



Master Forschungspraktikum

---

BK.MDS

# Molekulardynamik-Simulationen

---

Praktikanten: Philip Marszal

Betreuer:

Gruppe:

Durchgeführt:

E-Mail: `philip.marszal@stud.uni-goettingen.de`



# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
<b>2</b>	<b>Theorie</b>	<b>1</b>
2.1	Van-der-Waals Gas . . . . .	1
2.2	Proteine . . . . .	1
2.3	Molekulardynamik Simulation . . . . .	2
2.4	Principal Component Analysis . . . . .	2
<b>3</b>	<b>Durchführung</b>	<b>3</b>
3.1	Simulation von Argon . . . . .	3
3.2	Simulation von Protein G B1 . . . . .	3
3.3	Analyse der Bewegung eines Proteins . . . . .	4
<b>4</b>	<b>Auswertung</b>	<b>4</b>
4.1	Simulation von Argon . . . . .	4
4.1.1	Gasförmiges Argon . . . . .	4
4.1.2	Flüssiges Argon . . . . .	7
4.2	Simulation von Proteinen . . . . .	9
4.2.1	Protein G B1 . . . . .	9
4.2.2	PCA von T4-Phagen Lysozym . . . . .	10
<b>5</b>	<b>Diskussion</b>	<b>14</b>
5.1	Simulation von Argon . . . . .	14
5.2	Simulation des Proteins G B1 . . . . .	17
5.3	Analyse der Bewegung des Lysozyms . . . . .	17

# 1 Einleitung

## 2 Theorie

### 2.1 Van-der-Waals Gas

Beim van-der-Waals Gas handelt es sich um eine Erweiterung des idealen Gases. Im Gegensatz zum idealen Gas wird angenommen, dass die Teilchen des Gases eine endliche Ausdehnung haben und eine zwischen den Teilchen wirkende anziehende Kraft wirkt. Aus diesen Annahmen lässt sich eine Modifikation der idealen Zustandsgleichung, die van-der-Waals-Gleichung 1, herleiten [3]:

$$\left(p + \frac{a \cdot N^2}{V^2}\right) \left(\frac{V}{N} - b\right) = k_B T \quad (1)$$

Hierbei bezeichnet  $p$  den Druck,  $N$  die Teilchenzahl,  $V$  das Volumen,  $k_B$  die Boltzmann-Konstante und  $T$  die Temperatur. Die beiden Parameter  $a$  und  $b$  sind stoffabhängig und können experimentell bestimmt werden.

### 2.2 Proteine

Proteine gehören zu den Grundbausteinen von lebenden Organismen. Sie bestehen aus Ketten von Aminosäuren. Der Aufbau von Proteinen lässt sich in drei bis vier unterschiedliche Kategorien aufteilen. Die grundlegendste Struktur von Proteinen, die Primärstruktur, bezeichnet die Abfolge von Aminosäuren im Protein. In der Biologie kommen 20 verschiedene Aminosäuren vor. Die Reihenfolge der Aminosäuren bestimmt durch Anziehungskräfte der Aminorestgruppen zu einer charakteristischen Faltung des Proteins. Dabei bestimmen hauptsächlich Wasserstoffbrückenbindungen die Struktur, allerdings gibt es Aminosäuren wie Cystein, das kovalente Bindungen mit anderen Cystein-Molekülen eingehen kann [8].

Die Faltung des Proteins führt auf der nächsten Ebene zur Ausbildung der Sekundärstruktur. Die zwei am häufigsten anzutreffenden Sekundärstrukturen sind die  $\alpha$ -Helix und die  $\beta$ -Faltblatt Struktur.

Eine weitere Ebene darüber ist die Tertiärstruktur. Sie beschreibt die Anordnung der Sekundärstrukturen im dreidimensionalen Raum. Für Proteine, die nur aus einer Aminosäurekette bestehen, beschreibt sie die endgültige Form des Proteins[8]. Viele Proteine jedoch bestehen aus mehreren Aminosäureketten. Die Anordnung der Aminosäureketten zusammen wird durch die Quartärstruktur beschrieben.

## 2.3 Molekuldynamik Simulation

Die numerische Behandlung von Molekülen basiert zunächst auf der quantenmechanischen Beschreibung von Molekülen. Allerdings stößt die Berechnung von quantenmechanischen Systemen bereits ab einer Zahl von 10 Atomen an ihre Grenzen. Um dennoch sinnvolle numerische Ergebnisse in endlicher Zeit zu erhalten, werden für die Molekuldynamik Simulationen eine Reihe von Approximationen eingesetzt. So wird die Bewegung der Elektronen eines Atoms, von der Bewegung des Kerns, entkoppelt, indem angenommen wird, dass der Atomkern sich im Vergleich zu den Elektronen sehr langsam bewegt. Damit kann die Bewegung der Elektronen als die Bewegung in einem ruhenden Potential der Kerne betrachtet werden. Der Einfluss der Elektronen auf den Atomkern wird wiederum durch ein durch die Elektronen verursachtes effektives Potential beschrieben.

Dieses effektive Potential wird durch die Koordinaten des Atomkerns bestimmt. Es berücksichtigt sowohl Bindungen in einem Molekül als auch langreichweitige Wechselwirkungen wie die van-der-Waals-Kräfte.

Die Bewegung der Atomkerne kann darauf hin nach den klassischen Bewegungsgleichungen integriert werden. Hier wird der *leap frog* Algorithmus zur Integration verwendet, wie er ausführlich in [1] behandelt wird:

$$v\left(t + \frac{\Delta t}{2}\right) = v\left(t - \frac{\Delta t}{2}\right) + \frac{F(t)}{m}\Delta t, \quad (2)$$

$$r\left(t + \frac{\Delta t}{2}\right) = r(t) + v\left(t + \frac{\Delta t}{2}\right)\Delta t. \quad (3)$$

Des weiteren müssen für eine sinnvolle Simulation einige technische Grenzen gesetzt werden. Die Integrationsschrittweite wird den schnellsten erwarteten Fluktuationen entsprechend gewählt. In diesem Versuch sind dies Fluktuationen von Wasserstoffatomen, die auf einer Zeitskala von  $10^{-14}$  s geschehen. Als Schrittweite wird deswegen  $\Delta t = 10^{-15}$  s gewählt. Um Randeffekte zu vermeiden werden periodische Randbedingungen gewählt, wodurch prinzipiell ein unendlich großes System simuliert wird.

Die Molekuldynamik Simulation wird in diesem Versuch mit dem Softwarepaket *GROMACS* [4] durchgeführt.

## 2.4 Principal Component Analysis

Die Hauptkomponentenanalyse oder PCA (engl. Principal Component Analysis) ist eine Methode der Datenanalyse, die aus einem Datensatz Linearkombinationen von Variablen erzeugt die maximalen Einfluss auf die Dynamik haben. Dazu wird

zunächst die Kovarianz-Matrix des Datensatzes bestimmt. Anschließend wird diese diagonalisiert und die Diagonalelemente ihrer Größe nach geordnet. Nun werden die Daten in die Basis mit diagonalen Kovarianz-Matrix transformiert. Die Basisvektoren dieser Basis entsprechen den Hauptkomponenten oder PCs (engl. Principal Components). Das zugehörige Diagonalelement der Kovarianz-Matrix entspricht der auf dieser Achse auftretenden Varianz der Daten.

Die Hauptkomponenten mit der größten Varianz entsprechen den Variablen mit der stärksten Aussagekraft über den Zustand des Systems. Vernachlässigt man Hauptkomponenten geringer Varianz, so lässt sich die Dimensionalität des Datensatzes stark verringern ohne Information einzubüßen [1].

In diesem Versuch wird es hilfreich sein komplexe Proteinbewegungen auf zwei essentielle Bewegungen zu reduzieren um die Arbeitsweise des Proteins zu untersuchen.

## 3 Durchführung

### 3.1 Simulation von Argon

Um sich mit dem Simulationsprogramm vertraut zu machen und eventuelle Schwächen der Simulation aufzuzeigen, wird zunächst ein simples Argon-Gas simuliert. Untersucht wird das Verhalten von Argon beim Abkühlen und Erhitzen. Dazu wird das Gas von 100 K auf 25 K abgekühlt. Dies wird für Zeiträume von 10 ns, 5 ns und 1 ns durchgeführt. Die Simulation liefert Details zur Energieverteilung und Temperatur des Argons. Die Temperatur des Gases wird mittels des *Berendsen*-Thermostats [?] erniedrigt.

Das Sieden von Argon wird untersucht, indem das Ergebnis der Abkühlung aus dem ersten Schritt, von 25 K auf 100 K erwärmt wird. Dies findet über einen Zeitraum von 10 ns statt.

Als nächstes wird die Anzahl der Argon Atome auf 216 erhöht und das Argon von 100 K auf 0 K innerhalb von 1 ns abgekühlt.

### 3.2 Simulation von Protein G B1

Zur Simulation des Proteins G B1 wird eine Struktur aus der *protein data bank* verwendet. Um die heruntergeladene Struktur verwenden zu können, muss die Topologie-Datei mithilfe von *GROMACS* erstellt werden. Das Protein wird nun in einem Simulationskasten mit Wassermolekülen umgeben. Die Kanten der Box haben dabei einen Mindestabstand von 7 Å von der Oberfläche des Proteins. Auch hierfür wird eine Funktion von *GROMACS* verwendet. Da das Protein G

B1 4-fach negativ geladen ist, werden vier Wassermoleküle durch positiv geladene Natrium Ionen ersetzt.

Da es beim Laden der Struktur aus der .pdb Datei und dem Hinzufügen von Wasser zum Überlappen von Atomen kommen kann, muss nun eine Energieminimierung durchgeführt werden. Dabei werden die Koordinaten der Atome so variiert, dass sie in einem Minimum der potentiellen Energie zur Ruhe kommen. Der Unterschied zur herkömmlichen Molekulardynamik Simulation liegt hier darin, dass es sich um ein Minimum der potentiellen Energie handelt und nicht der freien Energie.

Nach der Minimierung der Energie, wird das Protein über längere Zeit simuliert. Hierbei wird das Kraftfeld *AMBER99SB protein, nucleic AMBER94* aus [6] verwendet.

### 3.3 Analyse der Bewegung eines Proteins

Die Bewegung von T4-Phage Lysozym wurde bereits im Voraus simuliert, da eine detaillierte Simulation zu lange für diesen Versuch dauert. Die Bewegung der einzelnen Atome wird mit PCA auf Hauptbewegungen reduziert, und anschließend mit experimentellen Daten verglichen.

## 4 Auswertung

### 4.1 Simulation von Argon

#### 4.1.1 Gasförmiges Argon

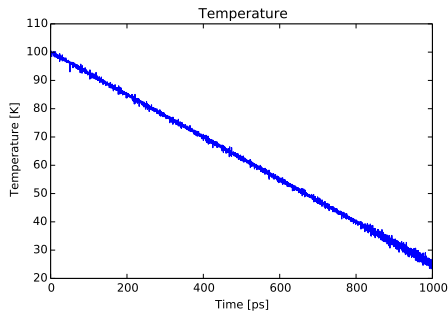
Der Temperaturverlauf des Argons für verschiedene Zeiträume der Abkühlung ist in Abb. 1a) bis 1c) dargestellt. Die dazugehörige potentielle Energie ist in Abb. 2a) bis 2c) zu sehen. Zusätzlich ist in den Abbildungen 1c) und 2c) der Vergleich zwischen Erwärmen und Abkühlen des Argon Gases zu sehen.

Aus dem Verlauf der potentiellen Energie lässt sich der Siedepunkt des Argongases bestimmen. Dazu wird eine Logistische Kurve,

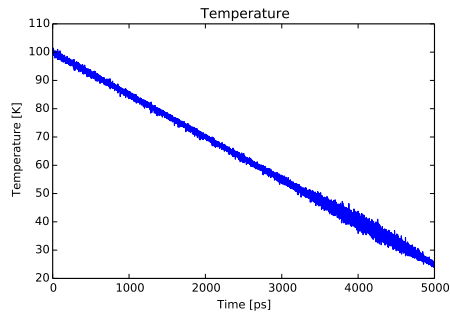
$$f(x) = \frac{L}{1 + \exp[-k(x - x_0)]} + y_0 \quad (4)$$

an die Messwerte gefittet. Zum Fitten wird eine *python-scipy* Funktion benutzt, die Gebrauch von dem *Levenberg-Marquardt* Algorithmus macht [7].

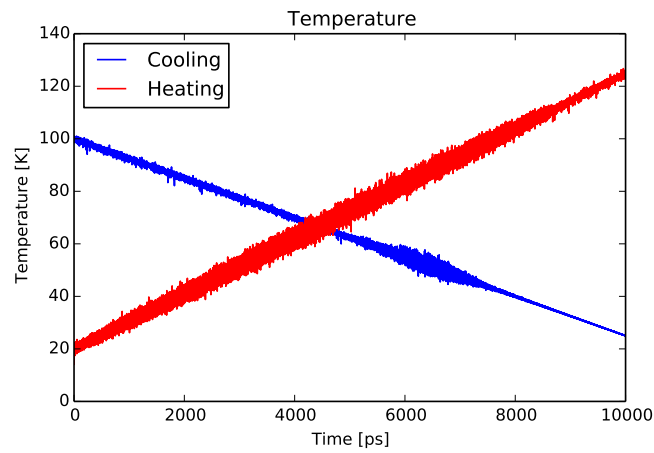
Das Maximum der Ableitung der Kurve entspricht dem Siedepunkt. Es liegt bei  $x = x_0$ . Der Siedepunkt konnte aus den drei verschiedenen Abkühlvorgängen bestimmt werden, die Ergebnisse sind in Tabelle 1) aufgeführt.



(a) Temperaturverlauf für das Abkühlen in 1 ns.



(b) Temperaturverlauf für das Abkühlen in 5 ns.



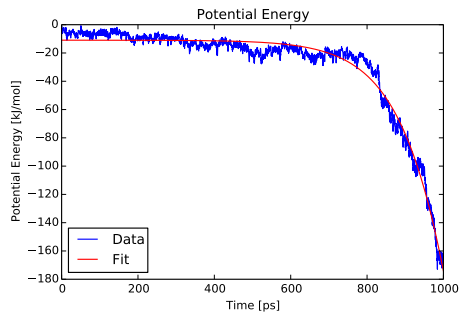
(c) Temperaturverlauf für das Abkühlen und Erwärmen in 10 ns.

Abbildung 1: Temperaturverläufe der Simulationen von gasförmigen Argon beim Abkühlen bzw. Erwärmen.

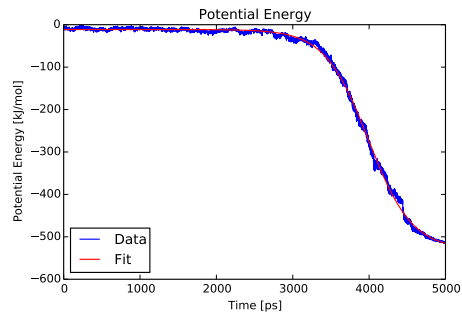
Tabelle 1: Siedepunkte wie sie aus den verschiedenen Simulationen bestimmt wurden. Zur Bestimmung der Temperatur wurde ein linearer Zusammenhang zwischen Temperatur und Zeit angenommen der aus dem Versuchsaufbau und Abb. 1a) bis 1c) hervorgeht.

Abkühlzeit [ns]	$x_0$ [ps]	Temperatur [K]
1	1150(20)	14(2)
5	3990(1)	40.15(1)
10	6239.6(2)	53.203(2)

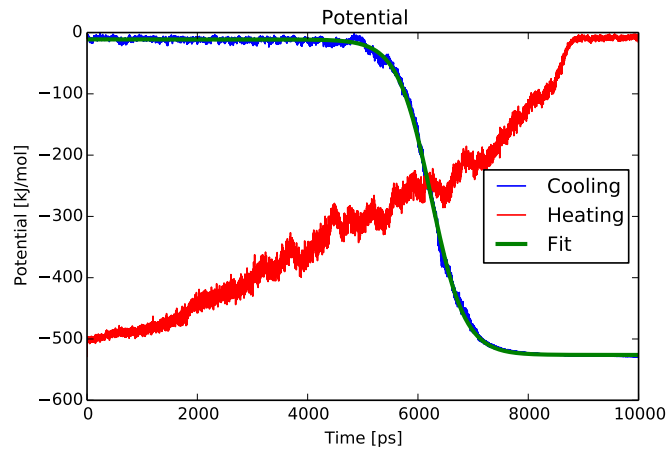




(a) Verlauf der potentiellen Energie für das Abkühlen in 1 ns.

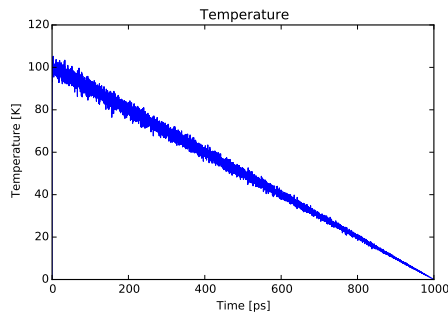


(b) Verlauf der potentiellen Energie für das Abkühlen in 5 ns.

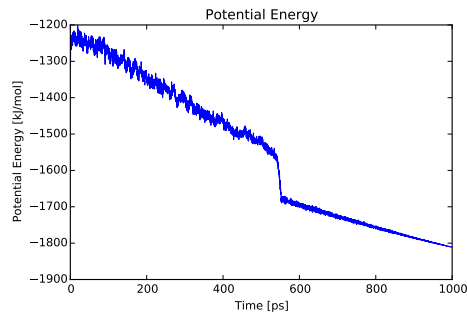


(c) Verlauf der potentiellen Energie für das Abkühlen und Erwärmen in 10 ns.

Abbildung 2: Potentielle Energie der Simulationen von gasförmigen Argon beim Abkühlen bzw. Erwärmen.



(a) Verlauf der Temperatur beim Abkühlen von flüssigem Argon von 100 K auf 25 K, bei einem Druck von 5.8 bar.



(b) Verlauf der potentiellen Energie beim Abkühlen von flüssigem Argon von 100 K auf 25 K, bei einem Druck von 5.8 bar.

Abbildung 3: Gefrieren von Argon.

Der Siedepunkt von Argon ist stark vom Druck abhängig. Der Druck in dem simulierten System lässt sich unter Verwendung der van-der-Waals Gleichung bestimmen. Die Simulationsbox hat eine Kantenlänge von 7.7395 nm mit 100 Argon Atomen in der Box. Dies ergibt mit Werten für  $a = 1.355 \text{ L}^2\text{bar/mol}^2$  und  $b = 0.03201 \text{ L/mol}$  [2] einen Druck von ca. 2.8 bar bei 100 K. Bei diesem Druck hat Argon einen Siedepunkt von ca. 98 K [5]. Bei einem atmosphärischem Druck von 1 bar liegt der Siedepunkt bei 87 K.

#### 4.1.2 Flüssiges Argon

Als nächstes wird das Gefrieren von flüssigem Argon betrachtet. Der Druck des Gases bei 100 K entspricht etwa 5.8 bar, der Siedepunkt bei diesem Druck liegt bei ca. 107 K [5]. In Abb. 3a) und 3b ist der Verlauf der Temperatur und der potentiellen Energie zu sehen. Das Gefrieren des Argons setzt bei einer Temperatur von ca. 45 K ein. Es lässt sich außerdem die Anzahl der Atome in einem bestimmten Abstand eines Atoms bestimmen. Dies ist in Abb. 4) dargestellt. Die Verteilung ist zu Beginn in der flüssigen Phase und am Ende der Simulation in der festen Phase bestimmt worden. Zudem lässt sich die Diffusionskonstante des Argons bestimmen. Dazu wird die Einstein'sche Diffusionsrelation auf die durchschnittliche quadratische Abweichung vom Startpunkt bezogen. *GROMACS* liefert hierfür ein Ergebnis von  $2.5(1) \cdot 10^{-5} \text{ cm}^2/\text{s}$ , der experimentelle Wert liegt bei  $2.43 \cdot 10^{-5} \text{ cm}^2/\text{s}$ .

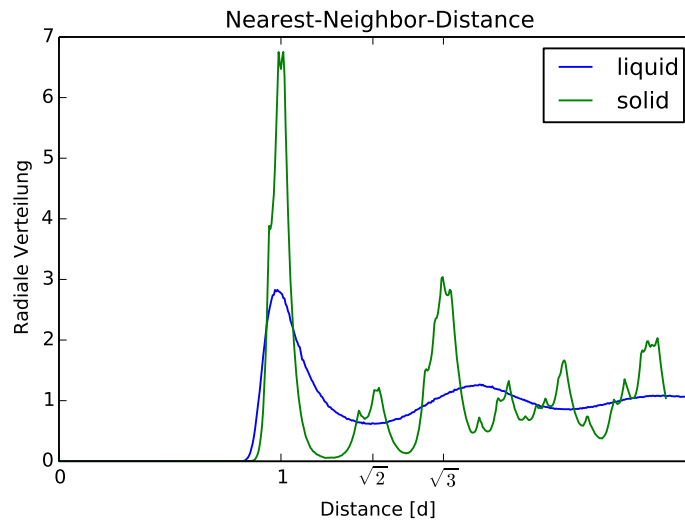


Abbildung 4: Radiale Verteilung der Argonatome als Funktion des Abstandes von einem Atom. Der Abstand ist in Einheiten des Durchmessers eines Argonatoms dargestellt ( $d = 0.376$  nm). Die Verteilung wurde zu Beginn der Simulation und nach Ende der Simulation als Mittel über ein Zeitintervall festgestellt. Das Auftreten der Peaks im Festkörper entspricht dem theoretisch erwarteten *fcc*-Gitter eines Argon-Kristalls.



Abbildung 5: Dreidimensionale Darstellung der B1 Domäne des Proteins G. Gelb:  $\beta$ -Faltblatt; Violett:  $\alpha$ -Helix.

## 4.2 Simulation von Proteinen

### 4.2.1 Protein G B1

Eine mit *VMD* erzeugte 3d-Darstellung der B1 Domäne des Proteins G ist in Abb. 5) zu sehen. Die gelben Pfeile stellen  $\beta$ -Faltblatt Strukturen da, und die violette Spirale eine  $\alpha$ -Helix Struktur. In grau sind Verbindungen zwischen den Sekundärstrukturen dargestellt. Wie stark die durchschnittlichen Positionen der Atome des Proteins von ihrer Anfangsposition abweichen, ist in Abb. 6a) als Funktion der Zeit aufgetragen. Der starke Anstieg zu Beginn der Simulation wird durch Minimierung der freien Energie verursacht. Im Gegensatz zur Minimierung der Energie wird nun nicht mehr nur die potentielle Energie minimiert. Das Verhalten der potentiellen Energie wird in Abb. ??) veranschaulicht. Die Fluktuationen der einzelnen Aminosäuren ist in Abb. 6b dargestellt. Zu erkennen sind deutliche Fluktuationen an den Residuen 1, 11, 14, 21, 24 und 42.

Die mit dem Lösungsmittel in Kontakt stehende Oberfläche des Proteins, lässt sich Abb. 6d) entnehmen. Der hohe Anteil an hydrophober Oberfläche hängt mit der Funktion des Proteins als Membranprotein zusammen. Die Tatsache, dass so-

wohl die hydrophile als auch die hydrophobe Oberfläche nur schwach fluktuieren bedeutet, dass sich das Protein nicht stark faltet.

Die durchschnittliche Ausdehnung des Proteins, der Gyrationradius ist in Abb. 6e) dargestellt.

Das Verhalten der Sekundärstrukturen des Proteins G B1 wird in Abb. 6f) als Funktion der Zeit farbkodiert aufgetragen. Man kann erkennen, dass die Sekundärstrukturen größtenteils stabil sind und nur an ihren Grenzen schwach fluktuieren.

#### 4.2.2 PCA von T4-Phagen Lysozym

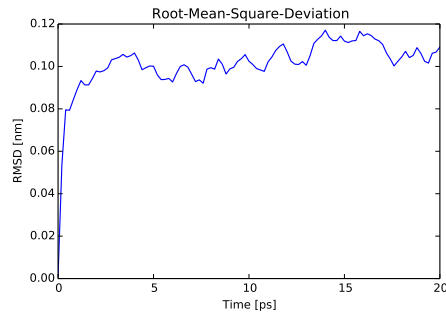
Die PCA der vorbereiteten Bewegung von T4-Phagen Lysozym liefert Hauptkomponenten mit den in Abb. 8) aufgetragenen Varianzen. Obwohl es mehr als 480 Atome im Lysozym gibt und damit mehr als 1400 Hauptkomponenten, wird ersichtlich, dass die Bewegung nur in wenigen linear unabhängigen Richtungen stattfindet. Hauptkomponenten mit einer dazugehörigen Varianz nah bei null sind auf die starke Bindung zwischen Atomen einer Aminosäure zurückzuführen und auf Symmetrien, wie die Translationsinvarianz des Systems. Die kumulative Varianz der ersten beiden Hauptkomponenten umfasst bereits mehr als 90% der gesamten Varianz der Atompositionen.

In Abb. 9) sieht man die Bewegung des Lysozyms entlang der ersten acht Hauptkomponenten. Aus den Projektionen kann der Einfluss der Hauptkomponenten auf die gesamte Bewegung abgeschätzt werden. Es ist allerdings einfacher nur die Hauptkomponenten zur Visualisierung und Analyse zu verwenden, die den größten Beitrag zur Varianz der Atompositionen liefern. Da es zum Visualisieren der Bewegung simpler ist nur zwei Dimensionen zu betrachten werden nur die ersten beiden Hauptkomponenten zur weiteren Analyse herangezogen.

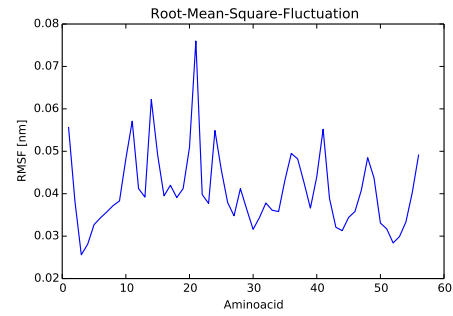
In Abb. 7) sind die Bewegungen entlang der ersten beiden Hauptkomponenten in der 3d-Darstellung des Lysozyms skizziert (rot: erste Hauptkomponente; blau: zweite Hauptkomponente).

Einen besonders guten Einblick in die Dynamik des Proteins liefert die Auftragung der ersten Hauptkomponente gegen die zweite. Dadurch wird die Energielandschaft in der sich das Protein bewegt angedeutet. In Abb. 10) wird die Bewegung des simulierten Lysozyms auf die Ebene, die von den ersten beiden Hauptkomponenten aufgespannt wird, projiziert.

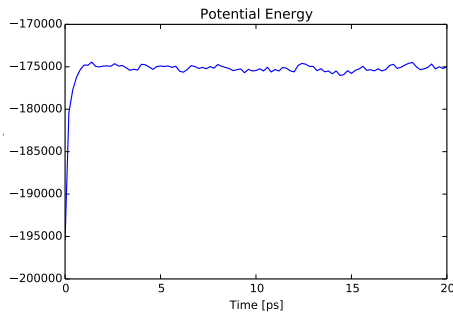
Da die PCA nicht von der zeitlichen Reihenfolge der Atompositionen abhängt, kann die Simulation des Lysozyms mit experimentell bestimmten Werten verglichen werden. Dazu wird für eine Reihe experimenteller Strukturanalysen des Lysozyms eine PCA durchgeführt. In Abb. 11) sieht man drei Simulationen mit unterschiedlichen Anfangsstrukturen auf die ersten beiden Hauptkomponenten



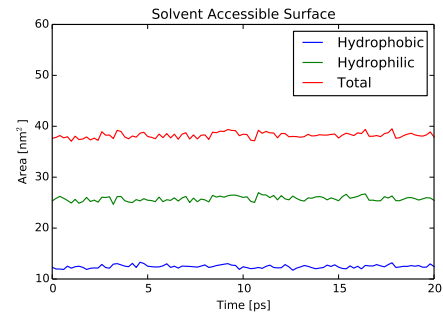
(a) Durchschnittliche Abweichung der Atompositionen als Funktion der Zeit.



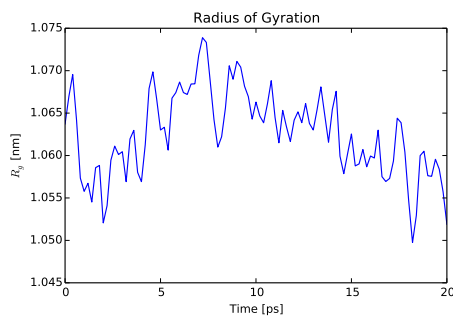
(b) Durchschnittliche Fluktuationen der einzelnen Aminosäuren.



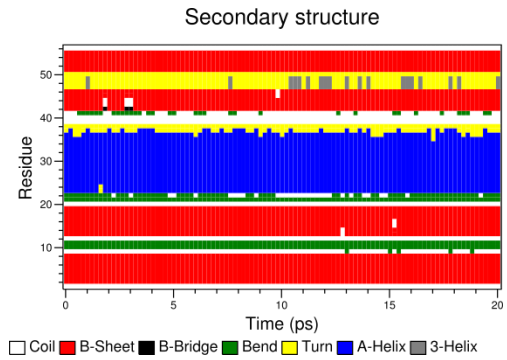
(c) Anstieg der potentiellen Energie (kJ/mol) als Resultat der Minimierung der freien Energie.



(d) Mit der Lösung in Kontakt stehende Fläche des Proteins.



(e) Gyrationradius des Proteins in Abhängigkeit von der Zeit.



□ Coil ■ B-Sheet ■ B-Bridge ■ Bend ■ Turn ■ A-Helix ■ 3-Helix

(f) Verhalten der Sekundärstrukturen als Funktion der Zeit.

Abbildung 6: Simulation des Proteins G B1 über einen Zeitraum von 20ps.

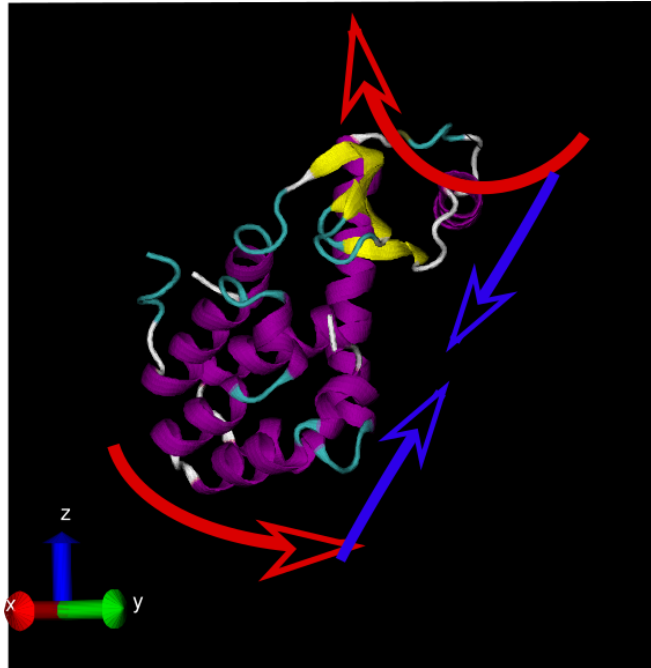


Abbildung 7: Skizze der Scher- und Klappbewegung des Lysozyms. Die erste Hauptkomponente entspricht der Scherbewegung (rot), die zweite einer Klappbewegung (blau).

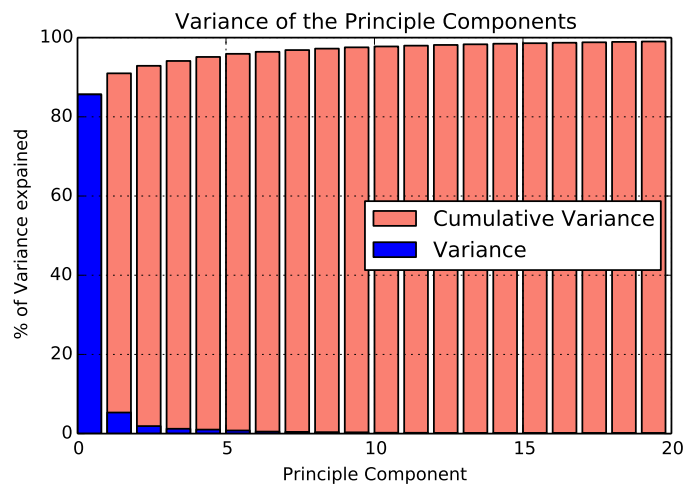


Abbildung 8: Varianz und kumulative Varianz der ersten 20 Hauptkomponenten. Schon nach der fünften Komponente ist kein merkbarer Zuwachs mehr zu erkennen. Die ersten beiden Hauptkomponenten entsprechen einer Varianz von 90.98%.

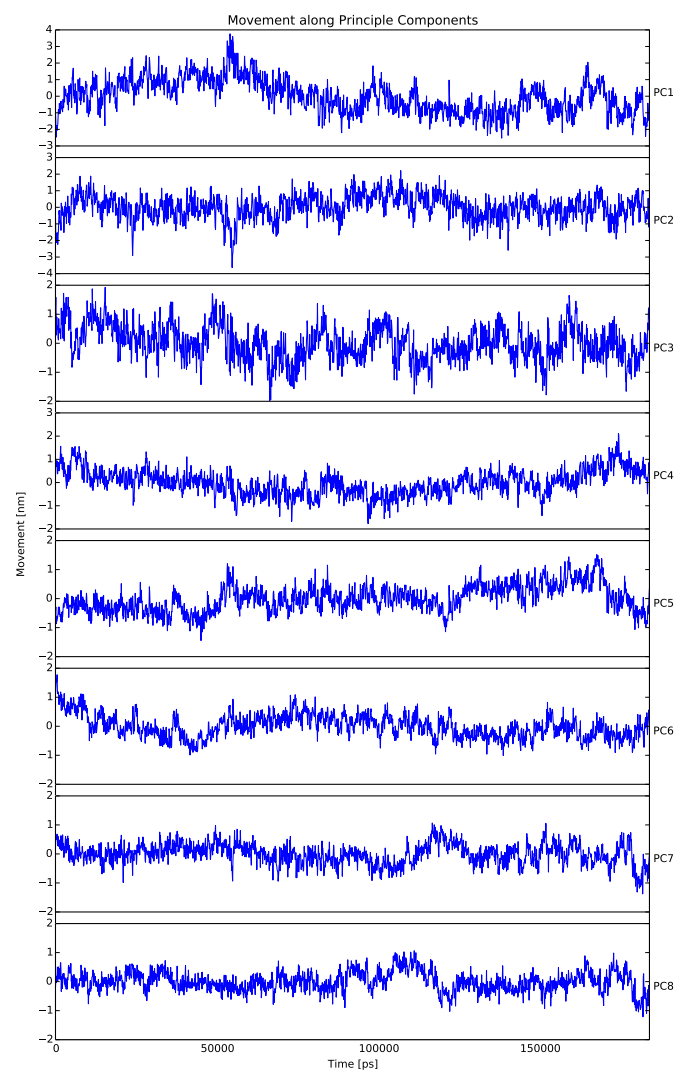


Abbildung 9: Projektion der Bewegung auf die ersten acht Hauptkomponenten in Abhängigkeit von der Zeit.



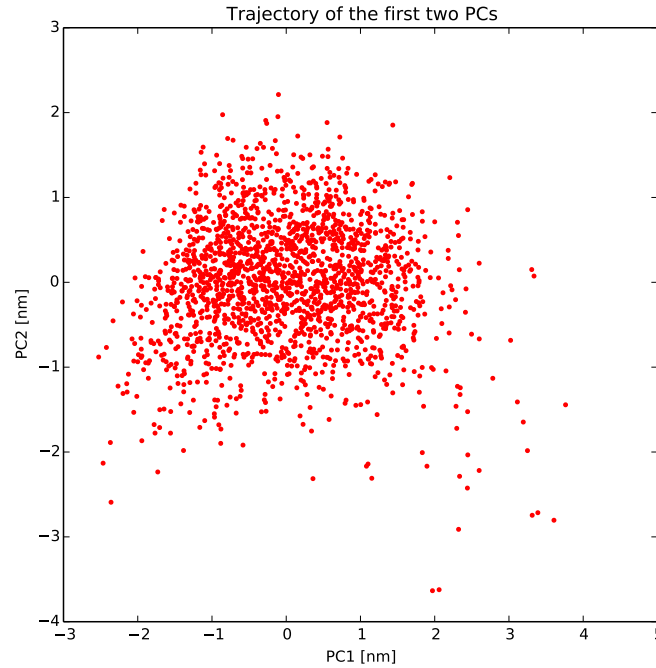


Abbildung 10: Projektion der Bewegung des T4-Phagen Lysozyms auf die ersten beiden Hauptkomponenten. Es wird ein zentraler Cluster der Bewegung deutlich.

der experimentellen Daten projiziert. Deutlich wird, dass Simulation 1 und 2 sich nur in lokalen Minima bewegen, während Simulation 3 die Energiebarriere der beiden Cluster überwindet und sich in beiden Minima bewegt. Die experimentellen Daten lassen diese lokalen Minima nur erahnen machen aber deutlich, dass die ersten beiden Simulationen nicht genügend Informationen zur Beschreibung der Dynamik des Proteins liefern.

## 5 Diskussion

### 5.1 Simulation von Argon

Betrachtet man den Verlauf der potentiellen Energie des Argon so stellt man für jede der Simulationen einen abrupten Abfall fest. Dies ist durch den Phasenübergang gasförmig-flüssig zu erklären. Die potentielle Energie des Gases ist durch die Summe der einzelnen effektiven Potentiale bestimmt. Beim Kondensieren, verrin-

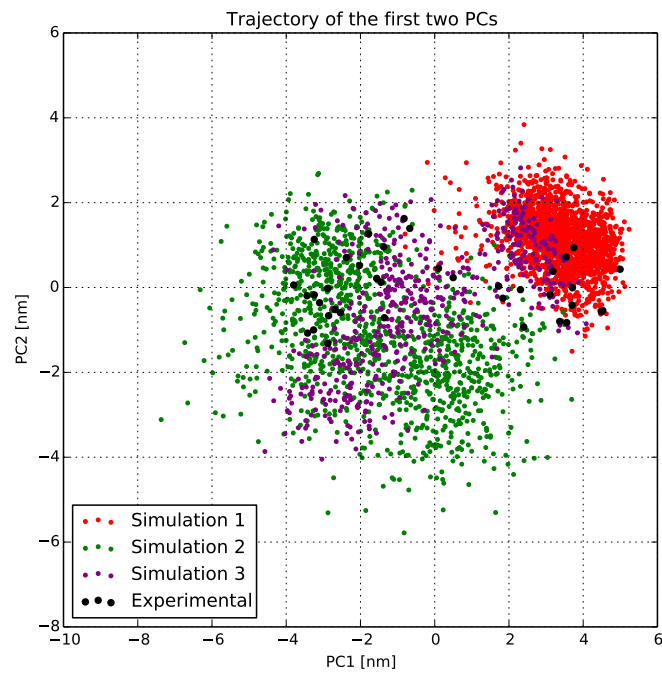


Abbildung 11: Vergleich der aufgezeichneten Bewegung des Lysozyms mit experimentellen Daten. Die Hauptkomponenten sind Resultate der PCA der experimentellen Daten. Es wird deutlich, dass Simulation 1 und 2 nicht den gesamten Raum der möglichen Bewegungen aufzeigen.

gern sich die Abstände zwischen den Atomen stark. Die einzelnen Argonteilchen sitzen nahezu im Minimum des Potentials fest. Deswegen ist es möglich gewesen den Siedepunkt des Argons mithilfe der potentiellen Energie zu bestimmen.

Die in Tab. 1 ermittelten Siedepunkte weichen stark von dem Literaturwert von 98 K [5] ab. Auch zueinander sind die ermittelten Werte stark unterschiedlich. Da der einzige Unterschied zwischen den Simulationen der Zeitraum der Abkühlung ist, liegt es nahe, dass dieser die entscheidende Rolle für den Siedepunkt spielt.

Besonders deutlich ist die Abweichung der 1 ns Simulation. Auffällig ist, dass das Minimum der potentiellen Energie über diesen Zeitraum nicht erreicht wird. Sogar der mittels dem Fit ermittelte Siedepunkt wird nicht erreicht. Hier wird deutlich, dass der Algorithmus zur Temperatursenkung nicht für sehr kleine Zeitskalen anwendbar ist. Unter Umständen jedoch kann das Ergebnis für kleine Zeitskalen verbessert werden, wenn kleinere Integrationsschritte verwendet werden, um den Einfluss der Temperatursenkung auf das Resultat der Integration zu minimieren. Ein gravierender Unterschied ist zwischen dem Erhitzen und Abkühlen festzustellen. In Abb. 2c) ist zu erkennen, dass die Kurve der potentiellen Energie beim Erhitzen des flüssigen Argons ungenauer ist und keiner logistischen oder ähnlichen Kurve entspricht. Die Bestimmung des Siedepunktes für diese Kurve war leider nicht mit einem Fit an eine Logistische Kurve möglich, weswegen nur eine Schätzung möglich ist. Der Punkt der Maximalen Steigung ist bei ca. 8300 ps zu erkennen was einer Temperatur von ca. 105 K entspricht. Obwohl dieser Wert näher an dem Literaturwert liegt als die zuvor bestimmten Werte, ist dennoch anzunehmen, dass aufgrund der Kurvenform das Ergebnis wesentlich ungenauer ist.

Der Schmelzpunkt von Argon kann erneut durch den rapiden Abfall der potentiellen Energie in Abb. 3b) bestimmt werden und liegt bei ca. 50 K. Der Literaturwert für den Schmelzpunkt von Argon bei 1 bar liegt bei 83.9 K [2]. Da Argon keine Anomalie besitzt, wie z.B. Wasser, ist für einen höheren Druck von 5.8 bar auch ein höherer Schmelzpunkt zu erwarten. Das Ergebnis widerspricht also den theoretischen und experimentellen Ergebnissen. Auch hier ist möglicherweise die rapide Abkühlung und systematische Fehler der verwendeten Algorithmen die Ursache für die starke Abweichung.

Ein Phänomen, welches deutlich zu erkennen ist, ist die Kristallisation von Argon. Dies wird durch Abb. 4) deutlich. Im flüssigen Zustand ist die Radiale Verteilung der Atome unscharf. Einzelne Argon Atome können sich noch um das Minimum des effektiven Potentials bewegen. Nach dem Erstarren jedoch, befinden sich die meisten Argon Atome im exakten Minimum des Potentials. In Abb. 4) wird das dadurch sichtbar, dass die Peaks der Kurve viel schmäler werden. Dass sich die Atome in einem Gitter anordnen wird dadurch deutlich, dass das zweite Maximum der Kurve ca. beim  $\sqrt{2}$ -fachen Abstand des ersten Maximums befindet. Die

Atome im zweiten Maximum sind also nächste Nachbarn von Atomen des ersten Maximums.

## 5.2 Simulation des Proteins G B1

Die Simulation der B1 Domäne des Protein G zeigt, dass sich die Bewegung auf eine Handvoll von Aminosäuren beschränkt. Die Form des Proteins bleibt zum Großteil unverändert. Da das benutzte Kraftfeld nicht dem aus der Versuchsanleitung entspricht, ist auch die Abweichung der Residuen mit der größten Fluktuation aus Abb. 6b unterschiedlich zu den in der Anleitung erwähnten Residuen 1, 11, 21, 38. Der zu erwartende Anstieg der Root-Mean-Square-Deviation der Atome, ist in Abb. 6a) deutlich zu erkennen und spiegelt auch den Anstieg der potentiellen Energie in Abb. ??) wieder.

Einen besonders guten Einblick in die Rolle und das Vorkommen des Proteins lässt die Verteilung der hydrophoben und hydrophilen Gruppen auf der Oberfläche des Proteins zu. Da der Größte Teil der Oberfläche hydrophob ist, kann man darauf schließen, dass das Protein hauptsächlich in einer ebenfalls hydrophoben Umgebung anzutreffen ist. Dies ist wahrscheinlich die Membran des Bakteriums. Unklar ist jedoch ob sich das Protein tatsächlich im Minimum der freien Energie befindet, da Abb. 6a) möglicherweise nicht um einen konstanten Wert schwankt. Es ist über den gesamten Zeitraum der Simulation ein Anstieg zu verzeichnen. Um eine genaue Aussage über die natürliche Bewegung dieses Proteins treffen zu können müsste die Simulation über einen längeren Zeitraum stattfinden.

## 5.3 Analyse der Bewegung des Lysozyms

Betrachtet man die Animation der Bewegung des Lysozyms entlang der ersten beiden Hauptkomponenten so erkennt man eine Scherbewegung und eine Klappbewegung. Da diese Bewegungen hauptsächlich um das aktive Zentrum des Proteins stattfinden, liegt die Vermutung nahe, dass sie grundlegend für die Zersetzung des Substrates sind.

Abb. 11) macht deutlich wie wichtig Anfangsbedingungen und Simulationsdauer für die vollständige Beschreibung der Bewegung eines Proteins sind. Bereits zwei experimentelle Strukturen liefern nur etwa die Hälfte des möglichen Bewegungsraumes bei einer Simulation über 200 ns. Zudem geben die experimentellen Daten auch nicht unbedingt die reale Struktur des Proteins wieder. Des weiteren können auch experimentelle Daten unvollständig sein, sodass eventuell ein weiterer Cluster im Phasenraum von Abb. 11) existiert. Allerdings könnten hier weitere Bewegungen durch Simulationen erschlossen werden.

Ein interessantes Phänomen ist die Existenz von Clustern, die die experimentellen

Daten andeuten und die Simulation 3 aufzeigt. Diese könnten als zwei unterschiedliche Zustände des Proteins interpretiert werden, die während der Substratverarbeitung durchlaufen werden und für kurze Zeit stabil sind. Die Funktion dieser Zustände könnte unter Umständen dadurch aufgeklärt werden, dass man die Form und Lage des Substrates zu Zeitpunkten bestimmt an denen das Protein in einem der beiden Zustände ist. Damit wäre ein tiefer Einblick in die Arbeitsweise des Enzyms möglich.

## Literatur

- [1] C.M. Bishop. *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer, 2006.
- [2] Chemical Rubber Company. *CRC Handbook of Chemistry and Physics*. Number Bd. 79, Teile 1998-1999. CRC Press.
- [3] W. Demtröder. *Experimentalphysik 1: Mechanik und Wärme*. Springer-Lehrbuch. Springer Berlin Heidelberg, 2006.
- [4] GROMACS. <http://www.gromacs.org/>.
- [5] M. Henderson and M. S. Wertheim. Phase diagram and pv isotherms of argon. *The Journal of Chemical Physics*, 51(12):5420–5429, 1969.
- [6] Viktor Hornak, Robert Abel, Asim Okur, Bentley Strockbine, Adrian Roitberg, and Carlos Simmerling. Comparison of multiple amber force fields and development of improved protein backbone parameters.
- [7] Wanzer Drane R. H. Barham. An algorithm for least squares estimation of nonlinear parameters when some of the parameters are linear. *Technometrics*, 14(3):757–766, 1972.
- [8] J.B. Reece. *Campbell Biology*. Benjamin Cummings / Pearson, 2011.