

Strukturelle Bioinformatik (Masterstudiengänge Bioinformatik/Biochemie):

Lösungen der Übungsaufgaben zum Vorlesungsteil P. Güntert

1. Proteinkristallographie.

a) Geben Sie die 4 experimentellen und rechnerischen Hauptschritte einer Proteinkristallstrukturbestimmung an. Welcher dieser Schritte ist im Allgemeinen der unvorhersehbarste?

1. Proteinherstellung
2. Kristallisation (unvorhersehbarster Schritt)
3. Messung der Beugungsmuster
4. Datenauswertung

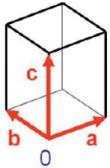
b) Welches sind die wichtigsten 4 Schritte der kristallographischen Datenauswertung?

1. Bestimmung der Einheitszelle und Raumgruppe
2. Phasenbestimmung
3. Modellbau
4. Verfeinerung der Phasen und der Struktur

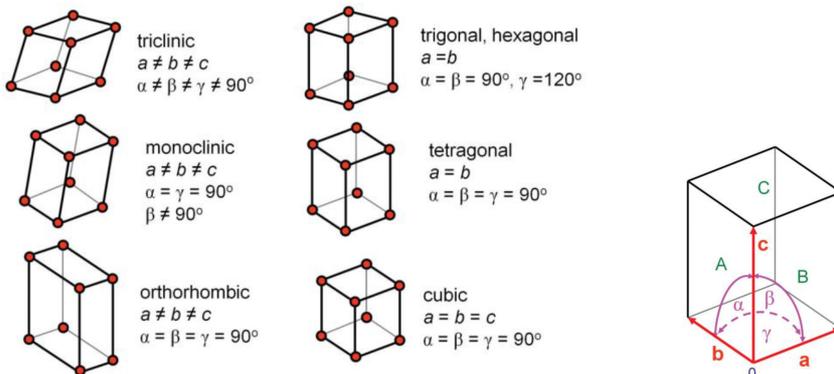
2. Proteinkristallographie.

a) Welche Größen definieren die Einheitszelle eines Kristalls?

Die drei Vektoren \vec{a} , \vec{b} , \vec{c} , die eine Einheitszelle aufspannen.



b) Welche primitiven 3-dimensionalen Gittertypen gibt es? Wie werden sie charakterisiert?



c) Warum besteht das Beugungsmuster eines Kristalls aus einzelnen Punkten und nicht aus einer kontinuierlichen Verteilung?

Im Beugungsmuster gibt es signifikante Intensität nur dort, wo die von jedem Gitterpunkt gestreuten Wellen konstruktiv interferieren, d.h. wenn die Bragg-Bedingung erfüllt ist. Bei einem (perfekten, unendlich großen) periodischen Gitter ist das nur an Punkten des reziproken Gitters der Fall.

3. Proteinkristallographie.

a) Was ist ein Strukturfaktor?

Eine komplexe Größe $F(hkl)$, die die Punkte des Beugungsmusters beschreibt. Die Punkte werden durch die Millerschen Indizes hkl indiziert.

b) Welche Größe wird in einem Röntgenbeugungsmuster gemessen?

Die Amplitude (der Absolutbetrag) des (komplexen) Strukturfaktors.

c) Wie hängen Strukturfaktoren und Elektronendichte mathematisch zusammen (Formel!)?

$$F(\mathbf{k}) = \int_R \rho(\mathbf{r}) e^{2\pi i \mathbf{r} \cdot \mathbf{k}} d\mathbf{r}$$

$$\rho(\mathbf{r}) = \int_{R^*} F(\mathbf{k}) e^{-2\pi i \mathbf{r} \cdot \mathbf{k}} d\mathbf{k}$$

$\rho(\mathbf{r})$ electron density at position \mathbf{r} in real space R
 $\rho(\mathbf{r}) \in \mathbb{R}$ is real
 $F(\mathbf{k})$ structure factor at position \mathbf{k} in reciprocal space R^*
 $F(\mathbf{k}) \in \mathbb{C}$ is complex with (measurable) amplitude $|F(\mathbf{k})|$ and (not measurable) phase $\alpha(\mathbf{k})$, i.e.
 $F(\mathbf{k}) = |F(\mathbf{k})| e^{i\alpha(\mathbf{k})}$

4. NMR Spektroskopie.

a) Was wird in einem NMR Experiment gemessen?

Die magnetische Resonanz der Spins der Atomkerne. D.h. elektromagnetische Strahlung, die von Kernspins abgestrahlt wird, die sich in einem äußeren Magnetfeld bewegen (präzessieren).

b) Wie kommt man von einem zeitabhängigen Signal auf das Spektrum?

Durch Fouriertransformation des FIDs (free induction decay).

c) Was versteht man unter „through-bond“ und „through-space“ NMR Experimenten?

Through-bond Experimente führen zu Kreuzsignalen, zwischen Atomen, die (direkt oder indirekt) über wenige (≤ 3) kovalente Bindungen miteinander verbunden sind. Through-space Experimente liefern Kreuzsignale zwischen Atomen die räumlich (aber nicht notwendigerweise auch in der kovalenten Struktur) benachbart sind (typischerweise Abstände $\leq 5 \text{ \AA}$ in NOESY Experimenten).

5. NMR Spektroskopie.

a) Wie kommt ein mehrdimensionales (2D, 3D) NMR Spektrum zustande?

Ein normales (1-dimensionales) NMR Experiment wird wiederholt, wobei eine Wartezeit in der Pulssequenz, die der Messung des FIDs (free induction decays) vorangeht, schrittweise erhöht wird. Dadurch erhält man ein Signal, das von 2 Zeitkoordinaten abhängt: $S(t_1, t_2)$, wobei t_1 der schrittweise erhöhten Wartezeit und t_2 dem zeitlichen Verlauf des FIDs entspricht. Werden zwei Wartezeiten in der Pulssequenz unabhängig voneinander erhöht, entsteht ein 3-dimensionales Spektrum usw.

b) Was ist ein Kern-Overhauserereffekt (NOE)?

Durch die Kreuzrelaxation, d.h. dem gemittelten Effekt der Dipol-Dipol Wechselwirkungen, der durch die zufällige Brownsche Bewegung des Moleküls in der Lösung erzeugt wird, wird Magnetisierung von einem Spin auf einen in räumlicher Nähe befindlichen Spin übertragen. Der NOE führt zu Kreuzsignalen im NOESY Spektrum und ist die Grundlage der Distanzmessung und damit der Strukturbestimmung mit NMR.

c) Wie hängt die Intensität eines NOESY Kreuzsignals mit der entsprechenden interatomaren Distanz zusammen?

Die Intensität (Integral) des Kreuzsignals ist umgekehrt proportional zur sechsten Potenz des Abstands, $I \propto d^{-6}$. Gilt für den (hypothetischen) Fall eines isolierten Spinpaars in einem starren Molekül. In der Praxis (Moleküle mit vielen Spinpaaren und internen Bewegungen) nur näherungsweise.

6. NMR Spektroskopie.

a) Was sind die wichtigsten Eingabedaten für eine NMR Strukturberechnung?

Obere Distanzschranken aus NOESY Kreuzsignalen.

b) Was ist eine mehrdeutige Distanzeinschränkung (ambiguous distance restraint)?

Eine Distanzeinschränkung, die von mehreren Atompaaren stammt bzw. stammen könnte. Die

c) Wie lautet eine typische Zielfunktion, die während einer NMR Strukturberechnung minimiert wird?

$$T = \sum_{\text{upper limits}} \Delta_u^2 + \sum_{\text{lower limits}} \Delta_l^2 + \sum_{\text{torsion angle restraints}} \Delta_a^2 + \dots$$

upper limits distance (NOEs) lower limits distance (steric) torsion angle restraints

d) Was ist „simulated annealing“ (in einer NMR Strukturberechnung)?

Eine Strukturberechnung mit MD Simulation bei einer hohen Temperatur zu beginnen (um effizient den Konformationsraum abzusuchen) und dann die Temperatur langsam auf einen tiefen Wert zu senken (um kleinere und kleinere Einzelheiten der Struktur genau festzulegen). Simulated annealing verringert die Gefahr, in einem lokalen Minimum „hängen zu bleiben“, drastisch.

e) Warum berechnet man im Allgemeinen mehrere Konformere anstatt einer Struktur?

Um einen Anhaltspunkt dafür zu erhalten, wie genau die experimentellen Daten die Struktur des Proteins bestimmen.

7. Klassische Mechanik.

a) Wie ist die Lagrangefunktion eines 1-dimensionalen klassischen mechanischen Systems mit dem Potenzial $V(x)$ definiert?

$$L = E_{\text{kin}} - E_{\text{pot}} = T - V = \frac{1}{2} m v^2 - V(x)$$

b) Wie lauten die Lagrange-Bewegungsgleichungen der klassischen Mechanik?

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{q}} - \frac{\partial L}{\partial q} = 0$$

c) Zeigen Sie, dass die Lagrange-Bewegungsgleichungen den Newtonschen Bewegungsgleichungen äquivalent sind, wenn man das System mit kartesischen Koordinaten beschreibt.

$$L = \frac{1}{2} m \dot{x}^2 - V(x) \Rightarrow \frac{d}{dt} \frac{\partial L}{\partial \dot{x}} - \frac{\partial L}{\partial x} = \frac{d}{dt} m \dot{x} + \frac{\partial V}{\partial x} \Rightarrow m \ddot{x} = - \frac{\partial V}{\partial x} = F$$

8. Klassische Mechanik.

a) Ein mechanisches System sei invariant gegenüber Verschiebungen entlang der x -Achse und Drehungen um die z -Achse. Welche Erhaltungsgröße hat es?

Die x -Komponente des Impulses und die z -Komponente des Drehimpulses.

b) Ein eindimensionaler harmonischer Oszillator ist durch das Potenzial $V(x) = \frac{1}{2} k x^2$ charakterisiert. Welche Erhaltungsgrößen hat er?

Nur die Gesamtenergie. Der Impuls ist nicht erhalten, weil das System nicht translationsinvariant ist. In einer Dimension gibt es keine Drehungen und damit keinen Drehimpuls.

9. Klassische Mechanik.

a) Wie lautet die allgemeine Lösung der klassischen mechanischen Bewegungsgleichungen für einen eindimensionalen harmonischen Oszillator mit dem Potenzial $V(x) = \frac{1}{2} k x^2$?

$$\text{Bewegungsgleichung (Newton): } m \ddot{x} = - \frac{\partial V}{\partial x} = -kx$$

Lösungsansatz: $x(t) = a \cos \omega t + b \sin \omega t$ erfüllt die Bewegungsgleichung.

Anfangsbedingungen: $x(0) = a$, $\dot{x}(t) = -a\omega \sin \omega t + b\omega \cos \omega t \Rightarrow \dot{x}(0) = b\omega$

Allg. Lösung: $x(t) = x(0) \cos \omega t + \frac{\dot{x}(0)}{\omega} \sin \omega t$

b) Wie lautet die spezielle Lösung, die zur Zeit $t = 0$ den Anfangsbedingungen $x(0) = a$ und $\dot{x}(0) = 0$ genügt?

$$x(t) = a \cos \omega t$$

c) Wie groß ist die Gesamtenergie des System aus b)?

Zur Zeit $t = 0$ ist die kinetische Energie 0 und die potenzielle Energie $\frac{1}{2}ka^2$. Weil die Gesamtenergie erhalten ist, ist sie immer gleich $\frac{1}{2}ka^2$.

10. Moleküldynamiksimulation.

a) In einem Protein mit N Atomen gebe es N_{bond} kovalente Bindungen und N_{nb} nichtbindende Paarwechselwirkungen. Wieviele kovalente Bindungen bzw. nichtbindende Paarwechselwirkungen gibt es in einem 3 mal größeren Protein mit $3N$ Atomen?

ca. $3N_{\text{bond}}$ kovalente Bindungen und $9N_{\text{nb}}$ nichtbindende Paarwechselwirkungen

b) Wieviele nichtbindende Wechselwirkungen gibt es etwa in einem Protein mit 1000 Atomen?

$$N(N-1)/2 = 1000(1000-1)/2 \approx 1000^2/2 = 5 \times 10^5$$

c) Wie kann der Rechenaufwand für nichtbindende Wechselwirkungen reduziert werden? Zeigen Sie die einzelnen Schritte.

Durch Cutoff in Verbindung mit Paarliten:

1. Finde alle Atompaaire mit Distanzen $\leq d_{\text{cut}} + \Delta$ und speichere diese Atompaaire in einer Paarlite. Dies erfordert die Berechnung der Distanz für alle Paare.

2. Während einer Reihe von (z.B. 50) MD-Schritten werden nur die Wechselwirkungen aus der Paarlite ausgewertet. Falls die Distanz $\geq d_{\text{cut}}$ ist, wird der Beitrag zur potenziellen Energie vernachlässigt. Der Rechenaufwand dafür ist proportional zur Länge der Paarlite und damit (für ein System, das viel größer als d_{cut} ist) proportional zur Molekülgröße (und nicht zum Quadrat davon).

3. Nach einer festgelegten Anzahl von MD Schritten oder wenn sich die Struktur signifikant verändert hat, ist die Paarlite nicht mehr aktuell. Sie muss erneuert werden, indem Schritt 1 wieder durchgeführt wird.

11. Moleküldynamiksimulation.

a) Wie lautet die Taylor Reihenentwicklung der Funktion $f(x) = \cos(2x)$ um die Stelle $x_0 = \pi$. Geben Sie die Formel bis zu Größen der Ordnung $O(x^2)$ an.

$$f(x) = f(\pi) + f'(\pi)(x - \pi) + \frac{f''(\pi)}{2}(x - \pi)^2 + O((x - \pi)^3) = 1 - 4(x - \pi)^2 + O((x - \pi)^3)$$

b) Wie funktioniert der Verlet Algorithmus zur numerischen Integration der Bewegungsgleichungen? Wie berechnen sich die Positionen $\vec{r}(t + \Delta t)$ aus den Grössen zu früheren Zeiten?

Man betrachtet die Taylorreihenentwicklung der Ortskoordinaten als Funktion der Zeit

$$r(t \pm \Delta t) = r(t) \pm \Delta t \dot{r}(t) + \frac{1}{2}\Delta t^2 \ddot{r}(t) \pm \dots$$

Durch Addition und Einsetzen der Newton Bewegungsgleichung folgt:

$$r(t + \Delta t) = 2r(t) - r(t - \Delta t) + \Delta t^2 F(t)/m$$

c) Wie können beim Verlet-Algorithmus die Geschwindigkeiten berechnet werden?

$$v(t) = \frac{r(t + \Delta t) - r(t - \Delta t)}{2\Delta t}$$