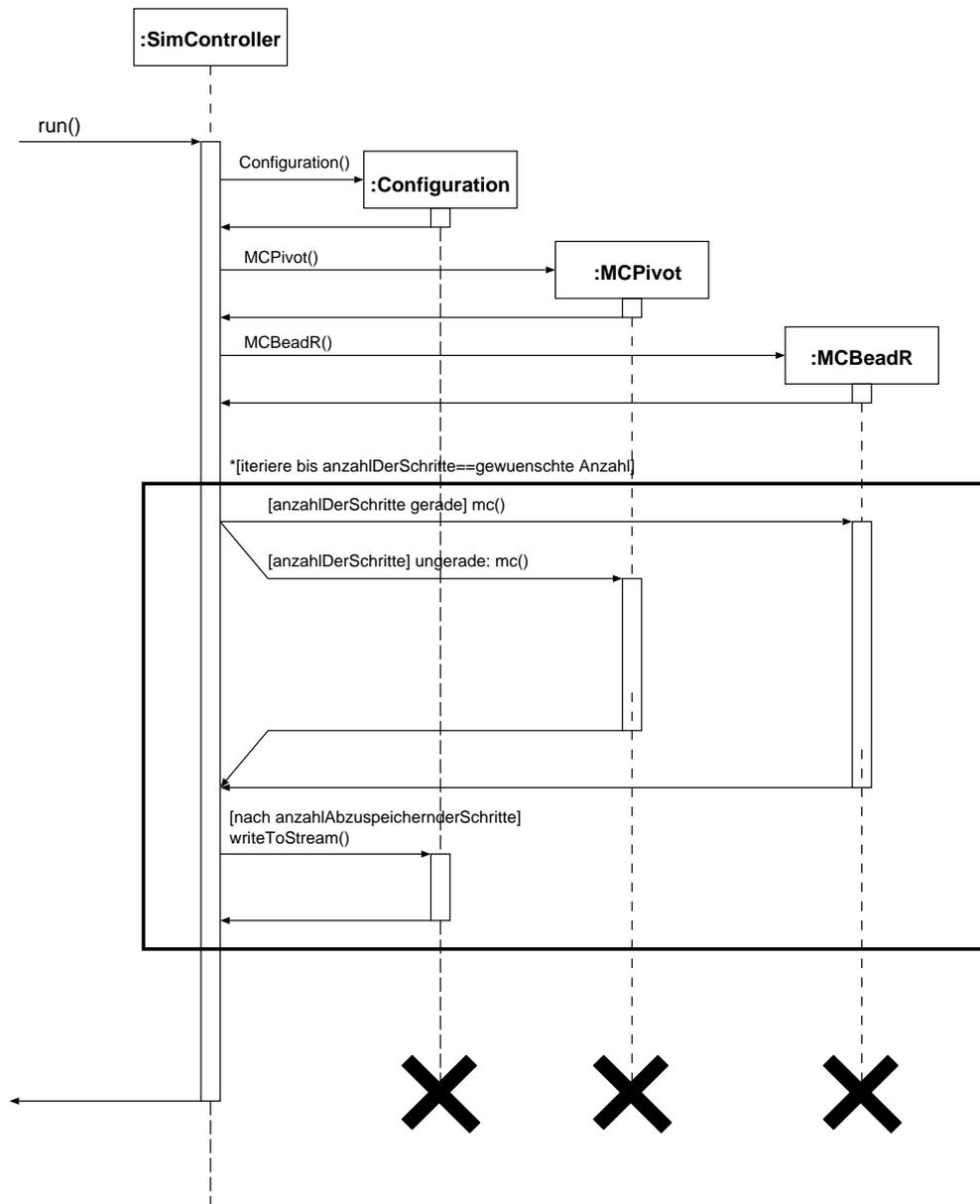
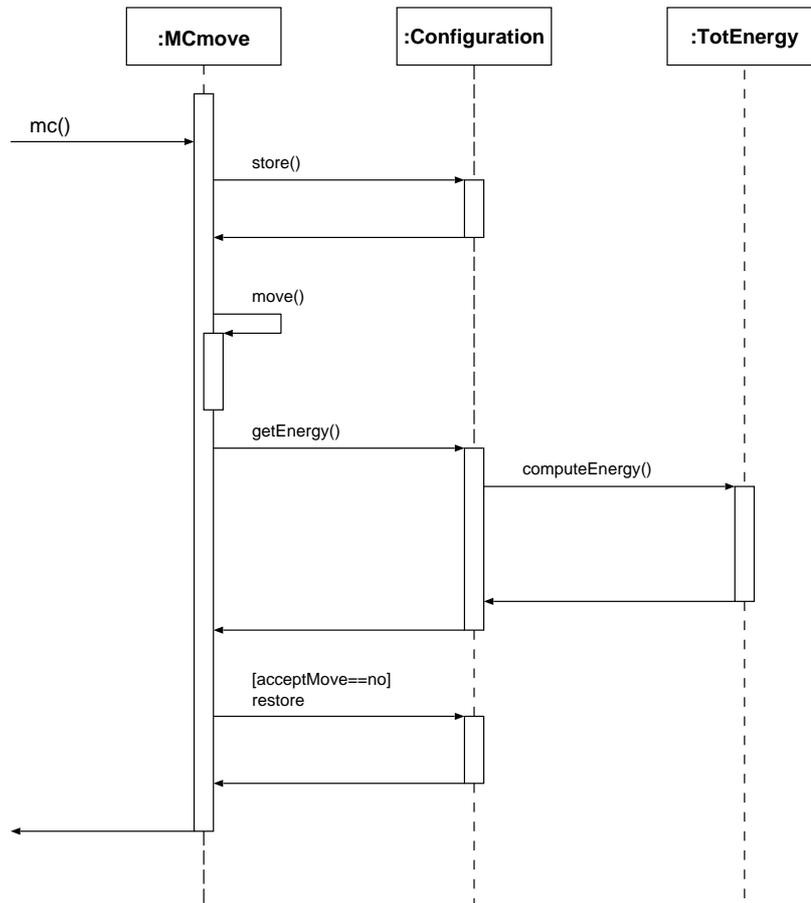


Abbildung B.8: Sequenzdiagramm der Funktion `run()` der Klasse `SimController`

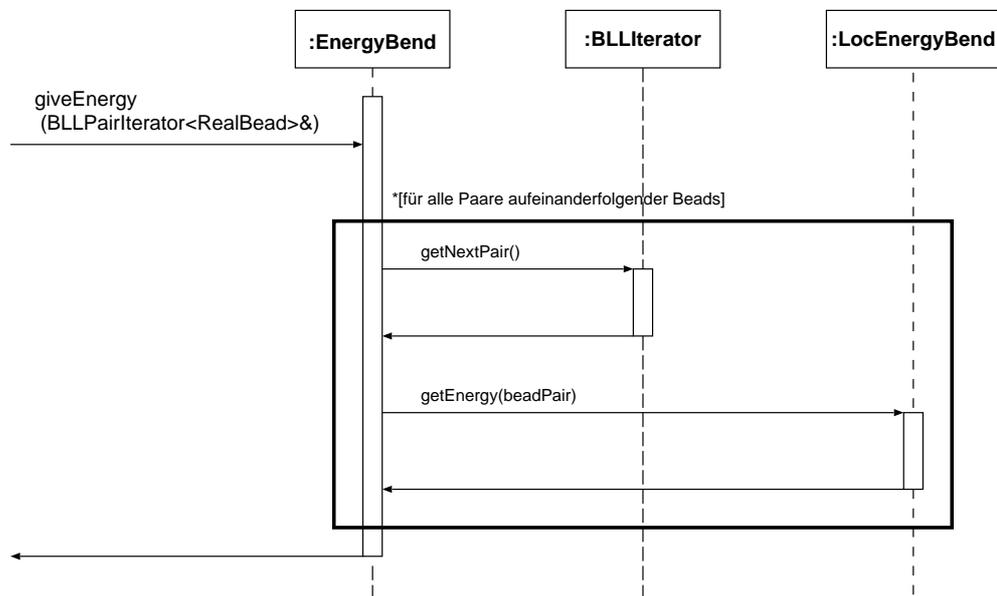
alle benötigten Objekte. In einer Schleife werden dann abwechselnd Pivot- und Rotations-Schritte durchgeführt, bis die vorgegebene Anzahl von Schritten erreicht ist.



**Abbildung B.9:** Sequenzdiagramm der Funktion `mc()` der Klasse `MCmove`

Das Sequenzdiagramm in Abbildung B.9 gibt den Ablauf eines Monte Carlo-Schritts wieder. Zunächst wird die ursprüngliche Konfiguration gesichert. Danach werden die Beads versetzt. Diese Versetzung ist in abgeleiteten Klassen implementiert. Anschließend wird die Energie der neuen Konfiguration bestimmt. Anschließend wird geprüft, ob die neue Konfiguration akzeptiert wird, oder ob sie verworfen wird und die ursprüngliche Konfiguration wiederhergestellt wird.

Exemplarisch für die Berechnung der Energieberechnung wird in Abbildung B.10 der Ablauf der Berechnung der Biegeenergie wiedergegeben. Der



**Abbildung B.10:** Sequenzdiagramm der Funktion `computeEnergy()` der Klasse `EnergyBend`

übergebene Iterator liefert Paare von Beads. Anschließend wird die elastische Energie zwischen beiden Segmenten berechnet. Dies geschieht für alle Paare von Beads, die benachbart sind und die elastisch wechselwirken.



# Abbildungsverzeichnis

1.1	Lehrbuchsicht der Kompaktierung der DNA im Zellkern . . . . .	2
1.2	Beispiel eines Nukleotids (DNA-Monomer): Desoxythymidin-5'-triphosphat (dTTP) . . . . .	3
1.3	Zwei Basenpaare einer DNA . . . . .	3
1.4	Räumliches Modell einer DNA-Doppelhelix . . . . .	4
1.5	Skizze des Nukleosoms . . . . .	6
1.6	Atomare Struktur des Nukleosom-Core-Partikels . . . . .	7
1.7	Struktur von Nukleosomen aus Hühner-Erythrozyten in einer cryo-elektronenmikroskopischen Aufnahme . . . . .	8
1.8	Struktur von Chromatinfibern aus Hühner-Erythrozyten in einer cryo-elektronenmikroskopischen Aufnahme . . . . .	10
1.9	Experimentell bestimmte Massenbelegungsdichte von Chromatin-Fibern in Abhängigkeit von der Salzkonzentration . . . . .	13
1.10	Sedimentationskoeffizienten von Nukleosomen-Oligomere verschiedener Länge . . . . .	14
1.11	Skizze des Solenoid-Modells . . . . .	19
1.12	Definition der Winkel $\alpha$ und $\beta$ . . . . .	20
1.13	Chromatinstrukturen nach dem Zick-Zack-Modell . . . . .	21
1.14	Chromatinstrukturen nach dem Zick-Zack-Modell mit $\alpha=37^\circ$ und $\beta=0^\circ$ . . . . .	22
2.1	Skizze einer diskotisch columnar hexagonalen Flüssigkristallphase . . . . .	28
3.1	Zur Diskretisierung der DNA . . . . .	34
3.2	Zur Definition des Biegungsvektors $\vec{B}_i$ . . . . .	35
3.3	Zur Diskretisierung des Nukleosoms . . . . .	35
3.4	Modelle der Kopplung von DNA und Nukleosom . . . . .	36
3.5	Geometrischen Struktur des Modells der Chromatinfiber . . . . .	37
3.6	Skizze eines Pivot-Schritts . . . . .	44
3.7	Skizze eines Rotations-Schritts . . . . .	45
3.8	Approximation einer Scheibe durch Kugeln . . . . .	47

3.9	Skizze zur Berechnung der Länge der DNA im Stamm . . . . .	49
4.1	Beschleunigung des parallelisierten Programms mit der Anzahl der Prozessoren . . . . .	58
5.1	Schematische Skizze zur Berechnung des Rückgrats der Fiber .	61
5.2	Skizze zur Berechnung der Anzahl der Nukleosomen, die zum Rückgrat der Fiber gehören. . . . .	62
6.1	Autokorrelationsfunktion der Energie einer Simulation von 100 Nukleosomen mit $\alpha=26^\circ$ , $\beta=110^\circ$ . . . . .	68
6.2	Tatsächlicher Öffnungswinkel in Abhängigkeit vom vorgegebenen Öffnungswinkel . . . . .	70
6.3	Massenbelegungsichte eines simulierten Polynukleosoms bei $\alpha=26^\circ$ in Abhängigkeit vom Verdrehungswinkel $\beta$ zwischen zwei aufeinanderfolgenden Nukleosomen. . . . .	71
6.4	Visualisierung einer simulierten Chromatinfiber mit $\alpha=26^\circ$ , $\beta=110^\circ$ . . . . .	72
6.5	Entwicklung der Energie und der Massenbelegungsichte eines simulierten Polynukleosoms im Verlauf einer Simulation, bei der mit einer gestreckten Konfiguration begonnen wurde. . . .	74
6.6	Kompaktierung eines Polynukleosoms ausgehend von einer gestreckten Konfiguration. . . . .	76
6.7	Mittelwert und Standardabweichung der Massenbelegungsichte eines simulierten Polynukleosoms in Abhängigkeit vom Parameter der Stärke der internukleosomalen Wechselwirkung $\epsilon_0$ . . . . .	77
6.8	Massenbelegungsichte eines simulierten Polynukleosoms in Abhängigkeit von der Salzkonzentration . . . . .	78
6.9	Winkel zwischen einlaufender und auslaufender Linker-DNA eines simulierten Polynukleosoms in Abhängigkeit von der Salzkonzentration . . . . .	79
6.10	Persistenzlänge einer Fiber mit $\alpha=26^\circ$ in Abhängigkeit von $\beta$ und dem Parameter der Stärke der internukleosomalen Wechselwirkung $\epsilon_{nuc}$ . . . . .	80
6.11	Mittlerer Durchmesser simulierter Polynukleosomen in Abhängigkeit von der Länge der Verbindungs-DNA . . . . .	81
6.12	Massenbelegungsichte eines simulierten Polynukleosoms mit einer Länge Linker-DNA von 33 bp bei $\alpha=26^\circ$ und $\beta=60^\circ$ in Abhängigkeit von der Ladungsichte der Linker-DNA. . . . .	82

---

6.13	Massenbelegungsichte eines simulierten Polynukleosoms mit einer Länge Linker-DNA von 22 bp und 33 bp bei $\alpha=26^\circ$ in Abhängigkeit vom Verdrehungswinkel $\beta$ zwischen Nukleosomen. Die DNA ist an den Nukleosomen-Stamm gekoppelt. . . . .	83
6.14	Visualisierung von simulierten Chromatinfibern mit Verbindungs-DNA der Länge 22 bp und 33 bp. Die DNA ist an den Nukleosomen-Stamm gekoppelt. . . . .	84
6.15	Massenbelegungsichte eines simulierten Polynukleosoms mit einer Verbindungs-DNA der Länge 15 bp bei $\alpha=30^\circ$ in Abhängigkeit von $\beta$ . Die DNA ist an den Nukleosomen-Core gekoppelt.	85
6.16	Visualisierung einer simulierten Chromatinfiber mit einer Verbindungs-DNA der Länge 40 bp, bei der die DNA an das Nukleosomen-Core gekoppelt ist . . . . .	86
7.1	Skizze zur Begründung, warum die Fiber bei Kopplung der Verbindungs-DNA an das Core-Partikel und großer Repeat-Länge nicht stabil ist. . . . .	92
B.1	Die wichtigsten graphischen Elemente der UML für die Darstellung von Programm Elementen. . . . .	103
B.2	Die wichtigsten graphischen Elemente der UML für die Darstellung von Beziehungen zwischen Elementen. . . . .	104
B.3	Use-Case-Diagramm des Systems . . . . .	105
B.4	Packetdiagramm des Systems . . . . .	105
B.5	Klassendiagramm des Moduls <code>support</code> . . . . .	107
B.6	Klassendiagramm des Moduls <code>base</code> . . . . .	108
B.7	Klassendiagramm des Moduls <code>analysis</code> . . . . .	110
B.8	Sequenzdiagramm der Funktion <code>run()</code> der Klasse <code>SimController</code>	111
B.9	Sequenzdiagramm der Funktion <code>mc()</code> der Klasse <code>MCmove</code> . . .	112
B.10	Sequenzdiagramm der Funktion <code>computeEnergy()</code> der Klasse <code>EnergyBend</code> . . . . .	113



# Tabellenverzeichnis

3.1	Debye Längen von Wasser ( $\varepsilon=80$ ) bei verschiedenen Konzentrationen von einwertigen Ionen . . . . .	40
3.2	In den Simulationen benutzte Parameter . . . . .	45
3.3	Länge der Linker-DNA und der Länge der Verbindungs-DNA für verschiedene Zelltypen . . . . .	50
4.1	Elementare und abgeleitete Einheiten im GROMACS-Einheitensystem . . . . .	55
4.2	Konstanten im GROMACS-Einheitensystem . . . . .	55
4.3	Mittelwert und Standardabweichung des End-zu-End-Abstands von simulierten Ketten von 25 Nukleosomen unter Verwendung verschiedener Zufallszahlengeneratoren. . . . .	56
4.4	Zeitpunkte der formalen und inhaltlichen Verifikation beim Prozeß des Programmtests. . . . .	59
5.1	Benutzte Parameter des Algorithmus zur Berechnung der Konturlänge. . . . .	63
6.1	Korrelationszeiten in Anzahl der Monte Carlo-Schritte von verschiedenen Eigenschaften . . . . .	67
6.2	Mittlerer Radius und Massenbelegungsdichte von Polynukleosomen mit $\alpha=26^\circ$ und $\beta$ normalverteilt mit Schwerpunkt bei $110^\circ$ und verschiedenen Verteilungsbreiten . . . . .	71
6.3	Mittelwert und Standardabweichung des End-zu-End-Abstands und der Massenbelegungsdichte in Abhängigkeit von der Startkonfigurationen. . . . .	75
6.4	Persistenzlänge von simulierten Polynukleosomen mit $\alpha=26^\circ$ und $\beta$ gaußverteilt mit Schwerpunkt bei $110^\circ$ bei verschiedenen Verteilungsbreiten . . . . .	79



# Literaturverzeichnis

- [Amd67] AMDAHL, G. M.: Validity of the single-processor approach to achieving large-scale computing capabilities. In: *Proceedings of the AFIPS Conference, Vol. 30*. New York, 1967
- [AT87] ALLEN, M. P.; TILDESLEY, D. J.: *Computer Simulation of Liquids*. Oxford, UK : Oxford University Press, 1987
- [AT93] ALLEN, M.P.; TILDESLEY, D.J.: *NATO ASI Series / Series C: Mathematical and Physical Sciences*. Bd. 397: *Computer Simulation in Chemical Physics*. Kluwer Academic Publishers, 1993
- [BB79] BAUMGÄRTNER, A.; BINDER, K.: Monte Carlo studies on the freely jointed polymer chain with excluded volume interaction. In: *J. Chem. Phys.* 71 (1979), Nr. 6, S. 2541–2545
- [BH88] BINDER, Kurt; HEERMANN, Dieter: *Monte Carlo Simulation in Statistical Physics. 2<sup>nd</sup>*. Berlin : Springer, 1992 (1988)
- [BHDW95] BEDNAR, Jan; HOROWITZ, Rachel A.; DUBOCHET, Jacques; WOODCOCK, Christopher L.: Chromatin Conformation and Salt-induced Compaction: Three dimensional structural information from cryoelectron microscopy. In: *The Journal of Cell Biology* 131 (1995), Nr. 6, S. 1365–1376
- [BHG<sup>+</sup>98] BEDNAR, Jan; HOROWITZ, Rachel; GRIGORIEV, Sergei; CARUTHERS, Lenny; HANSEN, Jeffrey; KOSTER, Abraham; WOODCOCK, Christopher: Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compactation of chromatin. In: *Proc. Natl. Acad. Sci.* 95 (1998), S. 14173–14178
- [BP72] BERNE, Bruce; PECHUKAS, Philip: Gaussian Model Potentials for molecular Interactions. In: *J. Chem. Phys.* 56 (1972), Nr. 8, S. 4213–4216

- [BPGK<sup>+</sup>86] BORDAS, J.; PEREZ-GRAU, L.; KOCH, M.H.; VEGA, MC; NAVE, C: The superstructure of chromatin and its condensation mechanism. I. Synchrotron radiation X-ray scattering results. In: *Eur. Biophys. J.* 13 (1986), Nr. 3, S. 157–173
- [BRJ99] BOOCH, Grady; RUMBAUGH, Jim; JACOBSON, Ivar: *Das UML-Benutzerhandbuch*. Bonn : Addison-Wesley, 1999
- [Bro78] BROWN, I.R.: Postnatal appearance of short DNA repeat length in neurons of the cerebral cortex. In: *Biochem. Biophys. Res. Comm.* 84 (1978), Nr. 2, S. 285–292
- [Bur97] BURKHARDT, Reiner: *UML – Unified Modeling Language*. Bonn : Addison Wesley Longman, 1997
- [But84] BUTLER, P.J.G.: A defined structure of the 30 nm chromatin fibre which accomodates different nucleosomal repeat lengths. In: *EMBO Journal* 3 (1984), Nr. 11, S. 2599–2604
- [Cas94] CASTRO, C.: *Measurement of the elasticity of single chromatin fibers: the effect of histone H1*, University of Oregon, Eugene, Dissertation, 1994
- [CCP78] CAMPBELL, A.M.; COTTER, R.I.; PARDON, J.F.: Light scattering measurements supporting helical structures for chromatin in solution. In: *Nucleic Acids Research* 5 (1978), Nr. 5, S. 1571–1580
- [Chi96] CHIRICO, Giuseppe: Torsional-Bending Infinitesimal Dynamics of a DNA Chain. In: *Biopolymers* 38 (1996), S. 801–811
- [CK90] CLARK, David J.; KIMURA, Takeshi: Electrostatic Mechanism of Chromatin Folding. In: *J. Mol. Biol.* 211 (1990), S. 883–896
- [CL94] CHIRICO, Giuseppe; LANGOWSKI, Jörg: Kinetics of DNA Supercoiling Studied by Brownian Dynamics Simulation. In: *Biopolymers* 34 (1994), S. 415–433
- [CL96] CHIRICO, Giuseppe; LANGOWSKI, Jörg: Brownian Dynamics Simulations of Supercoiled DNA with Bend Sequences. In: *Biophys. J.* 71 (1996), S. in press
- [CLR90] CORMEN, Thomas H.; LEISERSON, Charles E.; RIVEST, Ronald L.: *Introduction to Algorithms*. Cambridge, MA : MIT Press, 1990

- [Cor48] CORNER, J.: The second virial coefficient of a gas of non-spherical molecules. In: *Proc. R. Soc. London A* 192 (1948), S. 275–292
- [DAC<sup>+</sup>88] DUBOCHET, Jaques; ADRIAN, Marc; CHANG, Jiin-Ju; HOMO, Jean-Claude; LEPAULT, Jean; MCDOWALL, Alasdair W.; SCHULZ, Patrick: Cryo–Electron microscopy of vitrified specimens. In: *Quarterly Review of Biophysics* 21 (1988), Nr. 2, S. 129–228
- [DE86] DOI, M.; EDWARDS, S.: *The theory of polymer dynamics*. New York : Oxford University Press, 1986
- [DGS97] DELROW, Jeffrey; GEBE, Jahn; SCHURR, J. M.: Comparison of Hard–Cylinder and Screened Coulomb Interactions in the Modeling of Supercoiled DNAs. In: *Biopolymers* 42 (1997), S. 455–470
- [DMP90] DIMITROV, S.I.; MAKAROV, V.L.; PASHEV, I.G.: The Chromatin Fiber: Structure and Conformational Transitions as Revealed by Optical Anisotropy Studies. In: *Journal of Biomolecular Structure and Dynamics* 8 (1990), Nr. 1, S. 23–35
- [Ehr96] EHRLICH, Lutz: *Simulation der Struktur und Dynamik von Chromatin mit Brownscher Dynamik*. 1996. – Diplomarbeit, Universität Heidelberg
- [ELW94] EMERSON, A. J. P.; LUCKHURST, G. R.; WHATLING, S. G.: Computer–simulation studies of anisotropic systems. XXIII. The Gay–Berne discogene. In: *Molecular Physics* 82 (1994), Nr. 1, S. 113–124
- [EM78] ERMAK, D.L.; MCCAMMON, J.A.: Brownian dynamics with hydrodynamic interactions. In: *J. Chem. Phys.* 69 (1978), Nr. 4, S. 1352–1360
- [EMCL97] EHRLICH, Lutz; MÜNKEL, Christian; CHIRICO, Giuseppe; LANGOWSKI, Jörg: A Brownian dynamics model for the chromatin fiber. In: *Comput. Appl. Biosci.* 13 (1997), S. 271–279
- [EST92] ENGH, G. van d.; SACHS, R.K.; TRASK, B.J.: Estimating genomic distance from DNA sequence location in cell nuclei by a random walk model. In: *Science* 257 (1992), S. 1410–1412

- [Fel96] FELSENFEST, G.: Chromatin unfolds. In: *Cell* 86 (1996), S. 13–19
- [FH76] FREIRE, Juan J.; HORTA, Arturo: Mean reciprocal distances of short polymethylene chains. Calculation of the translational diffusion coefficient of n-alkanes. In: *J. Chem. Phys.* 65 (1976), S. 4049–4054
- [Fix78] FIXMAN, M.: Simulation of polymer dynamics I. General theory. In: *J. Chem. Phys.* 69 (1978), S. 1527–1359
- [FK76] FINCH, J.T.; KLUG, A.: Solenoidal model for superstructure in chromatin. In: *Proc. Natl. Acad. Sci. USA* 73 (1976), Nr. 6, S. 1897
- [GB81] GAY, J. G.; BERNE, B. J.: Modification of the overlap potential to mimic a linear site-site potential. In: *J. Chem. Phys.* 74 (1981), S. 3316–3319
- [Gen79] GENNES, P. G.: *Scaling Concepts in Polymer Physics*. Ithaca : Cornell Univ. Press, 1979
- [GGSR94] GRAZIANO, V.; GERMAN, S.E.; SCHNEIDER, D.K.; RAMAKRISHNAN, V.: Histone H1 is located int the interior of the chromatin 30-nm filament. In: *Nature* 368 (1994), S. 351–354
- [GHJV96] GAMMA, Erich; HELM, Richard; JOHNSON, Ralph; VLISSIDES, John: *Entwurfsmuster*. Bonn : Addison-Wesley, 1996
- [GI-96] GI-FACHGRUPPE 2.1.7. *Test, Analyse und Verifikation von Software*. 1996
- [GK94] GROSSBERG, Alexander; KHOKLOV, Alexei: *Statistical Physics of Macromolecules*. New York : American Institute of Physics, 1994 (AIP Series in Polymers and Complex Materials)
- [GP93] GENNES, P. G.; PROST, J.: *The Physics of Liquid Crystals*. 2<sup>nd</sup>. Oxford : Oxford University Press, 1993
- [GR87] GERCHMAN, S.E.; RAMAKRISHNAN, V.: Chromatin higher-order structure studied by neutron scattering and scanning transmission electron microscopy. In: *Proc. Natl. Acad. Sci.* 84 (1987), S. 7802–7806
- [Gri96] GRITZ, Larry: *Blue Moon Rendering Tools: User Guide*, 1996

- [Ham99] HAMMERMANN, Markus: *Strukturuntersuchungen mit Neutronenkleinwinkelstreuung an superhelikaler DNA und Oligonukleosomen*, Universität Heidelberg, Dissertation, 1999
- [HB92] Kap. 6 In: HEPPKE, Gerd; BAHN, Christian: *Bergmann-Schäfer — Lehrbuch der Experimentalphysik*. Bd. 5: *Flüssigkristalle*. Berlin : Walter de Gruyter, 1992, S. 389–445
- [HBK<sup>+</sup>98] HAMMERMANN, Markus; BRUN, Nathalie; KLENIN, Konstantin; MAY, Roland; TOTH, Katalin; LANGOWSKI, Jörg: Salt-Dependent DNA Superhelix Diameter Studied by Small Angle Neutron Scattering Measurements and Monte Carlo Simulation. In: *Biophys. J.* 75 (1998), S. 3057–3063
- [HMCL97] HOUCHEMANDZADEH, Bahram; MARKO, John F.; CHATENAY, Didier; LIBCHABER, Albert: Elasticity and Structure of Eukaryote Chromosomes Studied by Micromanipulation and Micropipette Aspiration. In: *The Journal of Cell Biology* 139 (1997), October, Nr. 1, S. 1–12
- [Hol89] HOLDE, Kensal E.: *Chromatin*. New York : Springer, 1989
- [Hol93] HOLST, M.J.: *The Poisson-Boltzmann Equation. Analysis and Multilevel Numerical Solution*, University of Illinois at Urbana-Champaign, Dissertation, 1993. — <ftp://ftp.scicomp.ucsd.edu/pub/mholst/pubs/Hols94d.ps.gz>
- [HSM<sup>+</sup>97] HAMMERMANN, Markus; STEINMAIER, Carola; MERLITZ, Holger; KAPP, Ulrike; WALDECK, Waldemar; CHIRICO, Giuseppe; LANGOWSKI, Jörg: Salt effects on the structure and internal dynamics of superhelical DNAs studied by light scattering and Brownian dynamics. In: *Biophys. J.* 73 (1997), S. 2674–2687
- [HZ95] HOLDE, Kensal van; ZLATANOVA, Jordanka: Chromatin Higher Order Structure: Chasing a Mirage? In: *J. Biol. Chem.* 270 (1995), Nr. 15, S. 8373–8376
- [HZ96] HOLDE, Kensal van; ZLATANOVA, Jordanka: What determines the folding of the chromatin fiber? In: *Proc. Natl. Acad. Sci. USA* 93 (1996), S. 10548–10555
- [Jac83] JACKSON, John D.: *Klassische Elektrodynamik*. Berlin : Walter de Gruyter, 1983

- [Kab86a] KABADI, Vinayak: Molecular Dynamics of Fluids: The Gaussian Overlap Model II. In: *Ber. Bunsenges. Phys. Chem.* 90 (1986), S. 327–332
- [Kab86b] KABADI, Vinayak: Statistical Mechanics of Non-Spherical Molecules: Spherical Harmonic Expansions on non-Spherical Surfaces II. Gay-Berne Gaussian Overlap Potential. In: *Ber. Bunsenges. Phys. Chem.* 90 (1986), S. 332–339
- [Kat99] KATRICH, Seva. *Modeling of chromatin structure at different levels of organization: Stretching of individual chromatin fibers.* <http://abagyan.scripps.edu/seva/stretch.html>. 1999
- [KB73] KUSHICK, Joseph; BERNE, Bruce: Methods for experimentally determining the angular momentum velocity relaxation in liquids. In: *J. Chem. Phys.* 59 (1973), Nr. 8, S. 4486–4490
- [KB75] KUSHICK, Joseph; BERNE, Bruce: Computer Simulation of anisotropic molecular fluids. In: *J. Chem. Phys.* 64 (1975), Nr. 4, S. 1362–1367
- [KB98] KRAJEWSKI, Wladyslaw A.; BECKER, Peter B.: Reconstitution of hyperacetylated, DNase I-sensitive chromatin characterized by high conformational flexibility of nucleosomal DNA. In: *Proc. Natl. Acad. Sci.* 95 (1998), S. 1540–1545
- [KBO98] KATRITCH, V.; BUSTAMANTE, C.; OLSON, W.K.: Pulling Single Chromatin Fibers – Experiment and Theory. In: *Biophys. J.* 74 (1998), Nr. 2, S. A230
- [KFKL95] KLENIN, Konstantin; FRANK-KAMENETSKII, Maxim; LANGOWSKI, Jörg: Modulation of Intramolecular Interactions in Superhelical DNA by Curved Sequences: A Monte Carlo Simulation Study. In: *Biophys. J.* 68 (1995), S. 81–86
- [KHNN90] KUBISTA, Mikael; HAGMAR, Per; NIELSEN, Peter E.; NORDEN, Bengt: Reinterpretation of Linear Dichroism of Chromatin supports a Perpendicular Linker Orientation in the folded State. In: *Journal of Biomolecular Structure and Dynamics* 8 (1990), Nr. 1, S. 37–54
- [Kle93] KLEINERT, Hagen: *Pfadintegrale*. Mannheim : BI Wissenschaftsverlag, 1993

- [KML98] KLENIN, Konstantin; MERLITZ, Holger; LANGOWSKI, Jörg: A Brownian Dynamics Program for the Simulation of Linear and Circular DNA and Other Wormlike Chain Polyelectrolytes. In: *Biophys. J.* 75 (1998), S. 3057–3063
- [Knu81] KNUTH, Donald E.: *The Art of Computer Programming, Volume 2. Seminumerical Algorithms.* Reading, MA : Addison-Wesley, 1981
- [KS81] KIRCKPATRICK, S.; STOLL, E.: A Very Fast Shift-Register Sequence Random Number Generator. In: *Journal of Computational Physics* 40 (1981), S. 517–526
- [LL97] LEFORESTIER, A.; LIVOLANT, F.: Liquid Crystalline Ordering of Nucleosome Core Particles under Macromolecular Crowding Conditions: Evidence for a Discotic Columnar Hexagonal Phase. In: *Biophys. J.* 73 (1997), S. 1771–1776
- [LMR<sup>+</sup>97] LUGER, Karin; MÄDER, Armin; RICHMOND, Robin; SARGENT, David; RICHMOND, Timothy: Crystal structure of the nucleosome core particle at 2.8 Å resolution. In: *Nature* 389 (1997), S. 251–260
- [LS93] LUCKHURST, G. R.; SIMMONDS, P. S. J.: Computer Simulation studies of anisotropic systems. XXI. Parametrization of the Gay-Berne potential for model mesogens. In: *Molecular Physics* 80 (1993), Nr. 2, S. 233–252
- [LZH93] LEUBA, Sanford H.; ZLATANOVA, Jordanka; HOLDE, Kensal van: On the Location of Histones H1 and H5 in the Chromatin Fiber. Studies with Immobilized Trypsin and Chymotrypsin. In: *J. Mol. Biol.* 229 (1993), S. 917–929
- [MC85] MORSE, Randall; CANTOR, Charles: Nucleosome core particles suppress the thermal untwisting of core DNA and adjacent linker DNA. In: *Proc. Natl. Acad. Sci.* 82 (1985), S. 4653–4657
- [Meh98] MEHRING, Carsten: *Modellierung der Dynamik von Interphaschromosomen.* 1998. – Diplomarbeit, Universität Heidelberg
- [MR95] MOTWANI, Rajeev; RAGHAVAN, Prabhakar: *Randomized Algorithms.* Cambridge : Cambridge University Press, 1995

- [MRBA96] MIGUEL, Enrique de; RIO, Elvira M.; BROWN, Julian T.; ALLEN, Michael P.: Effect of the attractive interactions on the phase behaviour of the Gay–Berne liquid crystal. In: *J. Chem. Phys.* 105 (1996), Nr. 10, S. 4234–4249
- [MRKL98] MERLITZ, Holger; RIPPE, Karsten; KLENIN, Konstantin; LANGOWSKI, Jörg: Looping Dynamics of Linear DNA Molecules and the Effect of DNA Curvature: A study by Brownian Dynamics Simulation. In: *Biophys. J.* 74 (1998), S. 773–779
- [MRR<sup>+</sup>53] METROPOLIS, N.; ROSENBLUTH, A.W.; ROSENBLUTH, M.N.; TELLER, A.H.; TELLER, E.: Equation of state calculations by fast computing machines. In: *J. Chem. Phys.* 21 (1953), S. 1087–1092
- [MS88] MEAL, N.; SOKAL, A.D.: In: *J. Stat. Phys.* 50 (1988), S. 109
- [MS96] MUSSER, David R.; SAINI, Atul: *STL Tutoraial and Reference Guide*. Reading, MA : Addison–Wesley, 1996
- [OKV90] OSIPOVA, T.N.; KARPOVA, E.V.; VOROB'EV, V.I.: Chromatin Higher–Order Structure: Two–Start Double Superhelix Formed by Zig–Zag Sharped Nucleosome Chain with Folded Linker DNA. In: *J. Biomol. Struct. Dyn.* 8 (1990), Nr. 1, S. 11–22
- [OL94] OSTASHEVSKY, J.Y.; LANGE, C.S.: The 30 nm Chromatin fiber as a Flexible Polymer. In: *Journal of Biomolecular Structure & Dynamics* 11 (1994), Nr. 4, S. 813–820
- [Par98] ParaSoft: *Insure++ Manual v 4.11*. 1998
- [Phi97] PHILO, John: *SEDNTERP: Software Description*. 1.01. Amgen Center. Thousand Oaks, CA 91320-1789. <http://www.jphilo.mailway.com>: Protein Chemistry 14-2-D. Amgen Inc., 1997
- [Pix89] Pixar: *The RenderMan interface and shading languag*. 1989
- [PTVF92] PRESS, W.H.; TEUKOLSKY, S.A.; VETTERLING, W.T.; FLANNERY, B.P.: *Numerical Recipes in C*. second. Cambridge : Cambridge University Press, 1992
- [Rei65] REIF, Frederick: *Fundamentals of Statistical and Thermal Physics*. New York : M<sup>c</sup>Graw-Hill, 1965

- [RF94] RÖMER, Hartmann; FILK, Thomas: *Statistische Mechanik*. Weinheim : VCH, 1994 (Konzepte der Theoretischen Physik)
- [RHMC98] RYDBERG, Björn; HOLLEY, William R.; MIAN, I. S.; CHATTERJEE, Alope: Chromatin Conformation in Living Cells: Support for a Zig-Zag Model of the 30 nm Chromatin Fiber. In: *J. Mol. Biol.* 284 (1998), S. 71–84
- [RU97a] RATIONAL; UML PARTNERS. *UML Notation Guide version 1.1*. <http://www.omg.org>. 1 September 1997
- [RU97b] RATIONAL; UML PARTNERS. *UML Proposal Summary version 1.1*. <http://www.omg.org>. 1 September 1997
- [RU97c] RATIONAL; UML PARTNERS. *UML Semantics and appendices version 1.1*. <http://www.omg.org>. 1 September 1997
- [RU97d] RATIONAL; UML PARTNERS. *UML Summary version 1.1*. <http://www.omg.org>. 1 September 1997
- [Rum91] RUMBOUGH, J.: *Object-Oriented Modeling and Design*. Prentice Hall, 1991
- [SBA<sup>+</sup>96] SPOEL, David van d.; BUUREN, Aldert van; APOL, Emile; MEULENHOF, Pietar; TIELEMAN, D. P.; SIJBERS, Alfons; DRUNEN, Rudi van; BERENDSEN, Herman: *GROMACS User Manual*. 1.3, 1996
- [SBB79] SUAU, Pedro; BRADBURY, E. M.; BALDWIN, John P.: Higher-Order Structures of Chromatin in Solution. In: *Eur. J. Biochem.* 97 (1979), S. 593–602
- [SCB96] SMITH, Steven; CUI, Yujia; BUSTAMANTE, Carlos: Overstretching B-DNA: The Elastic Response of Individual Double-Stranded and Single-Stranded DNA Molecules. In: *Science* 271 (1996), S. 795–799
- [Sch98] SCHWABE, Johannes: Team-Probleme. Amdahls Gesetz beschreibt die Leistungsbegrenzung von Multiprozessorsystemen. In: *c't* 26 (1998), S. 80–81
- [Sok96a] SOKAL, Alan D.: Monte Carlo methods for the self-avoiding walk. In: *Nuclear Physics B* S47 (1996), S. 172–179

- [Sok96b] SOKAL, Alan D.: *Monte Carlo Methods in Statistical Mechanics: Foundations and New Algorithms*. September 1996. – Lectures at the Cargèse Summer School
- [SS77] SCHELLMANN, J.A.; STIGTER, D.: Electrical double layer, zeta potential, and electrophoretic charge of double stranded DNA. In: *Biopolymers* 1415–1434 (1977)
- [Sti77] STIGTER, Dirk: Interactions of highly charged colloidal cylinders with applications to double stranded DNA. In: *Biopolymers* 16 (1977), S. 1435–1448
- [Str91] STRYER, Lubert: *Biochemie*. Heidelberg : Spektrum Akademischer Verlag, 2. neubearbeitete Auflage 1991
- [Tan95] TANENBAUM, Andrew S.: *Moderne Betriebssysteme*. München : Carl Hanser, 1995
- [TKK79] THOMA, F.; KOLLER, Th.; KLUG, A.: Involvement of the histone H1 in the Organization of the Nucleosome and of the Salt-Dependent Superstructures of Chromatin. In: *J. Cell. Biol.* 83 (1979), S. 403–427
- [Tra91] TRAUT, Walther: *Chromosomen. Klassische und molekulare Cytogenetik*. 1<sup>st</sup>. Berlin : Springer, 1991
- [Tra99] TRAVERS, Andrew: The location of the linker histone on the nucleosome. In: *Trends Biochem. Sci.* 24 (1999), S. 4–8
- [TT77] THOMAS, J.O.; THOMPSON, R.J.: Variation in chromatin structure in two cell types from the same tissue: a short DNA repeat length in cerebral cortex neurons. In: *Cell* 10 (1977), S. 633–640
- [VC95] VOLOGODSKII, A.; COZARELLI, N.: Modeling of long range electrostatic interactions in DNA. In: *Biopolymers* 35 (1995), S. 289–296
- [VS62] VERDIER, Peter H.; STOCKMAYER, W.H.: Monte Carlo Calculations on the Dynamics of Polymers in Dilute Solution. In: *J. Chem. Phys.* 36 (1962), Nr. 1, S. 227–235
- [WGHW93] WOODCOCK, C. L.; GRIGORYEV, S. A.; HOROWITZ, R. A.; WHITAKER, N.: A chromatin folding model that incorporates linker variability generates fibers resembling the native structures. In: *Proc. Natl. Acad. Sci.* 90 (1993), S. 9021–9025

- [WH95] WOODCOCK, C.L.; HOROWITZ, R.A.: Chromatin organization re-viewed. In: *Trends Cell Biol.* 5 (1995), S. 272–277
- [WMSL98] WEDEMANN, Gero; MÜNDEL, Christian; SCHÖPPE, Gunther; LANGOWSKI, Jörg: Kinetics of structural changes in superhelical DNA. In: *Phys. Rev. E* 58 (1998), Nr. 3, S. 3537–3546
- [Wol94] WOLFFE, Alan: *Regulation of Chromatin Structure and Function*. Austin, TX : R.G. Landes Company, 1994
- [Wol95] WOLFFE, Alan: *Chromatin: Structure and Function*. London : Academic Press, 1995
- [YEH+95] YOKOTA, Hiroki; ENGH, Ger van d.; HEARST, John E.; SACHS, Rainer K.; TRASK, Barbara J.: Evidence for the Organization of Chromatin in Megabase Pair-sized Loops Arranged along a Random Walk Path in the Human G0/G1 Interphase Nucleus. In: *Journal of Cell Biology* 130 (1995), September, Nr. 6, S. 1239–1249
- [YLW90] YAO, J.; LOWARY, P. T.; WIDOM, J.: Direct detection of linker DNA bending in defined-length oligomers of chromatin. In: *Proc. Natl. Acad. Sci. USA* 87 (1990), S. 7603–7607
- [YLW91] YAO, J.; LOWARY, P. T.; WIDOM, J.: Linker DNA bending induced by the core histones of chromatin. In: *Biochemistry* 30 (1991), S. 8408–8414
- [ZLH98] ZLATANOVA, Jordanka; LEUBA, Sanford H.; HOLDE, Kensal van: Chromatin Fibre Structure: Morphology, Molecular Determinants, Structural Transitions. In: *Biophys. J.* 74 (1998), S. 2554–2566