

ASE Übung 6 - Konservierung

Einführung

Das Ziel dieser Übung ist es sich mit Hilfe des Programms *UCSF Chimera* mit dem Kontext der Sequenz- und Strukturkonservierung vertraut zu machen. Anhand eines Beispielproteins soll eine Datenbanksuche nach ähnlichen Proteinsequenzen durchgeführt werden. Die Ergebnisse werden in einem multiplen Sequenzalignment betrachtet und auf Konservierung und Konsensus untersucht. Anschließend kann die Struktur hinsichtlich der Konservierung visualisiert werden.

UCSF Chimera

Hierbei handelt es sich um eine frei verfügbare Software zur Visualisierung und Analyse molekularer Strukturen und zugehöriger Daten (Sequenzen, chemischen Eigenschaften, etc.). Informationen und Dokumentation zu dem Programm sind auf folgender Seite verfügbar: <https://www.cgl.ucsf.edu/chimera/>

Starten Sie das Programm mit folgendem Kommandozeilenbefehl:

```
/usr/local/zbhtools/chimera/1.13/bin/chimera &
```

Sollte eine Meldung bzgl. Miskonfiguration von OpenGL erscheinen, klicken Sie auf Ja. Das Programm wird trotzdem funktionieren.

Über das Menü `File -> Fetch by ID...` ist es möglich Strukturmodelle direkt aus Datenbanken herunterzuladen und zu öffnen. Wählen Sie hier die *Protein Databank* (PDB) als Datenbank aus und tragen Sie bei ID das Modell `2c9v` ein. Hierbei handelt es sich um eine menschliche Kupfer/Zink Superoxiddismutase. Klicken Sie auf `Fetch`, um das Modell zu öffnen.

Im Hauptfenster können Sie nun die Struktur betrachten. Halten Sie die linke Maustaste gedrückt und bewegen Sie die Maus, um die Struktur zu drehen. Halten Sie die mittlere Maustaste (oder das Scrollrad) gedrückt und bewegen Sie die Maus, um die Struktur zu verschieben. Mit dem Scrollrad können Sie die Zoomstufe einstellen. Halten Sie den Mauszeiger über ein Atom oder ein Residue, um Informationen zu diesen anzeigen zu lassen. Verschaffen Sie sich einen Überblick über die Struktur mittels *Chimera*, sowie über das Molekül und das Modell mittels der PDB (<https://www.rcsb.org/>).

Initial wird die Struktur in der vereinfachten `ribbons` Darstellung präsentiert, welche die Sekundärstruktur und einige relevante Seitenketten und Ione darstellt. Über das Menü `Presets` lässt sich die atomare Darstellung mit `Interactive 2 (all atoms)` und die nach Hydrophobizität eingefärbte Oberfläche mit `Interactive 3 (hydrophobicity surface)` aktivieren.

BLAST Suche

Bei BLAST (*Basic Local Alignment Search Tool*) handelt es sich um ein Programm, welches für eine gegebene Sequenz ähnliche Sequenzen in einer Sequenzdatenbank findet und die statistische Signifikanz der Ähnlichkeit zurück gibt. Das Programm lässt sich für die etablierten Sequenzdatenbanken über ein Webinterface nutzen¹.

Dieser Webservice ist für Proteine innerhalb von *Chimera* verfügbar. Starten Sie eine BLAST Suche durch die folgenden Menüoptionen:

```
Tools -> Sequence -> Blast Protein
```

Es erscheint ein Fenster mit den Suchparametern. Unter dem Reiter *From Structure* lässt sich die Sequenz eines geöffneten Molekülmodells verwenden. Wählen Sie hier eine der Ketten des betrachteten Proteins.

Unter der Sequenzauswahl finden sich weitere Suchparameter. *Database* gibt die verwendete Datenbank an. Der *E-Value* (expect value) für eine gefundene Sequenz und ihren Ähnlichkeitsscore gibt an, wie viele Sequenzen mit diesem Score und der gegebenen Datenbankgröße rein zufällig gefunden werden können. Je geringer der E-Wert, desto signifikanter ist die Ähnlichkeit einer gefundenen Sequenz zur Zielsequenz. Hier dient der Parameter als ein Grenzwert, nach welchem gefundene Sequenzen gefiltert werden. *Matrix* setzt die Substitutionsmatrix fest, welche bei dem Vergleich der Sequenzen verwendet wird. Falls gefundene Strukturen aus mehreren Ketten bestehen und redundante Einträge entfernt werden sollen, kann die *List only best matching chain per PDB entry* Checkbox angeklickt werden.

Behalten Sie die Standardparameter und starten Sie die Suche, indem Sie auf *OK* klicken. Es erscheint ein Ergebnisfenster mit den in der PDB gefundenen Einträgen, welche nach einem angegebenen Ähnlichkeitsscore sortiert sind. Ebenfalls ist der Name des Modells mit einer Beschreibung und der *E-Value* gelistet.

Multipl. Sequenzalignment

Klicken Sie im BLAST Ergebnisfenster auf den Button *Show in MAV*. Es wird ein *MultiAlignViewer* Fenster geöffnet, in welchem das multiple Sequenzalignment der gefundenen Sequenzen präsentiert wird.

Im Header wird ein Histogramm zur Sequenzkonservierung und die Konsensussequenz angezeigt. Großbuchstaben zeigen dabei einen stärkeren Konsensus als Kleinbuchstaben und rot gefärbte Zeichen geben an, dass die Spalte zu 100% aus einem Element besteht. Die Histogrammwerte ergeben sich aus dem Verhältnis des Vorkommens des Konsensuselements zu der Gesamtanzahl der Elemente in der jeweiligen Spalte.

¹<https://blast.ncbi.nlm.nih.gov/Blast.cgi>

Betrachten Sie das Alignment. Wie entwickelt es sich, wenn Sie weiter rechts bzw. runter scrollen? Wie schätzen Sie die Sequenzkonservierung im Allgemeinen ein?

Konservierung

Wählen Sie im Menü des MultiAlignViewer folgendes aus:

```
Structure -> Render by Conservation
```

Es erscheint das Render/Select by Attribute Fenster, in welchem eine Färbung oder Auswahl von Atomen oder Residues nach einem bestimmten Attribut (z.B. Hydrophobizität) durchgeführt werden kann. Mit dem Alignment wurde die Sequenzkonservierung als Attribut `mapConservation` erstellt und bereits ausgewählt. Im Histogramm kann über die drei Farbbalken der Farbverlauf über die Attributswerte definiert werden. Behalten Sie alle Standardparameter und klicken Sie auf `Apply`. Betrachten Sie wieder die Struktur. Welche Residues sind stark konserviert? Spekulieren Sie über mögliche Gründe.

Lassen Sie sich die Oberfläche eines Monomers ebenfalls nach Sequenzkonservierung färben. Hierzu müssen die Ketten des Modells zunächst auf eigene Modelle aufgeteilt werden. Öffnen Sie die *Chimera*-Kommandozeile mit `Favorites -> Command Line`. Diese erscheint nun unter dem Hauptfenster. Geben Sie dort nun den Befehl `split` ein. Für jede der Ketten ist nun ein Modell erstellt worden.

Wählen Sie dann eine der Ketten aus (`Select -> Chain`) und wenden Sie die Oberflächendarstellung auf die Selektion an (`Actions -> Surface -> show`). Im Render/Select by Attribute Fenster können Sie nun die entsprechende Kette erneut nach Konservierung färben, so dass die Visualisierung auf die Oberfläche angewendet wird.

Sie sollten an zwei größeren Bereichen der Oberfläche stärkere Konservierung beobachten. Spekulieren Sie auch hier, warum dies strukturell relevant sein könnte. Die Färbung der Oberfläche nach einem anderen Attribut ist dabei hilfreich.

Fragen

1. Welche Vorhersagen lassen sich mit Hilfe konservierter Sequenzen treffen?
2. Welche Tendenzen bzgl. Sequenzkonservierung erwarten Sie, wenn das Innere eines Proteins mit dessen Oberfläche verglichen wird und warum?
3. Im gegebenen Beispiel wurde die Konservierung als Verhältnis des Konsensuselements zur Größe der Spalte berechnet. In welcher Hinsicht ist dieser einfache Ansatz problematisch? Was wäre eine bessere Alternative?