

# Übung 5: Strukturevaluation

GST, Wintersemester 2018/2019

## 1 Einleitung

Thema dieser Übung ist die Beurteilung der Qualität von Proteinstrukturmodellen. Folgende Merkmale können bei Proteinstrukturen auf Fehler hinweisen:

1. kollidierende Atome
2. ungewöhnliche Rückratwinkel
3. Seitenketten mit ungewöhnlichen Konformationen

Im Folgenden sollen zwei Strukturmodelle des menschlichen DJ-1 Proteins auf Vorkommen dieser drei Merkmale untersucht werden. Beide Strukturen wurden mittels Röntgenstrukturanalyse bestimmt, aber bei unterschiedlicher Auflösung. Glücklicherweise sind für beide Strukturen die Elektronendichtekarten, anhand derer sie erstellt wurden, frei zugänglich.

## 2 Die Strukturen in der PDB

Besuchen Sie die Internetseite der PDB ([rcsb.org](http://rcsb.org)) und suchen Sie nach den Strukturen mit den IDs 1j42 und 4rkw. Betrachten Sie auf der Übersichtsseite die Angaben zur Publikation, zum Experiment und zur Validierung. Wie unterscheiden sich die beiden Modelle?

## 3 Öffnen und Überlagern der Strukturen und ihrer Elektronendichtekarten in Chimera

Öffnen Sie Chimera mit dem Befehl

```
/usr/local/zbhtools/chimera/1.11/bin/chimera &
```

und öffnen Sie in Chimera eine Kommandozeile über

*Favorites* → *Command Line*

Geben Sie anschließend

```
open 4rkw
```

in der Chimera-Kommandozeile ein, um das Strukturmodell des Proteins aus der PDB zu laden.

Nun können sie mit dem Befehl

```
open edsID:4rkw
```

die dazugehörige Elektronendichte vom “Uppsala Electron Density Server” laden. Struktur und Karte sind korrekt überlagert und können zusammen bewegt werden. Wählen sie für die Struktur eine Darstellungsart, mit der Elektronendichte und Strukturmodell gut untersucht werden können. Mit dem Erscheinen der Elektronendichte öffnet sich außerdem das **Volume Viewer** Fenster. In diesem steuern Sie die Darstellung der Elektronendichte. Das Histogramm zeigt die Verteilung verschiedener Dichtewerte an und mit dem Marker steuern Sie den Wert, für den die Isolinien angezeigt werden. Um nur die Dichte anzuzeigen, welche das Molekül umgibt, kann im Menü des **Volume Viewer** das **Zone** Feld aufgerufen werden:

*Features → Zone*

Wenn Sie nun das Molekül auswählen und den erschienenen **Zone** Button klicken, wird die Darstellung der Dichte auf einen angegebenen Radius um die Auswahl beschränkt.

Laden Sie nun die Struktur 1j42 und anschließend die Elektronendichte dazu. Achten Sie dabei auf die hier beschriebene Reihenfolge. Sie sollten nun sehen, dass die zweite Struktur mit der Elektronendichte sich ein Stück entfernt vom ersten Modell befindet. Stellen sie bei beiden Elektronendichtekarten die gleiche **Step**-Zahl ein (am besten 1) und nutzen Sie auch für 1j42 das **Zone**-Tool. Um die Dichtekarten und die Modelle miteinander zu vergleichen sollen diese nun überlagert werden. Öffnen Sie zuerst

*Favorites → Model Panel*

um sicher zu sein, dass 1rkw und dessen Elektronendichte als Modelle 0 und 1 und 1j42 als Modelle 2 und 3 geöffnet sind. Nutzen Sie

*Tools → Structure Comparison → MatchMaker*

um Modell 2 auf Modell 1 (Referenzstruktur) zu überlagern. Geben Sie nun in der Kommandozeile

```
matrixcopy #2 #3
```

um die Transformation des Modells 2 auf die Elektronendichte (Modell 3) zu übertragen. Nun sollten sowohl die beiden Strukturen als auch deren Elektronendichte korrekt überlagert sein.

Vergleichen sie nun die beiden Elektronendichtekarten. Wie unterscheiden Sie sich? Wie sieht es zum Beispiel bei den Seitenketten mit aromatischen Ringen aus?

## 4 Auswertung von Rückratkonformationen

Um die Plausibilität von Strukturen zu prüfen werden oft die Rotationswinkel phi und psi des Proteinrückrats mit den Vorkommen dieser Winkel im Ramachandran-Plot verglichen. Ein Blick in den PDB-Eintrag zur Struktur 1j42 sollte ihnen bereits verraten haben, dass es in dieser Struktur ungewöhnliche Rückratkonformationen gibt (“Ramachandran Outliers”).

Lassen Sie sich die Ramachandran Plots (entsprechender Menu Button im *Model Panel*) für beide Strukturen anzeigen. Von potentiellen Outliern, d.h. den Punkten außerhalb der Konturen für den General case, stimmen bis auf einen alle mehr oder weniger überein (Winkel phi bei etwa  $41^\circ$  und psi bei etwa  $-166^\circ$  im Model 1j42). Mit einem Klick auf den Punkt wird der entsprechende Rest in der Struktur ausgewählt. Vergleichen Sie dort die Konformationen der beiden Strukturen. Was fällt ihnen auf? Wie verhält sich die Elektronendichte? Welche Konformation ist wahrscheinlicher?

## 5 Atomkollisionen

Laut den Angaben auf der PDB-Internetseite soll es in der Struktur 1j42 auch Kollisionen von Atomen (“Clashes”) geben. Einige lassen sich mit Chimera finden. Starten Sie:

*Tools* → *Structure Analysis* → *Find Clashes/Contacts*

Unter *Atoms to Check* können Sie Atome zum Prüfen auf Clashes auswählen. Wählen Sie dazu alle Atome von 1j42 mit der normalen Selektion aus und klicken Sie auf *Designate*. Wählen Sie bei *Check designated atoms against* die Option *themselves*. Behalten Sie alle anderen Einstellungen und klicken Sie auf OK. Ihnen werden nun die Kollisionen in der Struktur angezeigt. Um wie viele handelt es sich? Welche Distanzen sehen Sie? Was beobachten Sie hinsichtlich der Elektronendichte? Suchen Sie nun nach Kollisionen in der anderen Struktur. Was ist der offensichtliche Grund für das Ergebnis?

## 6 Seitenkettenrotamere

Um einzuschätzen ob die Seitenketten eines Modells plausibel sind, werden sie oft mit Rotameren aus einer Datenbank. Ausreißer, also Rotamere die nicht oder nur sehr selten vorkommen, können ein Hinweis auf fehlerhafte Strukturen sein.

Im Folgenden befassen wir uns ausschließlich mit Valin Aminosäuren. Aus Rotamerdaten aus einer Datenbank soll eine diskrete Wahrscheinlichkeitsverteilung erstellt werden. Die Valin-Konformationen im Modell 1j42 sollen unter Berücksichtigung dieser Wahrscheinlichkeitsverteilung betrachtet werden.

### 6.1 Rotationswinkel exportieren

Wählen Sie alle Valin-Vorkommnisse in 1j42 mit dem Befehl `select #2:val` aus. Starten Sie

*Tools* → *Depiction* → *Render by Attribute*

und von dort aus:

*File* → *Save Attributes*

Wählen Sie bei *residues* die Winkel *chi1* als Attribut aus und aktivieren Sie unbedingt die Checkbox *Restrict save to current selection, if any*, so dass nur Daten für die ausgewählten Valin Reste gespeichert werden. Speichern Sie die Daten in ihrem Home Verzeichnis mit einem sinnvollen Dateinamen. Öffnen Sie die Datei mit dem Texteditor *Kate* und versuchen Sie das Format zu verstehen, in dem Attribute aus Chimera exportiert (und später wieder in Chimera importiert) werden können.

## 6.2 Rotamerdaten

Kopieren Sie die Datei `/home/petersen/teaching/structure_evaluation/rota500-val.tab` in ein lokales Verzeichnis. Öffnen Sie die Datei anschließend mit dem Programm *LibreOffice Calc*. Bei dem Inhalt der Datei handelt es sich um einen Ausschnitt der *“Penultimate Rotamer Library” (SC Lovell, JM Word, DC Richardson (2000))*. Wählen Sie beim Textimport die Sprache *English (UK)*. Wählen Sie den Doppelpunkt als Trennsymbol für Spalten und aktivieren Sie die erweiterte Zahlenerkennung. Betrachten Sie die Tabelle. Welche Spalten liegen vor und was bedeuten die jeweiligen Einträge? Wofür könnte man die Daten, die zusätzlich zu den Winkeln vorliegen, nutzen?

## 6.3 Wahrscheinlichkeiten für diskrete Winkelklassen berechnen

Im Folgenden ist es zur Übersicht sinnvoll die Spalten, die erstellt werden sollen, je unter einer Zelle zu erstellen, die mit einem zugehörigen Titel versehen ist. Nutzen Sie den Funktions-Assistenten (über den Button *function wizard* mit dem Funktionssymbol) um die gefragten Funktionen zu erstellen und aktivieren Sie dort immer die Checkbox *Matrix (Array)* damit die Funktionsergebnisse immer die ausgewählten Bereiche füllen.

Die *chi1* Winkel sollen mit einer diskreten Verteilung modelliert werden. Entscheiden Sie sich für eine geeignete Anzahl gleich großer Intervalle um den Wertebereich von  $-180^\circ$  bis  $180^\circ$  zu teilen und leiten Sie die Intervallgröße her. Erstellen Sie neben Rotamerdaten eine Spalte von Klassen, in die Sie die oberen Intervallsgrenzen eintragen. Bei vier Intervallen und einer Intervallgröße von  $90^\circ$  würden entsprechend die Zahlen  $-90$ ,  $0$ ,  $90$  und  $180$  untereinander stehen. Erstellen Sie direkt rechts daneben eine Spalte mit den Häufigkeiten der Winkel, die in das jeweilige Intervall fallen. Verwenden Sie dafür die Funktion *FREQUENCY()*. Der erste Parameter der Funktion ist der Bereich mit den Winkelwerten (*chi1*) und der Zweite der Bereich mit den Intervallgrenzen. Erstellen Sie wiederum rechts daneben eine Spalte mit den relativen Häufigkeiten (bzw. Wahrscheinlichkeiten) der Winkel. Teilen Sie dafür den Bereich der absoluten Häufigkeiten durch seine Summe (Funktion *SUM()* und  $\$$  um Zeile und Spalte in einer Formel zu fixieren).

Welche absoluten und relativen Häufigkeiten beobachten Sie? Reicht die Menge der Daten aus, um die relativen Häufigkeiten als Wahrscheinlichkeiten zu bezeichnen? Wie entwickeln sich die Werte über die Intervalle hinweg? Zur Übersicht können Sie mit Hilfe eines Balkendiagramms ein Histogramm erstellen.

## 6.4 Wahrscheinlichkeiten für gemessene Winkel

Kopieren Sie die Daten (Residue Nummer und Winkel, nicht die ersten Header Zeilen) aus der mit *Chimera* erstellten Datei der Valin Winkel in die Tabelle rechts neben den Wahrscheinlichkeiten. Wählen Sie die Textimport Einstellungen wie zuvor aus, nur setzen Sie dieses mal den Tabulator als Spaltentrennsymbol. Zunächst sollen die präzisen Werte auf ein Vielfaches Ihrer gewählten Intervallgröße aufgerundet werden. Erstellen Sie dazu rechts neben den Winkeln eine Spalte, auf die Sie die Funktion *CEILING.MATH()* anwenden. Der erste Parameter ist der Bereich, der aufgerundet werden soll (die eben importieren Winkel) und der zweite ist die Zahl auf deren Vielfache aufgerundet wird (Intervallgröße). Vergewissern Sie sich stichprobenartig, ob das Aufrunden korrekt funktioniert (insbesondere bei negativen Zahlen).

Nun kann die Vorhersage der Winkelwahrscheinlichkeiten stattfinden. Kopieren Sie zunächst die Spalte der Residue Nummern in eine neue Spalte (dies wird später relevant). Rechts daneben sollen die Wahrscheinlichkeitswerte für die entsprechenden Residues erstellt werden. Für die Zuordnung kann die Funktion *VLOOKUP()* verwendet werden. Der erste Parameter ist dabei ein Bereich von Schlüsselwerten, für die in einem anderen Bereich zugehörige Wertepaare entsprechend rausgesucht werden sollen. Wählen Sie dafür die zuvor aufgerundeten Winkel aus. Der zweite Parameter ist ein Bereich von mehreren Spalten, in dem sich die Wertepaare befinden, die rausgesucht werden sollen. Die erste Spalte muss dabei die Schlüsselwerte des ersten Parameters enthalten. Wählen Sie dazu die drei an einander grenzenden Spalten mit den Intervallgrenzen, absoluten Häufigkeiten und Wahrscheinlichkeiten aus. Der dritte Parameter gibt an, in welcher Spalte des eben ausgewählten Bereichs die Daten liegen, die zu den Schlüsselwerten gehören. Wählen Sie hier die Zahl **3**, da dies die Spalte mit den Wahrscheinlichkeitswerten sein soll.

Es sollten nun die entsprechenden Winkelwahrscheinlichkeiten den Residue Nummern zugeordnet sein (vergewissern Sie sich wieder stichprobenartig). Kopieren Sie diese beiden Spalten und ersetzen Sie damit den Datenbereich in der aus *Chimera* erstellten Winkeldatei. Am Anfang jeder Zeile muss ein Tab-Zeichen vorliegen. Dies können Sie mit dem Texteditor *Kate* erzeugen, indem Sie zunächst in der Fußleiste unten rechts den Einrückungsmodus von Leerzeichen auf Tabulatoren ändern, dann den Datenbereich auswählen und mit Drücken der **Tab** Taste einmal einrücken. Ändern Sie in der ersten Zeile den Attributnamen von *chi1* entsprechend (z.B. zu *chi1\_probabilities*) und speichern Sie die Datei.

Starten Sie nun in *Chimera*

*Tools* → *Structure Analysis* → *Define Attribute*

und wählen Sie die eben gespeicherte Datei aus. Es erscheint wieder das Render by Attribute Menu, in dem jetzt das neue Attribut der Winkelwahrscheinlichkeiten vorliegt. Färben Sie die Valin Residues nach dem Attribut.

Was beobachten Sie? Wo liegen potentielle Outlier? Laden Sie die Struktur von höherer Qualität 4rkw und vergleichen Sie die Valin Seitenketten. Wo unterscheiden sie sich stark? Was sind die vorhergesagten Wahrscheinlichkeiten an diesen Stellen?

## 7 Aufgaben

Beantworten Sie die folgenden Fragen. Schicken Sie ihre Lösungen bis zum **9.12.2018** als pdf-Datei an **petersen@zbh.uni-hamburg.de**.

Verwenden Sie als Dateinamen **GST2018\_VornameNachname\_5.pdf** und als Email-Betreff **GST2018\_VornameNachname\_5**.

- Wie unterscheiden sich die Elektronendichtekarten zu den beiden untersuchten Modellen?
- Wie unterscheiden sich die beiden untersuchten Modelle hinsichtlich ihrer Qualität? Erläutern Sie dies (kurz) mit Bezug auf die untersuchten Kriterien.
- Woran konnte man bei abweichenden lokalen Strukturen erkennen welche Struktur eher richtig sein könnte?