



Übung 4: Sidechain Rotamers

Strukturqualität

Attributexport

Laden Sie die Struktur der menschlichen Deglycase **DJ-1 (1j42)** in **Chimera**. Verschaffen Sie sich einen Überblick über das Molekül und die Qualität der Struktur. Im Folgenden befassen wir uns ausschließlich mit Valin Aminosäuren. Wählen Sie alle Vorkommnisse dieser im Molekül aus. Starten Sie

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

und von dort aus:

Fie → **Save Attributes**

Wählen Sie für **residues** das wen 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 Hilfe eines Texteditors (**Kate**) und versuchen Sie das Format zu verstehen, in dem Attribute aus **Chimera** exportiert (und später wieder in **Chimera** importiert) werden können.

Rotamerdaten

Öffnen Sie die beigegefügte Datei *rota500-val.tab* mit dem Programm *LibreOffice Calc*. Hierbei handelt es sich um einen Ausschnitt der „*Penultimate Rotamer Library*“ (SC Lovell, JM Word, DC Richardson (2000))¹. 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?

Diskrete Klassen und Wahrscheinlichkeiten

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 um die gefragten Funktionen zu erstellen und aktivieren Sie dort immer die Checkbox *Matrix (Array)* damit die Funktionsergebnisse immer die ausgewählten Bereiche füllen.

Die χ^1 Winkel sollen mit einer diskreten Verteilung modelliert werden. Entscheiden Sie sich für eine geeignete Anzahl gleich großer Intervalle in die 360° (betrachtet von -180° bis 180°) geteilt werden und leiten Sie die Intervallgröße her. Erstellen Sie neben den Rotamer Daten eine Spalte von Klassen, in die Sie obere Intervallsgrenzen eintragen. Bei vier Intervallen und einer Intervallgröße von 90° 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 Winkeldaten (χ^1) 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()*).

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.

¹ kinemage.biochem.duke.edu/databases/rotamer.php

Winkel Wahrscheinlichkeitsvorhersage

Kopieren Sie die Daten (Residue Nummer und Winkel, nicht die ersten Header Zeilen) aus der mit **Chimera** erstellen 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 Intervallsgrenzen, 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?

Aufgaben

Bitte beantworten Sie die folgenden Fragen möglichst kurz und schicken Sie ihre Antworten an olzhabaev@zbh.uni-hamburg.de als Anhang im plain text oder pdf Format bis zum 04.12.2017, 08:00 Uhr. Fügen Sie dabei bitte das Stichwort [AST] dem Betreff zu.

- a) Es wurde ein Datensatz von χ_1 Winkeln der Valin Seitenkette verwendet. Nennen Sie einige sinnvolle Kriterien, die die Strukturen haben sollten, aus denen man so einen Datensatz erstellt, wenn man die Winkelwahrscheinlichkeiten modellieren möchte.
- b) Es wurde nur der Winkel χ_1 verwendet. Warum? Was ist bei der Modellierung von Wahrscheinlichkeitsverteilungen zu bedenken, wenn man weitere χ Winkel einbindet?
- c) Es wurde mit einer Anzahl gleich großer Intervalle für die Winkelklassen gearbeitet. Gibt es dabei potentiell Probleme bei der Repräsentation der wahren Verteilung der Winkel? Was wären mögliche Lösungen?