

## Einführung

Das Ziel dieser Übung ist es sich mit Hilfe von *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.

Starten Sie *Chimera* und öffnen Sie die Struktur der Kupfer/Zink Superoxiddismutase 2c9v. Verschaffen Sie sich einen Überblick über die Struktur und Funktion des Moleküls. Offensichtlich handelt es sich um eine Homodimerstruktur. Beide Monomere sind Teil des selben Modells. Oft ist es für gewisse Analysen einfacher die Ketten jeweils einem eigenen Modell zuzuweisen. Führen Sie dies mit dem Kommandozeilenbefehl `split` durch.

## 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 nutzen<sup>1</sup>.

Dieser Webservice ist für Proteine innerhalb des *Chimera* Programms verfügbar. Starten Sie eine BLAST Suche durch die folgenden Menu Optionen:

```
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...` Checkbox angeklickt werden.

Behalten Sie die Standardparameter und starten Sie die Suche, indem Sie auf `OK` klicken. Es erscheint ein Ergebnisfenster mit in der *Protein Datenbank* (<https://www.ebi.ac.uk/pdbe/>) gefundenen Einträgen, welche nach einem angegebenen Ähnlichkeitsscore sortiert sind. Ebenfalls ist der Name des Modells mit einer Beschreibung und der `E-Value` gelistet.

---

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

## Multiples Sequenzalignment

Klicken Sie 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 haben 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.

Betrachten Sie das Alignment. Wie entwickelt es sich, wenn Sie weiter rechts bzw. runter scrollen? Welche Residues sind Ihrer Meinung nach für die Struktur relevant?

## Konservierung

Wählen Sie dort im Menü des MultiAlignViewer folgendes aus:

Structure -> Render by Conservation

Es erscheint das Render/Select by Attribute Fenster. Mit dem Alignment wurde die Sequenzkonservierung als Attribut mavConservation erstellt und bereits ausgewählt. Visualisieren Sie die Sequenzkonservierung und 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. Wählen Sie dazu eine der Ketten aus und wenden Sie die Oberflächendarstellung auf die Selektion an (surface selection). 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 anderem 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?