

4. Monte-Carlo-Algorithmus zur Simulation von Nukleinsäuren

4.1. Definition des molekularen Modells

4.1.1. Reduktion der Freiheitsgrade durch Einführung kollektiver und innerer Variablen

MC-Simulationen molekularer Systeme ermöglichen im Gegensatz zu MD-Simulationen auf einfachere Weise die Verwendung generalisierter Variablen, da bei MC-Simulationen nur die Energien der Konformationen nicht aber Kräfte und Trägheitstensoren zu berechnen sind. Wie in Kap. 3.1.2 beschrieben, führt die Verwendung innerer bzw. kollektiver Variablen im Vergleich zu einer Beschreibung in kartesischen Koordinaten zu einer deutlichen Reduktion der Zahl der Freiheitsgrade des zu behandelnden Systems. Das im Rahmen dieser Arbeit verwendete und im folgenden beschriebene molekulare Modell beschreibt die Nukleinsäurestrukturen im Raum kollektiver Variablen, ergänzt durch einen Satz innerer Variablen. Das hier vorgestellte Modell basiert auf einer lokalen Beschreibung der Nukleinsäurestrukturen, d.h. daß alle eingeführten Variablen lokal auf der Ebene der Nukleotide definiert werden. Diese Reduktion auf eine lokale Beschreibung der Nukleinsäurestrukturen ermöglicht die Definition von MC-Übergängen in Form lokaler Konformationsänderungen.

Die Basen der Nukleotide werden aufgrund ihrer konjugierten π -Elektronensysteme als starre Körper betrachtet. Die Basen besitzen somit als starre Körper keine inneren Freiheitsgrade. Sämtliche Bindungslängen, Valenz- und Torsionswinkel sind fixiert. Die Lage und Orientierung dieser starren Körper werden durch drei Translationen T_i in die drei Raumrichtungen und drei Rotationen R_i um drei senkrecht aufeinander stehende Achsen variiert. Diese Translationen T_i und Rotationen R_i werden auf der Basis der in Kap. 2.2 eingeführten körperfesten Koordinatensysteme der Basen definiert. Eine Ausnahme von der Beschreibung der Basen als starre Körper stellt, wie in Abb. 2.1 ersichtlich, lediglich die am Thymin vorhandene Methylgruppe dar, deren Wasserstoffatome frei um die C-N-Bindung rotieren können.

Darüberhinaus werden alle Bindungslängen innerhalb der Nukleinsäurestruktur als konstant betrachtet. Die Valenzwinkel hingegen sind variabel mit Ausnahme der exozyklischen Valenzwinkel. Als endozyklische und somit variable Valenzwinkel werden in dieser Arbeit neben den am Furanosering beteiligten endozyklischen Valenzwinkeln auch die zur Kettenbildung der Schweratome des Nukleinsäure-Rückgrats beitragenden Valenzwinkel bezeichnet. Somit sind alle am Kettenschluß nach Abb. 2.4 beteiligten Valenzwinkel endozyklisch, alle nicht daran beteiligten exozyklisch. Sämtliche exozyklischen Valenzwinkel, also die Bindungswinkel zwischen Atomen, die die Valenzen der Ketten und Ringe innerhalb des Desoxyribose-Phosphodiester-Rückgrats bildenden Atome

absättigen, werden wie die Bindungslängen als fixierte Größen betrachtet. Beispiele dafür sind der Valenzwinkel zwischen den Phosphatsauerstoffen $\angle(\text{OP1-P-OP2})$ und der Valenzwinkel zwischen den Wasserstoffen am C5'-Atom $\angle(\text{H5'-C5'-H5''})$. Grundlage der Wahl der fixen Werte für Bindungslängen und exozyklische Valenzwinkel sind experimentelle Daten und quantenchemische Rechnungen. Bei der Modellbildung wird davon ausgegangen, daß das MC-Sampling dieser Größen keinen wesentlichen Einfluß auf die Strukturbildung hat, so daß im Interesse einer Reduktion der Freiheitsgrade des Systems die Bindungslängen und exozyklischen Valenzwinkel als fixierte Größen betrachtet werden.

4.1.2. Molekulares Modell mit iterativem Kettenschluß

Die Lage und Orientierung der als starre Körper betrachteten Basen verändert sich ständig im Laufe der MC-Simulation. Dieser Bewegung der Basen wird durch ein flexibles Desoxyribose-Phosphodiester-Rückgrat Rechnung getragen. Die Bewegung einer individuellen, in Abb. 4.1 als blauer Kreis dargestellten Base als starrer Körper wird durch sechs Freiheitsgrade ($3T_i, 3R_i$) beschrieben. Die durch zwei weitere Freiheitsgrade (P, A) charakterisierte Umfaltung des Furanosringes wird in Kap. 4.1.4 erläutert. Diese acht Freiheitsgrade eines Nukleotids führen bei einer Variation notwendigerweise zum Kettenbruch, der wegen der in diesem Algorithmus definierten monomeren Einheiten, wie in Abb. 4.1 dargestellt, an der O5'-C5'-Bindung liegt. Da das bewegte Nukleotid j am 5'-Ende mit dem Nukleotid $j - 1$ und am 3'-Ende mit dem Nukleotid $j + 1$ benachbart ist, führt die Bewegung des Nukleotids j , wie in Abb. 4.1 ersichtlich, zum zweifachen Kettenbruch ober- und unterhalb dieses Nukleotids jeweils an den O5'-C5'-Bindungen.

Um die kovalenten Bindungen zwischen den O5'- und C5'-Atomen mit der Bindungslänge $d=1,44 \text{ \AA}$ wiederherzustellen und damit den Kettenschluß nach der Bewegung des Nukleotids herbeizuführen, wird gefordert, daß unter der Bedingung des Kettenschlusses

$$|\vec{r}_{O5'} - \vec{r}_{C5'}| = |\vec{r}'_{O5'} - \vec{r}'_{C5'}| = d \quad (4.1)$$

die Summe der Abweichungen aller am Kettenschluß beteiligten Valenzwinkel ω_i von ihren definierten Sollwerten ω_i^0 minimal wird:

$$F = \sum_{i=1}^4 (\omega_i - \omega_i^0)^2 + \sum_{i=1}^4 (\omega'_i - \omega_i^0)^2 \rightarrow \min \quad (4.2)$$

Die Sollwerte ω_i^0 sind die im Kraftfeld AMBER 3.19 aufgrund experimenteller Daten mit $\omega_1^0 = \omega_2^0 = \omega_3^0 = 120,5^\circ$ und $\omega_4^0 = 109,5^\circ$ festgelegten Werte für die Minima der Deformationsenergien der jeweiligen Valenzwinkel [105, 106]. Die Valenzwinkel ω_i befinden sich im Phosphodiester-Rückgrat zwischen den Pentosen der Nukleotide j und $j - 1$, während sich die Valenzwinkel ω'_i zwischen den Pentosen der Nukleotide j und $j + 1$ befinden. Die endozyklischen Valenzwinkel sind in dem molekularen Modell abhängige Variablen. Die Minimum-Forderung 4.2 führt allerdings zu einer eingeschränkten Variabilität der endozyklischen Valenzwinkel ω_i um die geforderten Sollwerte ω_i^0 . Das molekulare Modell bietet auch die Möglichkeit, wahlweise die Sollwerte ω_i^0 der in der

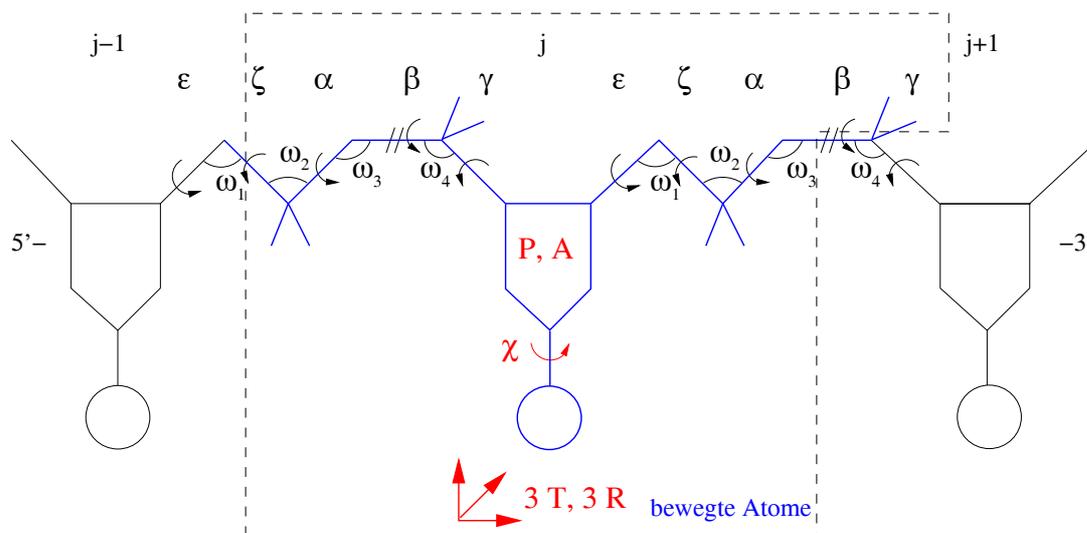


Abb. 4.1.: Schematische Darstellung des molekularen Modells mit der Darstellung des bewegten Fragments in blau und der MC-Variablen in rot

Minimum-Forderung benutzten Valenzwinkel durch MC-Übergänge zu variieren. Dadurch würde sich die Zahl der Freiheitsgrade je Nucleotid um 4 auf 12 MC-Variablen je Nucleotid erhöhen.

Die Torsionswinkel ε , ζ , α , β , γ und χ werden infolge des Kettenschlusses nach jeder MC-Bewegung der Base als abhängige Variablen variiert. Der Torsionswinkel δ ist ebenfalls eine abhängige Variable, die durch die in Kap. 4.1.4 diskutierte Umfaltung des Zuckers variiert wird, an der das C5'- und das O3'-Atom beteiligt sind. Die Minimum-Forderung 4.2 unter den Zwangsbedingungen 4.1 enthält eine Kopplung der beiden Kettenschlüsse auf der 5'- und der 3'-Seite des bewegten Nucleotids. Diese Kopplung führt auch zur Variation des Glykosidwinkels χ durch den Kettenschluß, allerdings als abhängige Variable.

Die Minimum-Forderung 4.2 ist gleichbedeutend mit der Forderung nach dem Verschwinden der ersten Ableitungen der Funktion $F(\{x_i\})$ nach den Variablen x_i , wobei zusätzlich die Matrix der 2. Ableitungen positiv definit sein muß. Die vier Valenzwinkel in 4.2 auf beiden Seiten des bewegten Nucleotids sind gekoppelte Variablen. Dabei können aber zwei der jeweils vier Valenzwinkel als Funktionen der beiden anderen Valenzwinkel, zweier Torsionswinkel und des Glykosidwinkels χ ausgedrückt werden, so daß die Größe F in der Minimum-Forderung 4.2 eine Funktion von jeweils zwei Valenz- und zwei Torsionswinkeln auf jeder Seite des bewegten Nucleotids sowie des die Kopplung herstellenden Glykosidwinkels χ ist. Damit ergeben sich 9 Variablen x_i , von denen die Summe der quadratischen Abweichungen der 8 Valenzwinkel ω_i in der Minimum-Forderung 4.2 abhängig ist. Die Auswahl der zu eliminierenden Valenzwinkel kann dabei nach Zweckmäßigkeit erfolgen. Werden die Valenzwinkel ω_3 , ω_4 , ω'_3 und ω'_4 laut Abb. 4.1 ersetzt, ist die Größe F in 4.2 eine Funktion der Valenzwinkel ω_1 , ω_2 , ω'_1 und ω'_2 , der Torsionswinkel $\tau_1 = \varepsilon$, $\tau_2 = \zeta$, $\tau'_1 = \varepsilon'$ und $\tau'_2 = \zeta'$ sowie des Glykosidwinkels χ .

Auch die Zwangsbedingungen 4.1 für die Korrektur der O5'-C5'-Bindungen lassen sich als Funktionen der gewählten jeweils 5 Variablen ausdrücken:

$$d = |\vec{r}_{O5'-C5'}| = f_1(\omega_1, \omega_2, \tau_1, \tau_2, \chi) \quad (4.3)$$

$$d = |\vec{r}'_{O5'-C5'}| = f_2(\omega'_1, \omega'_2, \tau'_1, \tau'_2, \chi) \quad (4.4)$$

Mit Hilfe dieser so formulierten Zwangsbedingungen kann jeweils eine der 5 Variablen auf jeder Seite des bewegten Nukleotids eliminiert werden. Auch diese Auswahl kann nach Zweckmäßigkeit erfolgen, mit Ausnahme des Glykosidwinkels χ , der, weil er die Kopplung der beiden Kettenschlüsse herstellt, nicht eliminiert werden kann. Wird der Torsionswinkel τ_2 eliminiert, wird die Größe $F(x_i)$ in der Minimum-Forderung 4.2 zu einer Funktionen von 7 Variablen x_i :

$$F(\omega_1, \omega_2, \tau_1, \chi, \omega'_1, \omega'_2, \tau'_1) \rightarrow \min \quad (4.5)$$

Die Ableitung der Funktion F nach diesen 7 Variablen führt damit zu 7 Gleichungen:

$$\frac{dF(\{x_i\})}{dx_i} = 0 \quad \text{mit} \quad \{x_i\} = (\omega_1, \omega_2, \tau_1, \chi, \omega'_1, \omega'_2, \tau'_1) \quad (4.6)$$

Dieses nichtlineare 7×7 -Gleichungssystem wird iterativ durch das Newton-Verfahren gelöst. Alternativ dazu wurden Lagrangesche Multiplikatoren in ein System mit 9 statt 7 unabhängigen Variablen eingeführt, da dann die beiden Torsionswinkel τ_2 und τ'_2 nicht durch die Zwangsbedingungen eliminiert worden sind. Die Lagrangeschen Multiplikatoren erweitern das 7×7 -Gleichungssystem zu einem 11×11 -Gleichungssystem. Diese Erweiterung des hier erläuterten Algorithmus führte nicht zu einer Reduktion der ohnehin schon sehr wenigen nicht möglichen Kettenschlüsse, so daß im Interesse einer Rechenzeiterparnis beim Kettenschluß die Lösung des 7×7 -Gleichungssystems 4.6 in der ersten hier vorgestellten Variante des molekularen Modells für den Kettenschluß Verwendung gefunden hat.

8 MC-Variablen je Nukleotid:	
3 T_i , 3 R_i	Translationen und Rotationen des Nukleotids
P , A	Phase und Amplitude des Zuckers
Kettenschluß: Iterative Lösung eines gekoppelten 7×7 -Gleichungssystems	

Tabelle 4.1.: Molekulares Modell mit gekoppelten Kettenschlußbedingungen

Die Verwendung fester Sollwerte ω_i^0 schränkt allerdings die notwendige Reversibilität der MC-Übergänge ein. Erst eine Variation der Sollwerte ω_i^0 der Valenzwinkel als zusätzliche MC-Variablen führt zu einer Erhöhung des Anteils reversibler MC-Übergänge.

Entkopplung der Kettenschlußbedingungen

Wird auch der Glykosidwinkel χ als MC-Variable im MC-Sampling frei variiert, können die beiden Kettenschlüsse auf der 5'- und der 3'-Seite des bewegten Nukleotids unabhängig voneinander behandelt werden, so daß die Minimum-Forderung 4.2 in zwei entkoppelte Forderungen bei Erfüllung der Zwangsbedingungen 4.1 übergeht:

$$F_1 = \sum_{i=1}^4 (\omega_i - \omega_i^0)^2 \rightarrow \min \quad \text{und} \quad F_2 = \sum_{i=1}^4 (\omega'_i - \omega_i^0)^2 \rightarrow \min \quad (4.7)$$

Aufgrund der Entkopplung sind die Funktionen F_1 und F_2 nur jeweils von 4 Variablen abhängig. Werden die Zwangsbedingungen 4.1 wiederum zur Eliminierung einer Variablen benutzt, sind in diesem entkoppelten Fall zwei 3×3 -Gleichungssysteme zu lösen:

$$\frac{dF_1(\{x_i\})}{dx_i} = 0 \quad \text{mit} \quad \{x_i\} = (\omega_1, \omega_2, \tau_1) \quad (4.8)$$

$$\frac{dF_2(\{x'_i\})}{dx'_i} = 0 \quad \text{mit} \quad \{x'_i\} = (\omega'_1, \omega'_2, \tau'_1) \quad (4.9)$$

Die Wahl der zu eliminierenden Variablen sollte auch hier nach Zweckmäßigkeit erfolgen. Auch wenn die Torsionswinkel τ_2 und τ'_2 eliminiert worden sind, enthält die Umsetzung im Algorithmus ein Auswahlkriterium, demzufolge diejenige Variable eliminiert wird, deren Koeffizient in den Zwangsbedingungen 4.3-4.4 den größten Absolutbetrag aufweist. Dadurch werden Divisionen durch sehr kleine Zahlen und somit eventuelle Singularitäten bei der Eliminierung verhindert.

Die Erhöhung der Zahl der 8 MC-Variablen bei gekoppelten Kettenschlußbedingungen um den Glykosidwinkel χ auf 9 MC-Variablen bei gleichzeitiger Entkopplung der Kettenschlußbedingungen vereinfacht den Algorithmus, da nun nur noch zwei 3×3 -Gleichungssysteme statt eines 7×7 -Gleichungssystems zu lösen sind. Der Kettenschluß wird durch die Einführung einer zusätzlichen MC-Variablen in diesem zweiten hier vorgestellten molekularen Modell somit signifikant vereinfacht.

9 MC-Variablen je Nukleotid:	
3 T_i , 3 R_i	Translationen und Rotationen des Nukleotids
P , A	Phase und Amplitude des Zuckers
χ	Glykosidwinkel (O1'-C1'-N-C6/C8)
Kettenschluß: Iterative Lösung zweier entkoppelter 3×3 -Gleichungssysteme	

Tabelle 4.2.: *Molekulares Modell mit entkoppelten Kettenschlußbedingungen*

Da die P- und O5'-Atome des Nukleinsäure-Rückgrats beim Kettenschluß bewegt werden, müssen natürlich diejenigen Atome mitbewegt werden, die die Valenzen an diesen beiden Atomen absättigen. Das sind am P-Atom zwei Phosphatsauerstoffe O1P und O2P sowie am C5'-Atom zwei Wasserstoffe H5' und H5''. Diese Atome werden unter Beibehaltung ihrer exozyklischen Valenzwinkel $\angle(OP1-P-OP2) = 116,1^\circ$ und $\angle(H5'-C5'-H5'') = 109,6^\circ$ in der Mittelsenkrechenebene der variierten endozyklischen Valenzwinkel $\angle(O3'-P-O5')$ sowie $\angle(O5'-C5'-C4')$ platziert.

Mikroreversibilität des Kettenschlusses

Um den Metropolis-Algorithmus anwenden zu können, müssen die beim Kettenschluß gefundenen Lösungen reversibel sein. Da der Kettenschluß iterative Lösungen der jeweiligen Gleichungssysteme verwendet, muß jede Lösung auf ihre Mikroreversibilität

getestet werden. Nur reversible Lösungen werden akzeptiert und in das MC-Sampling einbezogen. Zum Test der Mikroreversibilität wird jeder MC-Übergang nach erfolgtem Kettenschluß umgekehrt und getestet, ob der dann erfolgende Kettenschluß die Ausgangskonformation reproduziert. Nur ein solcher MC-Übergang wird akzeptiert. Der Algorithmus erzielt in nahezu allen MC-Übergängen einen reversiblen Kettenschluß. Der Anteil der wegen nicht erfüllter Mikroreversibilität nicht akzeptierten MC-Übergänge liegt im %-Bereich.

4.1.3. Molekulares Modell mit analytischem Kettenschluß

Der iterative Kettenschluß verwendet die der Ausgangskonformation am nächsten liegende Lösung, da nicht alle Lösungen gefunden und damit berücksichtigt werden können. Somit besitzen mit dem iterativen Kettenschluß MC-Übergänge zwischen auf der Potentialhyperfläche weit auseinander liegenden Zuständen eine sehr geringe Wahrscheinlichkeit. Um diese Einschränkung bei der Auswahl der Lösungen durch den Algorithmus des iterativen Kettenschlusses zu beseitigen, wurde eine Variante des Kettenschlusses entwickelt, die auf einer analytischen Behandlung basiert. Dieser analytische Kettenschluß erlaubt einerseits die Einbeziehung aller Lösungen in den Metropolis-Algorithmus und ermöglicht andererseits MC-Übergänge zwischen Zuständen unabhängig von dazwischen liegenden Energiebarrieren.

Um analytische Kettenschlußbedingungen formulieren und trotzdem die volle Flexibilität des Nukleinsäure-Rückgrats gewährleisten zu können, wird die Zahl der in Tab. 4.2 charakterisierten 9 MC-Variablen des molekularen Modells mit iterativem, entkoppeltem Kettenschluß um 3 weitere MC-Variablen auf 12 MC-Variablen je Nukleotid erhöht. Die Auswahl dieser 3 weiteren MC-Variablen aus den bisher als abhängige, durch den Kettenschluß variierten Variablen erfolgt nach Zweckmäßigkeit. Werden die bisher abhängigen Variablen ε , ω_1 und γ nun als freie MC-Variablen behandelt, muß infolge des Kettenschlusses lediglich die Position des O5'-Atoms berechnet werden. Zweckmäßig ist die Wahl dieser 3 Variablen deshalb, weil aufgrund dieser Wahl das O5'-Atom platziert werden muß, an welchem außer der für den Kettenschluß benutzten Atome keine weiteren Atome gebunden sind.

Das O5'-Atom muß aufgrund der als konstant angenommenen Bindungslängen auf dem Schnittkreis zweier Kugelsphären platziert werden, die in der schematischen Abb. 4.2 in der Ebene als Kreise dargestellt sind. Der Mittelpunkt einer dieser Kugelsphären mit dem Radius der P-O5'-Bindungslänge befindet sich am P-Atom, wohingegen der Mittelpunkt der anderen Kugelsphäre mit dem Radius der O5'-C5'-Bindungslänge sich am C5'-Atom befindet. Diese beiden Kugelsphären schneiden sich in einem Kreis, der in Abb. 4.2 nicht in der Zeichenebene liegt. Auf diesem Schnittkreis muß das O5'-Atom platziert werden. Da der Torsionswinkel γ als MC-Variable unabhängig vom Kettenschluß ist, wird durch diesen Torsionswinkel die Freiheit der Platzierung des O5'-Atoms auf zwei Punkte auf diesem Schnittkreis eingeschränkt. Diese zwei Punkte werden in der schematischen Abb. 4.2 für den Spezialfall ersichtlich, in dem die Mittelpunkte beider Kugelsphären in der Zeichenebene liegen und der Schnittkreis damit senkrecht zur Zeichenebene orientiert ist.

Die Berechnung dieser beiden Schnittpunkte resultiert in der Lösung einer quadratischen Gleichung. Mit der Reduktion des analytischen Kettenschlusses auf die Lösung einer quadratischen Gleichung gibt es für die Platzierung des O5'-Atoms entweder kei-

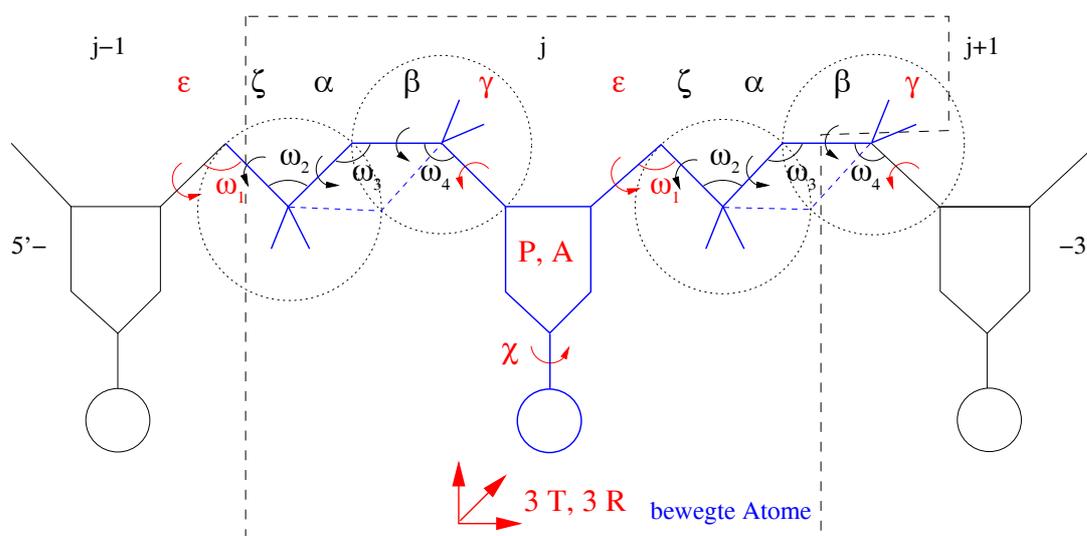


Abb. 4.2.: Schematische Darstellung des molekularen Modells mit der Darstellung des bewegten Fragments in blau und der MC-Variablen in rot

ne Lösung oder immer zwei Lösungen, die in Abb. 4.2 schematisch dargestellt sind. Damit lassen sich alle Lösungen durch den Metropolis-Algorithmus bewerten. Außerdem ist die Mikroreversibilität aufgrund der analytischen Lösung des Kettenschlusses immer erfüllt. Auch wenn sich die Zahl der MC-Variablen erhöht hat, so wurde doch mit diesem molekularen Modell eine weniger komplizierte Variante des Kettenschlusses gefunden, weshalb dieses Modell im weiteren Anwendung finden wird.

12 MC-Variablen je Nucleotid:	
$3 T_i, 3 R_i$	Translationen und Rotationen des Nucleotids
P, A	Phase und Amplitude des Zuckers
χ	Glykosidwinkel $O1'-C1'-N-C6/C8$
ε, γ	Torsionswinkel $C4'-C3'-O3'-P$ und $O3'-C5'-C4'-C3'$
ω_1	Valenzwinkel $C3'-O3'-P$
Kettenschluß: Analytische Lösung einer quadratischen Gleichung	

Tabelle 4.3.: Molekulares Modell mit analytischem Kettenschluß

Die Variation aller übrigen, in Abb. 4.2 ersichtlichen Torsionswinkel und endozyklischen Valenzwinkel als abhängige Variablen durch den Kettenschluß führt zur vollen Flexibilität des Nucleinsäure-Rückgrats. Die Phosphatsauerstoffe am P-Atom sowie die Wasserstoffe am C5'-Atom werden wie in Kap. 4.1.2 in den Mittelsenkrechenebenen der variierten endozyklischen Valenzwinkel unter Beibehaltung der exozyklischen Valenzwinkel platziert. Dadurch wird der Kettenschluß durch Wiederherstellung aller kovalenten Bindungen vervollständigt.

Auswahl einer analytischen Lösung

Das molekulare Modell mit analytischem Kettenschluß erzeugt durch die Lösung einer quadratischen Gleichung zwei exakte Lösungen, von denen eine ausgewählt werden muß. Dazu werden beide Lösungen mit dem Boltzmann-Faktor der jeweiligen Winkelenergien E_{ϑ} gewichtet. Die als Winkelenergie bezeichnete Energie E_{ϑ} ist dabei die Summe der Valenz- und Torsionswinkelbeiträge des dem bewegten Nukleotid i zuzuordnenden Nukleinsäure-Rückgrats. Die Energie E_{ϑ} kann lokal betrachtet werden, da sich durch einen Kettenschluß am Nukleotid i auch nur die Winkelenergie an diesem Nukleotid ändert. Die Berücksichtigung der Energie E_{ϑ} bei der Auswahl einer Lösung n erhöht die Effizienz des MC-Samplings, da aufgrund dieses Kriteriums diejenige Lösung mit dem energetisch vorteilhafteren Kettenschluß begünstigt wird. Dieses Procedere wird dadurch gerechtfertigt, daß im nach der Auswahl anzuwendenden Akzeptanzkriterium 4.12 die Auswahlwahrscheinlichkeit berücksichtigt wird [150].

Die ausgewählte Lösung n wird mit dem Boltzmann-Faktor gewichtet und durch die Summe der mit dem Boltzmann-Faktor gewichteten Energien beider möglichen Lösungen dividiert, um die Auswahlwahrscheinlichkeit $\sigma(m \rightarrow n)$ der Lösung n beim MC-Übergang vom Zustand m zum Zustand n zu berechnen [149, 150]:

$$\sigma(m \rightarrow n) = \frac{e^{-\beta E_{\vartheta}(n)}}{\sum_{k=1}^2 e^{-\beta E_{\vartheta}(k)}} \quad (4.10)$$

Die Auswahl einer Lösung n erfolgt dann, wenn für diese Lösung und eine Zufallszahl ξ gilt:

$$\sigma(m \rightarrow n) \geq \xi \quad (4.11)$$

Ist diese Bedingung nicht erfüllt, fällt die Wahl auf die zweite Lösung.

Akzeptanzkriterium der gewählten Lösung

Bei der Entscheidung über die Akzeptanz sind neben der Änderung der Gesamtenergie in einem MC-Übergang zwei weitere Größen zu berücksichtigen. Erstens wurde bei jedem Kettenschluß eine der beiden möglichen analytischen Lösungen ausgewählt, so daß auch die Auswahlwahrscheinlichkeit in die Entscheidung über die Akzeptanz mit eingehen muß [149, 150]. Und zweitens erfordert die MC-Simulation im Raum generalisierter Variablen die Berücksichtigung von Jacobi-Faktoren, die, wie in Kap. 3.2.5 erläutert, die Erhaltung von Volumenelementen bei der Transformation von kartesischen Koordinaten in diese generalisierten Variablen gewährleisten [149, 150, 151].

Außerdem wird der Jacobi-Faktor der ausgewählten Lösung n berechnet. Dabei werden die Jacobi-Faktoren lokal für jeden Kettenschluß und für jedes Nukleotid bestimmt. Die Jacobi-Faktoren des Kettenschlusses am bewegten Nukleotid vor und nach dem MC-Übergang werden wie die Auswahlwahrscheinlichkeiten in die Berechnung der Akzeptanzwahrscheinlichkeit P mit einbezogen [149, 150]:

$$P(m \rightarrow n) = \min \left\{ 1, \frac{\sigma(n \rightarrow m) J(n) e^{-\beta E(n)}}{\sigma(m \rightarrow n) J(m) e^{-\beta E(m)}} \right\} \quad (4.12)$$

Nach diesem Akzeptanzkriterium wird ein MC-Übergang immer dann akzeptiert, wenn das Verhältnis der Boltzmann-gewichteten Energien E vor und nach dem MC-Übergang unter Einbeziehung der Jacobi-Faktoren J beider Zustände und der Auswahlwahrscheinlichkeiten σ in $4.12 \geq 1$ ist. Ist das nicht erfüllt, entscheidet, wie in Kap. 3.2.3 erläutert, ein Zufallskriterium über die Akzeptanz.

Vorteile des analytischen Kettenschlusses

MC-Simulationen benötigen bei Anwendung des analytischen Kettenschlusses drei zusätzliche MC-Variablen gegenüber der Anwendung des in 4.1.2 vorgestellten iterativen Kettenschlusses. Die Erhöhung der Zahl der MC-Variablen von 9 auf 12 ermöglicht aber einen weniger komplizierten Kettenschluß. Beim analytischen Kettenschluß erfordert die Berechnung der Jacobi-Faktoren im Vergleich zum iterativen Kettenschluß nahezu keine zusätzliche Rechenzeit, da die dazu erforderlichen Größen bereits für den Kettenschluß berechnet worden sind. Der iterative Kettenschluß erfordert für jede Lösung den Test, ob die notwendige Bedingung der Mikroreversibilität erfüllt ist. Dieser Test erübrigt sich bei einer analytischen Lösung der Kettenschlußbedingung.

Der analytische Kettenschluß ermöglicht die Einbeziehung aller möglichen Lösungen, also der jeweils zwei alternativen Lösungen je Kettenschluß, in die Auswahlkriterien des Metropolis-Algorithmus, wodurch die Effektivität des MC-Samplings gesteigert wird. Diese Variante des Kettenschlusses ermöglicht aufgrund der Berücksichtigung der zwei auf der Potentialhyperfläche weit auseinander liegenden Lösungen MC-Übergänge zwischen Zuständen unabhängig von dazwischen liegenden Energiebarrieren. Eine Überwindung dieser Energiebarrieren zwischen zwei Zuständen ist beim iterativen Kettenschluß hingegen von dieser Barriere abhängig und erfordert im Vergleich zum analytischen Kettenschluß eine große Zahl von MC-Übergängen. Aus diesem Grund ermöglicht das effektivere MC-Sampling bei Anwendung des analytischen Kettenschlusses eine Equilibrierung der MC-Simulation in einer deutlich geringeren Zahl von MC-Makrozyklen.

Der analytische Kettenschluß erhöht somit im Vergleich zum iterativen Kettenschluß zwar die Zahl der MC-Variablen von 9 auf 12, benötigt aber weniger Rechenaufwand und erhöht die Effektivität des MC-Samplings, worin der wesentliche Vorteil besteht, denn die notwendige Equilibrierung der MC-Simulationen wird dadurch schneller erreicht. Aus diesen Gründen findet in den im Rahmen dieser Arbeit durchgeführten Rechnungen das in Abb. 4.2 dargestellte und in Tab. 4.3 zusammengefaßte molekulare Modell des analytischen Kettenschlusses Anwendung.

4.1.4. Behandlung der Zuckerfaltung

Nachdem ein molekulares Modell für die Bewegung der Basen in kollektiven Variablen und des Phosphodiester-Rückgrats in inneren Variablen einschließlich eines Kettenschlusses formuliert worden ist, wird das molekulare Modell nun um die Beschreibung der Flexibilität der Desoxyribose erweitert. Das molekulare Modell ermöglicht dann die Variation der gesamten Nukleinsäurestruktur.

Der gefaltete Zuckering kann bei konstanten Bindungslängen entweder durch die 5 endozyklischen Valenzwinkel ω_i oder die 5 endozyklischen Torsionswinkel τ_i beschrieben werden. Die exozyklischen Valenzwinkel der Furanose werden wie beim Nukleinsäure-

Rückgrat nicht variiert. Die Forderung konstanter Bindungslängen enthält die Forderung des Ringschlusses. Unter dieser Nebenbedingung ist die Konformation des Zuckers frei variierbar. Durch diese Zwangsbedingung wird die Zahl der 5 frei variierbaren endozyklischen Winkel auf 4 reduziert. Diese Zahl ergibt sich allgemein für die $2N - 6$ inneren Freiheitsgrade eines N -Ringes, dessen konstante Bindungslängen N Zwangsbedingungen ergeben. Somit kann der Faltungsmodus der Furanose durch 4 Freiheitsgrade ν_i beschrieben werden, wobei die Freiheitsgrade nach Zweckmäßigkeit, z. B. zwei Torsionswinkel τ_i und zwei Valenzwinkel ω_i , gewählt werden können.

Bei der Variation der Zuckerkonformation, also der Umfaltung des Zuckerrings infolge eines MC-Überganges, werden das C1'-Atom als im Raum fixiert betrachtet und alle anderen Atome, wie in Abb. 4.3 ersichtlich, bewegt. Diese 6 frei beweglichen Atome führen bei einer Beschreibung in kartesischen Koordinaten zu $3 \times 6 = 18$ Freiheitsgraden. Um diese 18 Freiheitsgrade durch 4 innere Freiheitsgrade ν_i in Gestalt endozyklischer Valenz- und Torsionswinkel ersetzen zu können, sind somit $18 - 4 = 14$ Zwangsbedingungen notwendig. Als Zwangsbedingungen eignen sich die 7 beteiligten Bindungslängen. Diese 7 Zwangsbedingungen können dann beispielsweise durch 3 auf die Orientierung der O1'-C1'-C2'-Gruppe, 2 auf die Orientierung des C5'-Atoms und 2 auf die Orientierung des O3'-Atoms wirkende Zwangsbedingungen ergänzt werden. Die Orientierung der O1'-C1'-C2'-Gruppe relativ zur C1'-N-Bindung ist wie die Orientierung der C4'-C5'- und C3'-O3'-Bindungen relativ zum Zuckerring nicht frei variierbar. Die 7 Zwangsbedingungen für die Bindungslängen und 7 Zwangsbedingungen für die Orientierungs-

freiheitsgrade ergeben die geforderten 14 Zwangsbedingungen c_j . Das in Kap. 3.2.5 eingeführte Volumenelement 3.75 kann für die Transformation der 18 kartesischen Koordinaten in die 4 endozyklischen Valenz- bzw. Torsionswinkel ν_i formuliert werden:

$$d\Omega = \prod_{i=1}^6 dx_i dy_i dz_i = J \prod_{i=1}^4 d\nu_i \prod_{j=1}^{14} dc_j \quad (4.13)$$

Die Bedingung des Ringschlusses ist in den auf alle am Ring beteiligten Bindungslängen wirkenden Zwangsbedingungen enthalten.

Reduktion der Freiheitsgrade durch Einführung von Phase und Amplitude

Im weiteren erfolgt eine Auswahl zweier Valenzwinkel ω_i und zweier Torsionswinkel τ_i , um diesen Satz von 4 Freiheitsgraden durch die in Kap. 2.2.3 eingeführten Variablen Phase P und Amplitude A [6, 19] unter Benutzung empirisch gewonnener Transformationsgleichungen zu ersetzen [152]. Grundlage der empirischen Transformationsgleichungen sind die mit dem Kraftfeld FLEX 3.18 generierten Oberflächen $\omega_i(P, A)$ und

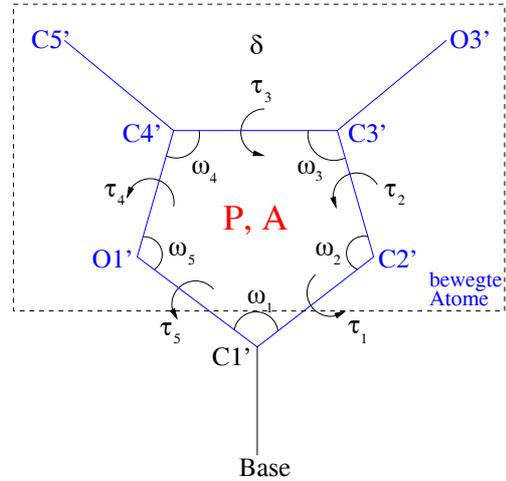


Abb. 4.3.: Molekulares Modell des Zuckerrings mit den bewegten Atomen in blau und den MC-Variablen in rot

$\tau_i(P, A)$, die jeweils in der Dimension der Phase P ein periodisches und in der Dimension der Amplitude A ein quadratisches Verhalten zeigen [152]:

$$\omega_i(P, A) : \quad (1 + A + A^2)(1 + \cos P + \sin P + \cos 2P + \sin 2P) \quad (4.14)$$

$$\tau_i(P, A) : \quad (1 + A + A^2)(1 + \cos P + \sin P) \quad (4.15)$$

Diese funktionalen Zusammenhänge wurden interpoliert, wodurch die Koeffizienten aller Terme der ausmultiplizierten Gleichungen bestimmt werden konnten [152]. Hierbei werden die 5 Freiheitsgrade $\omega_1, \omega_2, \omega_3, \tau_1$ und τ_2 betrachtet, weil in den Transformationsgleichungen zusätzlich eine Zwangsbedingung, nämlich die des Ringschlusses, approximativ enthalten ist. Diese 5 empirischen Gleichungen ermöglichen die approximative Transformation $(P, A) \rightarrow (\omega_1, \omega_2, \omega_3, \tau_1, \tau_2)$ und damit die Konstruktion der 5 Freiheitsgrade aus den zwei Freiheitsgraden Phase P und Amplitude A der reduzierten Beschreibung. Somit können die 4 in 4.13 eingeführten Freiheitsgrade ν_i und die Zwangsbedingung des Ringschlusses durch die zwei neuen Freiheitsgrade Phase P und Amplitude A unter der Bedingung der Erfüllung der an die funktionalen Formen 4.14 und 4.15 interpolierten Gleichungen ersetzt werden:

$$d\Omega = \prod_{i=1}^6 dx_i dy_i dz_i = J(P, A) dP dA \prod_{j=1}^{16} dc_j \quad (4.16)$$

Da die 18 kartesischen Freiheitsgrade nun durch die lediglich 2 Freiheitsgrade (P, A) ersetzt werden, sind nunmehr $18 - 2 = 16$ Zwangsbedingungen notwendig. Die 5 empirischen Gleichungen stellen dabei bereits 5 Zwangsbedingungen dar, die durch 11 weitere Zwangsbedingungen ergänzt werden müssen. Da mit den aus der Interpolation der funktionalen Zusammenhänge 4.14 und 4.15 hervorgehenden 5 empirischen Gleichungen ein approximativer Ringschluß, und zwar der C4'-O1'-Bindung bereits gefordert wird, sind nur 6 der 7 beteiligten Bindungslängen als Zwangsbedingungen geeignet. Als die verbleibenden 5 Zwangsbedingungen werden jeweils zwei auf die Orientierungen des O3'- und des C5'-Atoms relativ zum Zuckerring und eine auf die Orientierung der O1'-C1'-C2'-Gruppe wirkende Zwangsbedingungen gewählt. Die 16 Zwangsbedingungen c_j setzen sich somit zusammen aus den 5 empirischen Gleichungen, 6 auf Bindungslängen b_j und 5 auf Orientierungsfreiheitsgrade ϑ_j wirkende Zwangsbedingungen.

Jacobi-Faktoren bei der Zuckenumfaltung

Die Jacobi-Faktoren 3.77 der Transformation 4.16 werden damit zu Determinanten von 18×18 -Matrizen:

$$J^{-1}(P, A) = \left| \frac{\partial(P, A, \omega_1, \dots, \omega_3, \tau_1, \tau_2, b_1, \dots, b_6, \vartheta_1, \dots, \vartheta_5)}{\partial(x_1, \dots, x_6, y_1, \dots, y_6, z_1, \dots, z_6)} \right| \quad (4.17)$$

Im Gegensatz zu den leicht zu berechnenden Jacobi-Faktoren beim analytischen Kettenschluß, sind die Jacobi-Faktoren bei der Umfaltung der Furanose Determinanten von 18×18 -Matrizen. Somit ist die Berechnung der Jacobi-Faktoren bei der Zuckenumfaltung sehr rechenaufwendig, weshalb die Jacobi-Faktoren als Funktion $J(P, A)$ der Phase P und der Amplitude A bei der Initialisierung des Algorithmus als 2-dimensionales

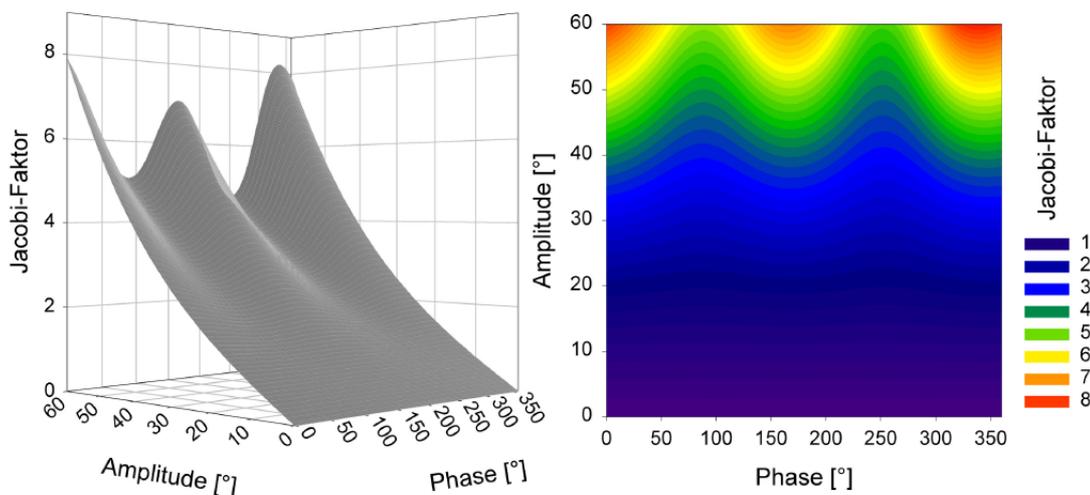


Abb. 4.4.: 3D-Darstellung (links) und Contour-Plot (rechts) der Jacobi-Faktoren der Zuckerkaltung als Funktion von Phase und Amplitude

Feld berechnet werden, so daß der dann jeweils benötigte Wert diesem in Abb. 4.4 dargestellten Feld entnommen werden kann.

Die graphische Darstellung 4.4 der Jacobi-Faktoren $J(P, A)$ als Funktion von Phase P und Amplitude A verdeutlicht den Einfluß der Jacobi-Faktoren $J(m)$ vor und $J(n)$ nach einem MC-Übergang auf das Akzeptanzkriterium des Metropolis-Algorithmus [149, 150]:

$$P(m \rightarrow n) = \min \left\{ 1, \frac{J(n) e^{-\beta E(n)}}{J(m) e^{-\beta E(m)}} \right\} \quad (4.18)$$

Der Wert der Jacobi-Faktoren steigt mit zunehmender Amplitude, wobei die Jacobi-Faktoren in der Nähe der Amplitude $A = 0^\circ$ gegen Null streben. Dieses Verhalten korrespondiert mit einem planaren Zuckerring im Grenzfall $A = 0^\circ$, so daß die Phase keinen Einfluß auf die Konformation hätte. Damit würden die Ableitungen der Phase nach den kartesischen Koordinaten des Zuckers $\rightarrow \infty$ streben. Die Jacobi-Faktoren tragen somit zusätzlich zum energetischen Kriterium zur Verhinderung kleiner Amplituden und zugleich planarer Zuckerkonformationen bei.

Außerdem besitzt die Funktion $J(P, A)$ zwei ausgeprägte Bergrücken für $A > 40^\circ$ jeweils bei Phasen $P \approx 170^\circ \pm 30^\circ$ und $P \approx 350^\circ \pm 50^\circ$. Die Akzeptanz eines MC-Übergangs $m \rightarrow n$ wird durch einen größer werdenden Jacobi-Faktor, also $J(n) > J(m)$, begünstigt. Somit wirken die Jacobi-Faktoren im Auswahlkriterium 4.18 des MC-Algorithmus zusätzlich zu den Energien durch die Ablehnung kleiner Amplituden und die Begünstigung der Faltungsmodi C2'-endo und C2'-exo/C3'-endo.

Der einzige beim Kettenschluß des Nukleinsäure-Rückgrats invariante Torsionswinkel δ wird infolge der Variation des Zuckerkaltungsmodus nun ebenfalls variiert. Die 7 Wasserstoffatome an den C1'-, C2'-, C3'-, C4'- und C5'-Atomen werden schließlich wie beim Nukleinsäure-Rückgrat die Phosphatsauerstoffe und Wasserstoffe am C5'-Atom nach einer erfolgten Umfaltung des Zuckerrings unter Erhaltung der exozyklischen Valenzwinkel in den Mittelsenkrechenebenen an dem jeweiligen Atom positioniert.

4.1.5. Explizite Gegenionen

Die polyanionischen Eigenschaften von Nukleinsäuren und der daraus resultierende Einfluß von Ionen auf Nukleinsäurekonformationen ist in Kap. 2.4.1 diskutiert worden. Manning beschrieb die Kondensation kationischer Gegenionen an der Oberfläche der DNA, um deren polyanionische Ladung partiell zu kompensieren [47]. Jayaram et al. konnten in MC-Simulationen die von Manning apostrophierte Kondensation der Gegenionen an der DNA-Oberfläche qualitativ bestätigen, wobei die DNA-Struktur fixiert und die Ionen frei beweglich waren [48, 49].

Das molekulare Modell wird um explizite Gegenionen erweitert, um dem signifikanten Einfluß der Gegenionen auf Nukleinsäurestrukturen Rechnung zu tragen. Als explizite Gegenionen werden einfach positiv geladene Natrium-Ionen dem System hinzugefügt. Die Zahl der verwendeten expliziten Gegenionen entspricht dabei der Zahl der Phosphatgruppen der Nukleinsäurestruktur, so daß das gesamte System Nukleinsäure-Gegenionen elektrisch neutral wird. Diese Elektroneutralität des zu modellierenden Systems kann als Vorteil des molekularen Modells gegenüber den bei Energieminimierungen angewandten molekularen Modellen bewertet werden. Außerdem geht die Poisson-Boltzmann-Gleichung bei expliziter Behandlung der Ionen in die Poisson-Gleichung über, wodurch Näherungen wie die Linearisierung der PB-Gleichung durch eine exakte Beschreibung ersetzt werden.

Die in das molekulare Modell der MC-Simulationen eingeführten expliziten Gegenionen sind im 3-dimensionalen Raum frei beweglich. Die Bewegung der Ionen wird durch drei Translationen t_i in die drei zueinander orthogonalen Raumrichtungen beschrieben. Der Betrag dieser Translationen t_i wird durch Zufallszahlen bestimmt. Die den Bewegungen der Ionen zuzuordnenden MC-Übergänge werden mit dem Metropolis-Algorithmus bewertet [143]. Über die Akzeptanz wird wie bei den MC-Übergängen infolge von Konformationsänderungen der Nukleinsäuren entschieden, wobei aufgrund der Bewegung der Ionen in kartesischen Koordinaten im Vergleich zu 4.18 keine Jacobi-Faktoren zu berücksichtigen sind:

$$P(m \rightarrow n) = \min \left\{ 1, \frac{e^{-\beta E(n)}}{e^{-\beta E(m)}} \right\} \quad (4.19)$$

Um auszuschließen, daß sich Ionen sehr weit von der DNA entfernen, wird um die DNA ein Zylinder gelegt, dessen Achse mit der linear approximierten helikalen Achse der DNA übereinstimmt. Die Ionen werden bei MC-Übergängen, die sie aus dem Inneren dieses Zylinders hinausbewegen würden, an der Mantelfläche des Zylinders bzw. seiner Grund- oder Deckfläche in das Innere des Zylinders reflektiert. Der Radius dieses Zylinders wird mit $R = 100 \text{ \AA}$ so gewählt, daß die DNA-Konzentration unter 1 mM liegt. Dieser bei allen MC-Simulationen im Rahmen dieser Arbeit angewandte Zylinderradius ist im Vergleich zu einem Radius $R \approx 10 \text{ \AA}$ einer DNA in B-Form hinreichend groß. Die Einschränkung der Beweglichkeit der Ionen auf das Innere dieses Zylinders ist gerechtfertigt, weil die Wechselwirkung eines Ions am Rande des Zylinders mit der DNA bereits vernachlässigt werden kann. Bei der Modellierung nicht zylindrischer Nukleinsäurekonformationen bzw. Komplexe mit Liganden oder Proteinen kann wahlweise eine Kugelsphäre gewählt werden, an deren Oberfläche die Ionen ins Innere reflektiert werden. Der Mittelpunkt dieser Kugelsphäre befindet sich im Schwerpunkt des molekularen Systems und bewegt sich mit diesem mit.

Die Einführung expliziter Gegenionen führt zur Elektroneutralität des Gesamtsystems, erhöht aber die Zahl der MC-Variablen um 3 Translationsfreiheitsgrade t_i pro explizitem Gegenion, wobei die Zahl der Gegenionen der Zahl der Phosphatgruppen der Nukleinsäure entspricht. Werden DNA-Fragmente untersucht, besitzen die Nukleotide an den 3'-Enden keine Phosphatgruppen, so daß die Zahl der Gegenionen dann der Zahl der beteiligten Nukleotide abzüglich der Nukleotide an den 3'-Enden entspricht. Bildet die polyanionische Nukleinsäure einen Komplex mit einem kationischen Liganden, kompensiert dieser bereits einen Teil der polyanionischen Ladung, weshalb dann wiederum die Zahl der Gegenionen um die Zahl der ganzen kationischen Ladungen des Liganden reduziert werden sollte. Zusätzlich zu den Gegenionen können weitere positiv wie negativ geladene Ionen dem System beigefügt werden, um Salzeffekte auf die Konformation der Nukleinsäuren zu untersuchen.

4.1.6. Molekulares Modell für die Ligandenbindung

Der Einführung eines molekularen Modells für die Bindung eines Liganden an die DNA wird die in Kap. 2.5.1 diskutierte Bindung des heterozyklischen Farbstoffes Metylenblau an die DNA zugrunde gelegt, die natürlich auf weitere DNA-bindende Liganden ausgedehnt werden kann. Aus der chemischen Struktur des Metylenblau lassen sich die in der MC-Simulation zu berücksichtigenden Freiheitsgrade des Liganden ableiten. Aus Abb. 4.5 wird das ausgedehnte π -Elektronensystem des Farbstoffs ersichtlich, das ihm einerseits seine in Kap. 2.6 diskutierten, medizinisch relevanten optischen Eigenschaften verleiht, andererseits seine innere Freiheit stark einschränkt, denn das π -Elektronensystem kann wie bei den Nukleinsäurebasen in guter Näherung als starrer Körper behandelt werden. Dieser starre Körper wird in der MC-Simulation wie die Basen mittels 6 Freiheitsgraden, nämlich 3 Translationen T_i^L und 3 Rotationen R_i^L , relativ zur DNA bewegt. Um diese Bewegungen ausführen zu können, muß ein körperfestes lokales Koordinatensystem des Liganden definiert werden. Die Berechnung dieses körperfesten Koordinatensystems des Liganden erfolgt parallel zu den Berechnungen der in Kap. 2.2.1 eingeführten körperfesten Koordinatensysteme $(\vec{j}_i, \vec{k}_i, \vec{l}_i)$ der Basen im Initialisierungsschritt des MC-Programms. In Abb. 4.5 sind die x- und die y-Achse dieses Koordinatensystems

dargestellt. Die z-Achse zeigt als Senkrechte zu diesen beiden Achsen mit ihrem Ursprung im Zentrum des Schwefel-Atoms aus der Ebene heraus. Dieses körperfeste Koordinatensystem findet sowohl für die Festlegung der MC-Bewegungen des Liganden als auch für die Berechnung der Lage und Orientierung des Liganden relativ zur helikalen Achse bzw. zu benachbarten Basen Verwendung [53].

Zusätzlich zu diesen 6 Freiheitsgraden weist Metylenblau 4 innere Rotationsfrei-

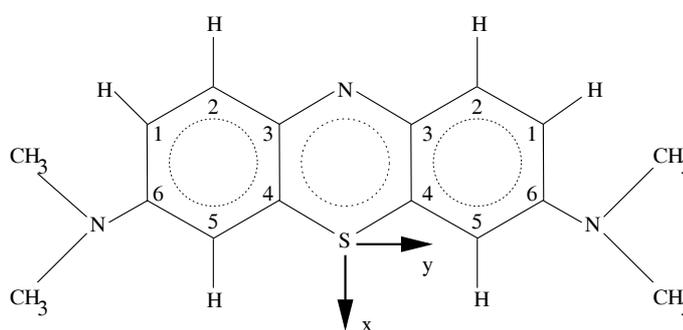


Abb. 4.5.: Chemische Struktur des DNA-bindenden Liganden Metylenblau und Definition des körperfesten Koordinatensystems

heitsgrade ψ_i auf, die die freie Rotation der Methylgruppen um die C-N-Bindungen ermöglichen. Diese 4 Rotationsfreiheitsgrade ψ_i werden wie die Rotation ϕ der Methylgruppe am Thymin als freie MC-Variablen behandelt. Das molekulare Modell führt somit zu 10 Freiheitsgraden des Liganden Methylenblau, wobei diese 10 MC-Variablen sich aus 6 äußeren und 4 inneren Freiheitsgraden des Liganden zusammensetzen.

Das molekulare Modell muß allerdings auch die totale Ladung des Liganden berücksichtigen, um auch für den Liganden-DNA-Komplex Elektroneutralität des Gesamtsystems DNA-Gegenionen-Ligand zu gewährleisten. Da das Methylenblau eine totale kationische Ladung $q = +e$ trägt, wird die Zahl der Gegenionen um ein Ion reduziert, womit die Elektroneutralität wieder hergestellt ist.

Dieses hier für Methylenblau entwickelte molekulare Modell läßt sich leicht auf andere nicht kovalent gebundene Liganden erweitern. Die 6 äußeren Freiheitsgrade erfordern für die Bewegung des Liganden die Definition eines körperfesten Koordinatensystems. Darüberhinaus sind die inneren Freiheitsgrade des Liganden zu analysieren und als MC-Variablen zu definieren.

4.1.7. Gesamtdarstellung des molekularen Modells in Relation zu vergleichbaren Modellen

Das für MC-Simulationen von Nukleinsäuren entwickelte molekulare Modell ist geometrisch charakterisiert worden und wird in Form seiner als MC-Variablen zu verwendenden Freiheitsgrade in Tab. 4.4 zusammengefaßt. Dieses molekulare Modell wird in dem MC-Algorithmus durch die Kraftfelder FLEX 3.18 und AMBER 3.19 zur Energieberechnung komplettiert [103, 105, 106]. Die in dieser Arbeit präsentierten ersten Resultate mit diesem MC-Algorithmus für Nukleinsäuren wurden unter Anwendung des Cornell et al. AMBER-Kraftfeldes [106] erzielt, dessen analytische Form in Kap. 3.1.2 mit Gleichung 3.19 angegeben und diskutiert worden ist.

Das molekulare Modell zeichnet sich durch die Verwendung der in Tab. 4.4 zusam-

MC-Variablen je Nukleotid, explizitem Gegenion bzw. des Liganden Methylenblau	
$3 T_i, 3 R_i$	Translationen und Rotationen des Nukleotids
P, A	Phase und Amplitude des Zuckers
χ	Glykosidwinkel O1'-C1'-N-C6/C8
ε, γ	Torsionswinkel C4'-C3'-O3'-P und O3'-C5'-C4'-C3'
ω_1	Valenzwinkel C3'-O3'-P
ϕ	Torsionswinkel der frei drehbaren Methylgruppe am Thymin
$3 t_i$	Translationen der expliziten Gegenionen
$3 T_i^L, 3 R_i^L$	Translationen und Rotationen des Liganden
4ψ	Torsionswinkel der frei drehbaren Methylgruppen des Liganden

Tabelle 4.4.: Aufstellung aller MC-Variablen des verwendeten molekularen Modells

mengefaßten MC-Variablen in Verbindung mit einem analytischen Kettenschluß und durch die Berücksichtigung der Jacobi-Faktoren sowohl beim Kettenschluß als auch bei der Variation der Zuckerfaltung aus. Zudem wird die polyanionische Ladung der Nukleinsäuren durch frei bewegliche explizite Gegenionen kompensiert. Damit befindet sich dieses molekulare Modell in Relation zu einigen anderen vergleichbaren molekularen Modellen:

Lokale Konformationsänderungen kettenförmiger Moleküle

Gö und Scheraga führten erstmals einen Algorithmus für Konformationsänderungen ketten- und ringförmiger Moleküle im Raum der endozyklischen Torsionswinkel als innere Variable ein, der die Invarianz der beteiligten Bindungslängen und Valenzwinkel voraussetzt [153]. Durch die Einführung dieser inneren Variablen konnten die Zahl der Freiheitsgrade molekularer Systeme und damit Rechenzeiten bei Simulationen im Vergleich zur Beschreibung in kartesischen Koordinaten deutlich reduziert werden. Gö und Scheraga ergänzten die Beschreibung in Torsionswinkeln durch einen Algorithmus zur Berechnung eines einen Ringschluß gewährleistenden Satzes endozyklischer Torsionswinkel [153]. Außerdem begrenzten Gö und Scheraga die Konformationsänderungen langer kettenförmiger Moleküle auf lokale Konformationsänderungen durch die Auswahl eines Satzes benachbarter, sich simultan ändernder Torsionswinkel [153].

Auch das in Tab. 4.4 zusammengefaßte molekulare Modell definiert lokale Konformationsänderungen des Nukleinsäure-Rückgrats durch Verwendung innerer Variablen bei invarianten Bindungslängen und exozyklischen Valenzwinkeln.

Konformationsänderungen von Nukleinsäuren unter Zwangsbedingungen im Raum innerer Variablen

Sklenar et al. entwickelten einen Algorithmus, der die Variation der Konformation von Nukleinsäuren im Raum innerer Variablen unter Berücksichtigung von Zwangsbedingungen ermöglicht [102]. Dieser Algorithmus berechnet einen Satz endozyklischer Torsions- und Valenzwinkel des Desoxyribose-Phosphodiester-Rückgrats unter der Zwangsbedingung, daß die Lage und Orientierung der Basen fixiert sind. Die Zwangsbedingungen werden in differentieller Form als Funktionen des gewählten Satzes innerer Variablen formuliert, so daß diese Differentialgleichungen durch sukzessive infinitesimale Rotationen (SIR) der als Variablen definierten Torsions- und Valenzwinkel erfüllt werden [102]. Dabei sind alle Bindungslängen und Valenzwinkel mit Ausnahme der endozyklischen Valenzwinkel des Zuckerrings invariant. Sklenar et al. formulierten eine Ringschlußbedingung des flexiblen Furanoseringes, dessen Konformation einschließlich des Ringschlusses durch 5 innere Variablen beschrieben werden kann [102].

Das in Tab. 4.4 dargestellte molekulare Modell benutzt ebenfalls innere Variablen. Ein wesentlicher Unterschied besteht aber darin, daß im SIR-Algorithmus nur Torsionswinkel, in dem in dieser Arbeit formulierten MC-Algorithmus aber auch die endozyklischen Valenzwinkel variiert werden. Der Kettenschluß erfolgt im SIR-Algorithmus numerisch, in dem hier vorgeschlagenen MC-Algorithmus analytisch, weshalb einige Torsions- und Valenzwinkel als MC-Variablen behandelt werden, wohingegen andere als abhängige Variablen durch den Kettenschluß variiert werden. Außerdem wird der Zucker durch Phase und Amplitude und nicht durch Torsions- oder Valenzwinkel beschrieben.

Energieminimierung von Nukleinsäuren im Raum helikaler und innerer Variablen

Lavery et al. kombinierten in dem für Nukleinsäuren entwickelten Energieminimierungs-Algorithmus JUMNA [104] die Beschreibung des Desoxyribose-Phosphodiester-Rückgrats in inneren Variablen in Analogie zum SIR-Algorithmus [102] mit einer Beschreibung der Lage und Orientierung der Basen als starre Körper in einem helikalen Achsensystem durch drei Translationen und drei Rotationen.

Die in dem molekularen Modell 4.4 für den MC-Algorithmus vorgeschlagenen kollektiven Variablen T_i und R_i beschreiben wie die in JUMNA verwendeten helikalen Variablen die Lage und Orientierung der Basen im Raum, sind jedoch nicht mit helikalen Variablen gleichzusetzen, da sie sich nicht auf ein helikales Achsensystem, sondern nur auf die lokalen körperfesten Achsensysteme der Basen beziehen.

Monte-Carlo-Simulationen von DNA im Raum innerer Variablen

Gabb et al. benutzten einen Satz innerer Variablen, der sich aus den endozyklischen Torsionswinkeln des Nukleinsäure-Rückgrats, den Glykosidwinkeln sowie der Phase und der Amplitude des Zuckers zusammensetzt und durch Translationen und Rotationen der beiden DNA-Stränge relativ zueinander ergänzt wird, als Variablen bei MC-Simulationen von Nukleinsäurestrukturen [154]. Im Gegensatz zu dem in dieser Arbeit vorgeschlagenen MC-Algorithmus erwies sich das MC-Sampling in diesen Variablen als sehr ineffektiv, da eine Vielzahl von MC-Übergängen zu sterischen Kollisionen führte, die sich dadurch erklären lassen, daß von Gabb et al. definierte MC-Variablen globale Konformationsänderungen generieren [154]. Sowohl die endozyklischen Torsionswinkel als auch erst recht die Translationen und Rotationen eines kompletten DNA-Stranges relativ zu dem zweiten DNA-Strang in einer Doppelhelix vermeiden nur bei sehr geringen Bewegungen sterische Kollisionen in einem MC-Übergang. Das ineffektive MC-Sampling verhindert eine Equilibrierung des Systems in der MC-Simulation.

Aus diesen Gründen basiert der in dieser Arbeit vorgeschlagene MC-Algorithmus auf lokalen Bewegungen auf der Ebene von Nukleotiden, die im Gegensatz zu dem MC-Algorithmus von Gabb et al. zwar Kettenschlüsse notwendig machen, dafür aber mit größeren lokalen Bewegungen ein effektives MC-Sampling und eine Equilibrierung ermöglichen. Gabb et al. behandeln die Zuckerkonformation wie in dem in Tab. 4.4 dargestellten molekularen Modell auch durch Phase und Amplitude [152, 154], berücksichtigen aber im gesamten Algorithmus keine Jacobi-Faktoren.

Jacobi-Faktoren bei MC-Simulationen von Polymeren im Raum innerer Variablen

Dodd et al. entwickelten einen MC-Algorithmus für langkettige Polymere, bei dem ein zufällig ausgewählter endozyklischer Torsionswinkel in einem MC-Übergang variiert wurde, der die Variation von benachbarten Torsionswinkel nach sich zog, um die Konformationsänderung lokal zu beschränken [149]. Bei diesen Konformationsänderungen bleiben Bindungslängen und Valenzwinkel invariant. Dodd et al. führten Jacobi-Faktoren ein, die den Einfluß der Zwangsbedingungen auf das MC-Sampling berücksichtigen [149]. Die als konzertierte Rotation (CONROT) bezeichnete Konformationsänderung von Polymeren im Raum endozyklischer Torsionswinkel wurde von Pant und Theodorou als die Bewegung zweier, ein Trimer einschließender Atome durch die Variation endozyklischer Torsionswinkel definiert, in dessen Folge das Trimer rekonstruiert werden mußte, um die Polymerkette wieder zu schließen [150]. Außerdem wurden MC-Übergänge definiert, die neue Kettenschlüsse generieren, wobei bei der Existenz mehrerer Lösungen eines Kettenschlusses Auswahlwahrscheinlichkeiten berücksichtigt werden [150]. Für diese beiden alternativen Varianten eines MC-Algorithmus für Polymere im Raum innerer Variablen wurden Jacobi-Faktoren eingeführt [150, 151]. Mavrantzas et al. zeigten, daß bei Einbeziehung der Jacobi-Faktoren in den MC-Algorithmus, auch für

große Kettenlängen gerade im Gegensatz zu MD-Simulationen eine sehr gute Equilibrierung struktureller und thermodynamischer Eigenschaften erzielt werden kann [151]. In Analogie zu der Einführung von Jacobi-Faktoren für Polymere werden in dem in dieser Arbeit benutzten MC-Algorithmus Jacobi-Faktoren definiert, die sowohl die Transformation des molekularen Systems von kartesischen Koordinaten in innere Variablen als auch den Einfluß der Zwangsbedingungen auf das MC-Sampling berücksichtigen.

Lokale Konformationsänderungen bei MC-Simulationen von Proteinen im Raum innerer Variablen ("window moves")

Knapp und Irgens-Defregger entwickelten einen MC-Algorithmus für Proteine, der ein effektives MC-Sampling durch lokale Konformationsänderungen ("window moves") des Protein-Rückgrats im Raum endozyklischer Torsionswinkel ermöglicht [155, 156]. Dabei sind die Bindungslängen und Valenzwinkel Invarianten, und die aufgrund ihrer π -Elektronensysteme annähernd planaren Peptidgruppen werden als starre monomere Einheiten behandelt. Die Freiheitsgrade des Polypeptidstranges sind damit auf die Rotationen um die C_α -C- und die C_α -N-Bindung beschränkt [156]. Lokale Konformationsänderungen betreffen drei aufeinander folgende monomere Einheiten, wodurch ein "window" definiert wird, dessen benachbarte monomere Einheiten nicht bewegt werden [156]. Dieses molekulare Modell erlaubt damit die Variation von 8 endozyklischen Torsionswinkeln in einem MC-Übergang, wobei nach dem molekularen Modell von Gō und Scheraga 6 Zwangsbedingungen erfüllt werden müssen. Knapp und Irgens-Defregger reduzierten diese Zahl auf 4 Zwangsbedingungen durch einen Kettenbruch erzeugende Variationen der Torsionswinkel und Einführung einer Kettenschlußbedingung [156]. Hoffmann und Knapp erweiterten diesen MC-Algorithmus für Proteine durch die Definition von Jacobi-Faktoren, die den Einfluß der Zwangsbedingungen des molekularen Modells auf das MC-Sampling berücksichtigen [157]. Dieser MC-Algorithmus im Raum innerer Variablen zeigte bei Verwendung des Kraftfeldes CHARMM [107] und konsequenter Einbeziehung der Jacobi-Faktoren ein effektives MC-Sampling und erwies sich aufgrund der Fähigkeit, MC-Übergänge zwischen Zuständen im Konfigurationsraum unabhängig von dazwischen liegenden Energiebarrieren realisieren zu können, als geeignete Alternative zu MD-Simulationen von Proteinen [158, 159, 160].

Ein Unterschied gegenüber dem in dieser Arbeit dargestellten MC-Algorithmus besteht darin, daß die molekularen Modelle den zu untersuchenden Systemen, also Proteinen bzw. Nukleinsäuren, angepaßt sind. Gemeinsamkeiten beider Algorithmen sind die Verwendung innerer Variablen als MC-Variablen, lokale Konformationsänderungen und die Berücksichtigung von Jacobi-Faktoren beim MC-Sampling. Unterschiede bestehen in den Kraftfeldern, der Verwendung kollektiver Variablen zur Beschreibung der Lage und Orientierung der Basen als monomere Einheiten und die Einführung expliziter Gegenionen zur Kompensation der polyanionischen Ladung der Nukleinsäuren.

MC-Simulationen des Arrangements expliziter Gegenionen an der DNA

Jayaram et al. untersuchten das Arrangement expliziter Gegenionen in der Umgebung einer B-DNA in MC-Simulationen frei im Raum beweglicher expliziter Gegenionen [48]. Die von Manning vorhergesagte Kondensation von Gegenionen an der Oberfläche der DNA [47] konnten Jayaram et al. dabei qualitativ erstmals unter Berücksichtigung

der atomaren DNA-Struktur bestätigen [48, 49]. Dieses Resultat konnte unabhängig von der Wahl der in Kap. 3.1.3 diskutierten Modelle zur Beschreibung des Lösungsmittels durch eine abstandsabhängige Dielektrizitätsfunktion [110] oder durch die FDPB-Methode [115] erzielt werden.

Die durch Jayaram et al. etablierte Einführung expliziter Gegenionen in MC-Simulationen [48] ist in Kombination mit einer impliziten Beschreibung des Lösungsmittels durch eine sigmoidale abstandsabhängige Dielektrizitätsfunktion 3.20 Bestandteil des in Tab. 4.4 dargestellten molekularen Modells. Der wesentliche Unterschied besteht darin, daß in dem in dieser Arbeit verwendeten MC-Algorithmus neben den freien Bewegungen der expliziten Gegenionen die Nukleinsäurestruktur nicht fixiert ist, sondern Konformationsänderungen der Nukleinsäuren der wesentliche Bestandteil des MC-Samplings sind.

Zusammenfassend läßt sich einschätzen, daß das im Rahmen dieser Arbeit benutzte molekulare Modell in bestimmten Aspekten Beziehungen zu den anderen Modellen aufweist, es jedoch die Grundlage des ersten für Nukleinsäuren entwickelten MC-Algorithmus ist, der mit lokalen Bewegungen überhaupt ein effektives MC-Sampling ermöglicht. Neben den in anderen molekularen Modellen variierten Torsionswinkeln werden auch die endozyklischen Valenzwinkel variiert, wodurch sich die Flexibilität der Strukturen und damit die Effektivität des MC-Samplings erhöht. Sowohl beim analytischen Kettenschluß als auch bei der Zuckerfaltung berücksichtigt dieser MC-Algorithmus Jacobi-Faktoren. Außerdem verbindet dieser MC-Algorithmus erstmals ein MC-Sampling der Nukleinsäurestruktur mit der freien Bewegung expliziter Gegenionen, wobei die implizite Beschreibung des Lösungsmittels auch ein effektives MC-Sampling der Bewegungen der expliziten Gegenionen gestattet. Schließlich integriert das molekulare Modell des vorgeschlagenen MC-Algorithmus die DNA-Bindung eines Liganden einschließlich der inneren Freiheitsgrade des Liganden.

4.2. Programmtechnische Implementierung

4.2.1. Initialisierung

Im Initialisierungsschritt des MC-Algorithmus werden alle notwendigen Parameter vereinbart, es wird die Startkonfiguration des zu untersuchenden molekularen Systems eingelesen, und es werden energetische und strukturelle Startwerte der Ausgangskonfiguration berechnet. Die festzulegenden Parameter gliedern sich in drei Gruppen:

Die Parameter der ersten Gruppe konkretisieren das molekulare Modell: So kann zwischen den Kraftfeldern FLEX 3.18 und AMBER 3.19 gewählt werden. Die abstandsabhängige Dielektrizitätsfunktion 3.20 wird durch $\varepsilon(r \rightarrow 0) = 1$ und $\varepsilon(r \rightarrow \infty) = 78$ sowie durch einen Anstieg $s = 0,356$ im sigmoidalen Bereich nach Hingerty et al. [110] charakterisiert. Die Erweiterung des molekularen Modells um explizite Gegenionen wird durch die Festlegung der Ionenradien $R_{Ion} = 3 \text{ \AA}$ und des Radius $R_{Zyl} = 100 \text{ \AA}$ eines in Kap. 4.1.5 eingeführten Zylinders ergänzt, an dessen Grenzflächen die Ionen bei MC-Übergängen, durch die sie ins Äußere gelangen würden, reflektiert werden.

Die Parameter der zweiten Gruppe sind für den Metropolis-Algorithmus relevant: Dazu gehören die in Kap. 4.2.3 diskutierten Maximal-, Minimal- und Startwerte für die Schrittweiten sämtlicher möglicher MC-Übergänge. Die das Boltzmann-Gewicht im Ak-

zeptanzkriterium bestimmende Temperatur wird mit $\beta^{-1} = k_B T = 0,6$ kcal/mol auf $T \approx 300$ K festgelegt. Ein Makrozyklus ist eine Abfolge von MC-Übergängen, in der jede MC-Variable einmal variiert wird, wobei über die Akzeptanz jeweils nach dem Metropolis-Algorithmus entschieden wird. Eine Gesamtzahl von $n = 10^6$ Makrozyklen hat sich für MC-Simulationen molekularer Systeme in der Größenordnung, wie sie in dieser Arbeit behandelt werden, als geeignet erwiesen.

Die Parameter der dritten Gruppe regeln technische Details des Simulationsprotokolls: Neben einem Listing, das das Verfolgen der MC-Simulation möglich macht, wird eine Trajektorie der Atomkoordinaten generiert, die der detaillierten Auswertung der Simulation dient. Die Trajektorie wird jeweils nach $\Delta n = 10$ Makrozyklen durch die momentanen Atomkoordinaten ergänzt. Die Wahl des Parameters $\Delta n = 10$ hat sich als Kompromiß zwischen der durch die Simulation erzeugten Datenmenge und der für die Auswertung der Simulation erforderlichen Genauigkeit als geeignet erwiesen. Das MC-Programm ermöglicht einen Restart nach einem Abbruch der Simulation, wozu eine Datei, die alle für einen Restart nötigen Informationen enthält, ständig aktualisiert wird. Ein Restart kann durch Parameter initialisiert werden.

Teil der Initialisierung ist das Einlesen der Startkonfiguration des zu untersuchenden molekularen Systems: Dazu werden für alle durch Namen und eine Nummerierung charakterisierten Atome i die kartesischen Koordinaten (x_i, y_i, z_i) , die für die Wahl der Kraftfeld-Parameter relevanten Atomtypen und Atomklassen, die Ladungen q_i sowie eine Bindungsmatrix des molekularen Systems eingelesen. Im Initialisierungsschritt werden die Startenergie und Startwerte der strukturellen Parameter berechnet. Um helikale Parameter berechnen und Konformationänderungen in MC-Übergängen realisieren zu können, ist zuvor die Berechnung der körperfesten Referenzsysteme $(\vec{j}, \vec{k}, \vec{l})$ der Basen und eines eventuellen Liganden notwendig. Die O5'-C5'-Bindungslänge im Nukleinsäure-Rückgrat der Startkonfiguration wird durch einen initialen Kettenschluß dem geforderten Wert angepaßt. Schließlich werden die Jacobi-Faktoren der Startkonfiguration berechnet.

4.2.2. Energieberechnung

Für die Energieberechnung wird in dieser Arbeit das Kraftfeld AMBER 3.19 als Standard definiert [106]. Das Kraftfeld AMBER enthält, wie in Kap. 3.1.2 bereits erläutert, die Terme E_{Ang} und E_{Tor} , die die Deformation der Valenz- und Torsionswinkel beschreiben, einen Term E_{LJ} , der die van-der-Waals-Wechselwirkung durch ein Lennard-Jones-Potential charakterisiert, und einen Term E_{elec} , der die elektrostatische Wechselwirkung durch ein Coulomb-Potential beschreibt. Bei den paarweisen Wechselwirkungen der Atome werden Wechselwirkungen direkter Nachbarn und nächster Nachbarn vernachlässigt. Van-der-Waals-Wechselwirkungen übernächster Nachbarn werden um den Faktor $f_1 = 2$ und elektrostatische Wechselwirkung dieser Nachbarn um den Faktor $f_2 = 1,2$ gedämpft [106]. Diese Charakteristik des Kraftfeldes AMBER erfordert im Initialisierungsschritt die Erstellung eines Feldes, aus dem für alle Atome die direkten, nächsten und übernächsten Nachbarn hervorgehen.

Die einzelnen Energiebeiträge werden getrennt voneinander berechnet und ergeben in der Summe die für das Akzeptanzkriterium des Metropolis-Algorithmus einzig relevante totale Energie:

$$E_{tot} = E_{Ang} + E_{Tor} + E_{LJ} + E_{elec} \quad (4.20)$$

Diese totale Energie ist beim Start der MC-Simulation eine potentielle Energie, wenn die Startkonformation durch Energieminimierung erzeugt worden ist. Der Beginn einer MC-Simulation führt deshalb zunächst zu einem Austausch zwischen potentieller und kinetischer Energie, so daß die totale Energie der Startkonfiguration sich in den ersten Makrozyklen in Richtung positiverer Energiewerte bewegt, da nach dem Gleichverteilungstheorem 3.33 jeder Freiheitsgrad des Systems im thermodynamischen Gleichgewicht eine mittlere kinetische Energie von $k_B T/2$ beiträgt.

Programmtechnisch ist die Energieberechnung so organisiert, daß nach einem MC-Übergang nur diejenigen Beiträge zu den Energietermen neu berechnet werden, die Atome oder kollektive Einheiten betreffen, die sich in dem MC-Übergang auch bewegt haben. Dadurch kann der Rechenaufwand bei der Energieberechnung ganz erheblich reduziert werden, denn schließlich erfordert die Bewegung weniger Atome auch nur die Berechnung der betroffenen Torsions- und Valenzwinkel sowie der paarweisen Wechselwirkungen der bewegten Atome mit den anderen Atomen des molekularen Systems.

In den in Kap. 5 diskutierten MC-Simulationen freier DNA-Dekamere erfolgte eine vermehrte Besetzung von B_I -Zuständen, die als Effekt der einfachen Beschreibung der elektrostatischen Wechselwirkung betrachtet wird. Deshalb wird die Energiebarriere der Torsionswinkel für jedes Nukleotid n modifiziert:

$$E_{Tor}(n) = E_{Tor}(n) + (\varepsilon(n) - \zeta(n) + 1,25)^2 \quad (4.21)$$

Diese Modifikation des Kraftfeldes ist für Konformationen im B_I -Zustand wegen $\varepsilon - \zeta \approx -1,25$ nahezu nicht wirksam. Wegen eines bei den in Kap. 7 diskutierten MC-Simulationen von Interkalatorkomplexen vereinzelt auftretenden Kinks der Target-DNA an der Interkalationstasche wird der Torsionswinkel β durch eine Modifikation des Kraftfeldes auf eine trans-Konformation festgelegt:

$$E_{Tor}(n) = E_{Tor}(n) + (|\beta(n)| - \pi)^2 \quad (4.22)$$

Befindet sich der Torsionswinkel β in der trans-Konformation ist auch diese Modifikation nahezu nicht wirksam. Experimentelle Strukturdaten zeigen, daß sowohl die B_I -Konformation als auch die trans-Konformation des Torsionswinkels β in der B-Form der DNA vorherrschend sind.

Die Equilibrierung einer MC-Simulation zeichnet sich durch einen konstanten Mittelwert der totalen Energie entlang der MC-Trajektorie aus:

$$\langle E_{tot} \rangle = \frac{1}{N} \sum_{i=1}^N E_{tot}(i) \quad (4.23)$$

Neben dem Mittelwert der totalen Energie liefert die Fluktuation σ der totalen Energie Erkenntnisse über die statistische Energieverteilung der repräsentativen Stichproben des MC-Samplings:

$$\sigma(E_{tot}) = \sqrt{\langle E_{tot}^2 \rangle - \langle E_{tot} \rangle^2} \quad (4.24)$$

Die Fluktuation ist dabei abhängig von der Temperatur $T \approx 300$ K. Eine Aufheizung des Systems würde zu einer Erhöhung der Energie-Fluktuation führen.

4.2.3. Dynamische Schrittweitenanpassung

Die in Kap. 3.2.3 eingeführte Schrittweite δ eines MC-Übergangs reguliert die Variation der jeweiligen MC-Variablen und der damit verbundenen Konformationsänderung hinsichtlich ihres Ausmaßes. Eine größere Schrittweite führt zu einer umfassenderen Konformationsänderung, die jedoch eher durch das Akzeptanzkriterium des Metropolis-Algorithmus abgelehnt werden könnte [143]. Daher muß die Schrittweite so gewählt werden, daß sie einerseits ein effektives MC-Sampling im Konformationsraum ermöglicht, andererseits aber noch die Akzeptanz von hinreichend vielen MC-Übergängen garantiert. Die Anpassung der Schrittweite δ an das Erreichen einer Akzeptanzrate von ca. 50% der MC-Übergänge hat sich als geeignet erwiesen [144].

Die Anpassung der Schrittweite erfolgt dynamisch während der MC-Simulation. Dazu wird nach jedem oder einer bestimmten Zahl von Makrozyklen die Akzeptanzrate neu berechnet. Ist die Akzeptanzrate $> 50\%$, wird die Schrittweite um 5% erhöht. Ist die Akzeptanzrate dagegen $< 50\%$, wird die Schrittweite um 5% verringert.

Das in Tab. 4.4 dargestellte molekulare Modell des MC-Algorithmus enthält 12 MC-Variablen je Nukleotid, eine zusätzliche je Thymin, 3 MC-Variablen je explizitem Gegenion und 10 MC-Variablen des Liganden Methylenblau. Die dynamische Schrittweitenanpassung muß natürlich für jede dieser MC-Variablen unabhängig voneinander erfolgen. Und außerdem werden gleiche MC-Variablen für die einzelnen Nukleotide bzw. Ionen individualisiert, da die Schrittweite von MC-Variablen eines individuellen Nukleotids bzw. Ions stark von der aktuellen Umgebung dieser zu bewegenden Einheit abhängig ist. Somit können die individuellen Schrittweiten eines molekularen Systems mit i monomeren Einheiten (Nukleotide, Ionen, Ligand) und j MC-Variablen in einer Matrix $\delta(i, j)$ dargestellt werden, deren Matrixelemente unabhängig voneinander der dynamischen Schrittweitenanpassung unterzogen werden.

Im Initialisierungsschritt des MC-Algorithmus werden die Maximal-, Minimal- und Startwerte der Schrittweiten der einzelnen MC-Übergänge vereinbart. Die MC-Simulation führt dann zur Individualisierung der aktuellen Schrittweiten infolge der dynamischen Anpassung. Die Startwerte δ_{Start} sind so gewählt, daß sich zu Beginn der MC-Simulation eine Akzeptanzrate von ca. 55% einstellt. Maximale Schrittweiten δ_{Max} werden festgelegt, um zu verhindern, daß das MC-Sampling in einem Bereich des Konformationsraumes, der große Schrittweiten zuläßt, die Schrittweite soweit hochgeregelt wird, daß in anderen Bereichen des Konformationsraumes keine MC-Übergänge akzeptiert werden. Beispielsweise sind die Ionen fernab der DNA weitgehend frei beweglich, was zu sehr großen Schrittweiten führen würde. Das MC-Sampling wäre aber für ein Ion, das mit großen Schrittweiten in die Nähe der DNA gelangte, stark eingeschränkt. Dagegen ist die Festlegung minimaler Schrittweiten δ_{Min} erforderlich, um zu gewährleisten, daß eine geringe Beweglichkeit in einem bestimmten Bereich des Konformationsraumes nicht durch die Generierung unendlich kleiner Schrittweiten ein weiteres MC-Sampling verhindert.

Die in 10^6 MC-Makrozyklen dynamisch angepaßten Schrittweiten δ der sich durch die Basen G, C, A und T unterscheidenden 4 Nukleotide sind in Tab. 4.5 angegeben: Die Translationen T_x entlang der in die große Furche zeigenden x -Achse erlauben größere Schrittweiten für die Purinbasen im Vergleich zu den Pyrimidinbasen, wohingegen sowohl die Translationen T_y als auch die Translationen T_z für alle Basen nahezu gleich sind. Die Translationen T_z in z -Richtung, also entlang der helikalen Achse, sind aber

4. Monte-Carlo-Algorithmus zur Simulation von Nukleinsäuren

MC-Variable	δ_{Start}	δ_{Max}	δ_{Min}	$\delta(G)$	$\delta(C)$	$\delta(A)$	$\delta(T)$
T_x [Å]	0,15	0,3	0,015	0,10-0,23	0,09-0,13	0,12-0,24	0,10-0,13
T_y [Å]	0,15	0,3	0,015	0,07-0,11	0,07-0,11	0,08-0,12	0,07-0,13
T_z [Å]	0,15	0,3	0,015	0,13-0,28	0,16-0,24	0,15-0,26	0,18-0,28
R_x [°]	1,5	3,0	0,15	1,3-2,3	1,2-2,5	1,2-2,0	1,2-2,2
R_y [°]	1,5	3,0	0,15	1,9-3,0	2,1-3,0	1,9-3,0	2,3-3,0
R_z [°]	1,5	3,0	0,15	0,5-0,7	0,5-0,7	0,5-0,7	0,4-0,7
P [°]	5,0	10,0	0,5	3,0-5,5	2,5-5,3	3,3-7,4	3,0-6,0
A [°]	2,5	5,0	0,25	1,4-1,8	1,4-4,3	1,7-2,2	1,4-4,4
χ [°]	2,5	5,0	0,25	1,8-2,7	1,8-3,6	1,8-2,7	1,8-4,1
ε [°]	5,0	10,0	0,5	4,0-6,4	4,3-6,7	4,0-6,0	4,0-5,5
γ [°]	5,0	10,0	0,5	9,3-10,0	9,4-10,0	8,2-10,0	9,0-10,0
ω_1 [°]	5,0	10,0	0,5	2,7-4,5	2,5-4,1	3,3-4,9	3,3-4,5
ϕ [°]	5,0	10,0	0,5	–	–	–	10,0
t_x, t_y, t_z [Å]	1,5	3,0	0,15	2,3-3,0		Explizite Gegenionen	
T_x^L [Å]	0,15	0,3	0,015	0,28-0,30		Ligand Methylenblau	
T_y^L [Å]	0,15	0,3	0,015	0,30			
T_z^L [Å]	0,15	0,3	0,015	0,15-0,28			
R_x^L [°]	1,5	3,0	0,15	2,5-3,0			
R_y^L [°]	1,5	3,0	0,15	3,0			
R_z^L [°]	1,5	3,0	0,15	3,0			
ψ_i [°]	5,0	10,0	0,5	10,0			

Tabelle 4.5.: *Start-, Maximal-, Minimal- und dynamisch angepaßte Schrittweiten der in Tab. 4.4 definierten MC-Variablen. Die dynamisch angepaßten Werte der Schrittweiten basieren auf 10^6 MC-Makrozyklen und sind für die Nukleotide, die expliziten Gegenionen und den Liganden angegeben. Die dynamisch angepaßten Schrittweiten der 4 Nukleotide G, C, A und T sind zum Vergleich einander gegenübergestellt.*

deutlich kleiner als die Translationen T_x und T_y in der Ebene der Basen. In den Rotationen um die drei Achsen unterscheiden sich die 4 Basen kaum. Die Schrittweiten für die Rotationen R_z der Basen um die helikale Achse sind kleiner als für die Rotationen R_x und R_y um die x - bzw. y -Achse. Die Phase P des Zuckers weist für Nukleotide, die die Basen A oder T enthalten, größere Schrittweiten auf als für Nukleotide, die die Basen G oder C enthalten. Dieses könnte auf die höhere Flexibilität von AT-Basenpaaren im Vergleich zu GC-Basenpaaren aufgrund der schwächeren Wasserstoffbrücken-Wechselwirkungen zurückzuführen sein. Die Amplituden A des Zuckers hingegen weisen wie auch die Glykosidwinkel χ größere Schrittweiten für Pyrimidinbasen enthaltende Nukleotide im Vergleich zu Nukleotiden auf, die Purinbasen enthalten. Die Schrittweiten der drei übrigen MC-Variablen ε , γ und ω_1 unterscheiden sich für die 4 Nukleotide nur geringfügig. Allerdings sind die Schrittweiten des Torsionswinkels γ stets größer als die Schrittweiten des Torsionswinkels ε . Die Rotationen ϕ der Methylgruppen an den Thyminen erreichen sehr schnell ihre maximale Schrittweite von 10° , was auf eine nahezu ungehinderte Rotation hinweist.

Die in Tab. 4.5 angegebenen, dynamisch angepaßten Schrittweiten δ sind durch MC-Simulationen von DNA-Dekameren generiert worden, deren Basensequenzen palindro-

me Symmetrien aufweisen. Dazu wurden MC-Simulationen von Dekameren mit alternierenden AT und GC Basensequenzen durchgeführt. Die große Beweglichkeit von Nukleotiden an den Rändern der DNA-Fragmente im Gegensatz zu Nukleotiden im Zentrum der Dekamere resultieren in größeren bzw. kleineren Werten für die dynamisch angepaßten Schrittweiten, weshalb in Tab. 4.5 Intervalle für δ angegeben sind. Lediglich die Rotationen ϕ der Methylgruppen am Thymin sind nicht durch ein Intervall für δ charakterisiert, da nicht das Akzeptanzkriterium des MC-Algorithmus sondern die zulässige maximale Schrittweite δ_{Max} die Beweglichkeit dieser MC-Variablen einschränkt.

Die expliziten Gegenionen sind mit dynamisch angepaßten Schrittweiten $\delta = 2,3-3,0 \text{ \AA}$ bei maximalen Schrittweiten von $\delta_{Max} = 3,0 \text{ \AA}$ sehr gut beweglich. Die expliziten Gegenionen tragen mit dieser hohen Beweglichkeit entscheidend zu einer schnellen Equilibrierung des Systems bei. Erst die Kombination einer impliziten Beschreibung des Lösungsmittels mit einer expliziten Beschreibung der Ionen ermöglicht diese hohe Beweglichkeit, die als einer der wesentlichen Vorzüge des verwendeten MC-Algorithmus angesehen werden kann.

Die dynamisch angepaßten Schrittweiten δ des Liganden Methylenblau weisen auf eine im Vergleich zu den Basen der DNA höhere Beweglichkeit des Liganden hin. Die Translationen T_x^L und T_y^L , die Rotationen R_y^L und R_z^L sowie die Rotationen der 4 Methylgruppen ψ_i sind durch die maximalen Schrittweiten δ_{Max} eingeschränkt. Diese gute Beweglichkeit in den Translationen und Rotationen im Vergleich zu den Basen beruht darauf, daß der Ligand nicht kovalent gebunden ist. Die Translationen T_z^L sind etwas eingeschränkt, da sterische Kollisionen bei Bewegungen des Liganden entlang der orthogonal zur Ebene des π -Elektronensystems orientierten z -Achse häufiger auftreten als bei Verschiebungen in der (x, y) -Ebene des planaren Liganden. Die Rotationen R_x^L sind ebenfalls etwas eingeschränkt, da bei Rotationen um die in Abb. 4.5 dargestellte x -Achse des MB sterische Kollisionen eher als bei Rotationen um die lange Achse des MB oder die zum planaren MB senkrechte Achse auftreten, die aufgrund nicht akzeptierter MC-Übergänge zu kleineren dynamisch angepaßten Schrittweiten führen. Die in Tab. 4.5 angegebenen dynamisch angepaßten Schrittweiten sind weitgehend unabhängig vom Bindungsmodus des MB mit der DNA. Die maximalen Schrittweiten für die Bewegungen des Liganden sind zunächst dennoch an den maximalen Schrittweiten für die Bewegungen der Basen orientiert, da die Nukleinsäurekonformationen auf die Bewegungen des Liganden reagieren.

4.2.4. Strukturelle Parameter

Die bei den MC-Simulationen generierten Trajektorien entsprechen statistischen Ensembles von Konfigurationen des molekularen Systems. Diese Konfigurationen sind für fortschreitende MC-Makrozyklen als Abfolge in den Trajektorien durch die kartesischen Koordinaten aller Atome des molekularen Systems protokolliert. Die statistischen Ensembles werden dahingehend ausgewertet, daß aus den kartesischen Koordinaten strukturelle Parameter berechnet werden, die die Interpretation der Resultate übersichtlicher und anschaulicher gestalten. Dazu werden strukturelle Parameter berechnet, die die Konformation der Nukleinsäuren, die Bewegung der Ionen oder den Bindungsmodus des Liganden charakterisieren.

Konformation der Nukleinsäuren

Die die Nukleinsäurekonformation beschreibenden strukturellen Parameter wurden in Kap. 2.2 eingeführt. Bei der Auswertung der statistischen Ensembles werden zunächst die in Abb. 2.3 dargestellten helikalen Parameter X-Displacement $dx(i)$, Y-Displacement $dy(i)$, Rise $Dz(i)$, Inclination $\eta(i)$, Tip $\vartheta(i)$ und Twist $\Omega(i)$ nach 2.1- 2.6 für die individuellen Basen i berechnet. In Analogie zum Curves-Algorithmus wird dazu die beste lineare Achse der Nukleinsäurekonformation berechnet und der in Kap. 2.2.1 erläuterten Definition der helikalen Parameter zugrunde gelegt [13, 14].

Neben den die Lage und Orientierung der Basen charakterisierenden helikalen Parametern wird die Konformation des Nukleinsäure-Rückgrats, wie in Kap. 2.2.3 eingeführt, durch die Phase $P(i)$ und Amplitude $A(i)$ des Zuckers, den Glykosidwinkel $\chi(i)$ sowie die in Abb. 2.4 dargestellten endozyklischen Torsionswinkel $\varepsilon(i)$, $\zeta(i)$, $\alpha(i)$, $\beta(i)$, $\gamma(i)$ und $\delta(i)$ der individuellen Nukleotide i beschrieben. Für diese strukturellen Parameter der Nukleinsäurekonfigurationen des statistischen Ensembles werden schließlich in Analogie zu 4.23 und 4.24 die Mittelwerte und Fluktuationen berechnet.

Bewegungen der Ionen

Die Bewegungen der expliziten Gegenionen sind bereits in Kap. 4.1.5 dahingehend diskutiert worden, daß sich die Ionen durch Translationen in die drei Raumrichtungen frei bewegen können, solange sie sich innerhalb eines Zylinders befinden, dessen Achse mit der helikalen Achse der Nukleinsäurestruktur zusammenfällt. Der Radius dieses Zylinders wird mit $R_{Zyl} = 100 \text{ \AA}$ gegenüber dem Radius $R \approx 10 \text{ \AA}$ einer B-DNA so groß gewählt, daß Wechselwirkungen von Ionen am Rand des Zylinders mit an der DNA kondensierten Ionen vernachlässigt werden können. Durch Reflexionen der Ionen an der Zylinderoberfläche verbleiben Ionen im Inneren des Zylinders, die durch MC-Übergänge den Zylinder verlassen würden.

Bei der Auswertung des statischen Ensembles wird hinsichtlich der expliziten Gegenionen im wesentlichen der Abstand der Ionen von der helikalen Achse der Nukleinsäurestruktur diskutiert. Dazu wird der Abstand der Ionen von der helikalen Achse in einer zu dieser Achse senkrechten Ebene berechnet, so daß eine Verteilung der Ionen über einer Distanz angegeben werden kann, die dem Radius R_{Zyl} des Zylinders entspricht, in dem sich die Ionen frei bewegen können. Die Kondensation der Gegenionen an der Oberfläche der DNA kann auf dieser Grundlage detailliert diskutiert werden.

Beschreibung des Liganden

Die Lage und Orientierung des Liganden wird in Analogie zu den Basen durch helikale Parameter charakterisiert. Die Berechnung helikaler Parameter des Liganden erfordert die Definition eines körperfesten Koordinatensystems $(\vec{x}, \vec{y}, \vec{z})$ des Liganden [53]. Der Ursprung \vec{O} dieses körperfesten Koordinatensystems des Liganden Methylenblau liegt im Mittelpunkt des Schwefelatoms. In diesem Referenzpunkt \vec{O} ist das körperfeste Koordinatensystem $(\vec{x}, \vec{y}, \vec{z})$ so definiert, daß die \vec{x} -Achse der kleinen Achse zwischen dem Schwefel- und Stickstoffatom im mittleren Ring des Methylenblau und die \vec{y} -Achse der großen Achse des Liganden entspricht. Die Lage dieser beiden Achsen ist in Abb. 4.5 dargestellt. Die \vec{z} -Achse steht orthogonal auf der Ebene des π -Elektronensystems des Methylenblau im Mittelpunkt des Schwefelatoms.

Die 4 helikalen Parameter des Liganden X-Displacement dx_L , Y-Displacement dy_L , Inclination η_L und Tip ϑ_L geben die Lage und Orientierung des Liganden in bezug auf die helikale Achse der Nukleinsäurestruktur an. Darüberhinaus charakterisieren die beiden helikalen Parameter Rise Dz_L und Twist Ω_L die Lage und Orientierung des Liganden relativ zu einem achsenbezogenen Koordinatensystem $(\vec{u}, \vec{v}, \vec{w})$ im Referenzpunkt \vec{P} eines zu wählenden Referenzbasenpaares. Als helikale Achse \vec{u} wird die in Analogie zum Curves-Algorithmus berechnete beste lineare Achse verwendet [13, 14]. Die \vec{v} - und \vec{w} -Achsen sind die zur helikalen Achse senkrechten Projektionen des achsenbezogenen Koordinatensystems des Referenzbasenpaares. Damit können die 6 helikalen Parameter des Liganden Methylenblau berechnet werden [53]:

$$dx_L = (\vec{O} - \vec{P}) \cdot (\vec{y} \times \vec{u})_n \quad \text{mit} \quad (\vec{y} \times \vec{u})_n = \frac{\vec{y} \times \vec{u}}{|\vec{y} \times \vec{u}|} \quad (4.25)$$

$$dy_L = (\vec{O} - \vec{P}) \cdot (\vec{u} \times (\vec{y} \times \vec{u})_n) \quad (4.26)$$

$$Dz_L = (\vec{O} - \vec{P}) \cdot \vec{u} \quad (4.27)$$

$$\eta_L = \arcsin(\vec{y} \cdot \vec{u}) \quad (4.28)$$

$$\vartheta_L = \begin{cases} \arccos((\vec{y} \times \vec{u})_n \cdot \vec{x}) & \text{wenn } (\vec{u} \cdot \vec{x}) \leq 0 \\ -\arccos((\vec{y} \times \vec{u})_n \cdot \vec{x}) & \text{wenn } (\vec{u} \cdot \vec{x}) > 0 \end{cases} \quad (4.29)$$

$$\Omega_L = \begin{cases} \arccos((\vec{y} \times \vec{u})_n \cdot \vec{v}) & \text{wenn } ((\vec{y} \times \vec{u})_n \cdot \vec{w}) \leq 0 \\ -\arccos((\vec{y} \times \vec{u})_n \cdot \vec{v}) & \text{wenn } ((\vec{y} \times \vec{u})_n \cdot \vec{w}) > 0 \end{cases} \quad (4.30)$$

Diese helikalen Parameter charakterisieren durch die Lage und Orientierung des Liganden relativ zur DNA den Bindungsmodus der DNA-Ligand-Bindung. Eine Bindung des Methylenblau durch Interkalation zwischen benachbarte Basenpaare oder eine Bindung in der kleinen bzw. großen Furche führt zu charakteristischen Werten der helikalen Parameter des Liganden [53]. Außerdem kann aus dem Rise Dz_L abgeleitet werden, mit welchen Basenpaaren der DNA der Ligand wechselwirkt. Die dynamischen Eigenschaften der unterschiedlichen Bindungsmodi können durch die Berechnung von Mittelwerten und Fluktuationen der helikalen Parameter in Analogie zu 4.23 bzw. 4.24 für in MC-Simulationen generierte statistische Ensembles diskutiert werden.

4.2.5. Erforderliche Rechenzeiten

Im Kap.5 wird gezeigt werden, daß für alle untersuchten Strukturen in der Größenordnung von DNA-Dekameren 10^6 MC-Makrozyklen ausreichend sind, um selbst bei extremen Startkonfigurationen nach $5 \cdot 10^5$ Makrozyklen eine gute Equilibrierung zu erreichen. Werden Energieminimum-Konformationen als Startkonfigurationen verwendet, stellt sich eine gute Equilibrierung bereits nach $2 \cdot 10^5$ Makrozyklen ein.

Die rechentechnisch aufwendige Energieberechnung wird dahingehend optimiert, daß nach MC-Übergängen nur diejenigen paarweisen Wechselwirkungen berechnet werden, an denen in dem jeweiligen MC-Übergang bewegte Atome beteiligt sind. Außerdem wird die Exponentialfunktion in der abstandsabhängigen Dielektrizitätsfunktion 3.20 gefittet. Damit werden auf AMD Athlon XP 1800+ Linux PCs (Taktfrequenz 1,533 GHz) für die MC-Simulation von Systemen in der Größenordnung von 20 Nukleotiden und damit ca. 650 Atomen Rechenzeiten von etwas mehr als 0,5 s je Makrozyklus und für die gesamte MC-Simulation mit 10^6 Makrozyklen nur 5-6 Tage benötigt.