

# Übungen zum Sequenzanalyse-Praktikum

Universität Bielefeld, SoSe 2015

Dr. Roland Wittler · M.Sc. Linda Sundermann

<http://wiki.techfak.uni-bielefeld.de/gi/Teaching/2015summer/SequaPrak>

[praktikum-seqan@CeBiTec.Uni-Bielefeld.DE](mailto:praktikum-seqan@CeBiTec.Uni-Bielefeld.DE)

**Übungsblatt 11 vom 29.06.2015**

**Abgabe bis Donnerstag, 24:00 Uhr.**

Benutze für diesen Zettel jeweils die *DIALIGN*-Version vom „Bielefelder Bioinformatik-Server“ (<http://bibiserv.techfak.uni-bielefeld.de/dialign>) und die *DIALIGN-Pfam*-Version vom „Göttingen Bioinformatics Compute Server“ (<http://dialign.gobics.de>).

## Aufgabe 1 (*DIALIGN 2*)

Auf der Praktikums homepage findest du die Datei *dialign\_sequences.fas*, mit der du in dieser Aufgabe arbeiten sollst.

1. Benutze die ersten beiden Sequenzen der Datei um mit *DIALIGN* ein Alignment mit Standardeinstellungen zu berechnen. Auf eine Validierung des Dateiformats kannst du verzichten. Betrachte nach der Berechnung den Output im „Custom Format“. Was bedeuten die kleinen und die großen Buchstaben im Alignment?
2. Wie viele Blöcke wurden aligniert? Fertige eine Skizze an.
3. Berechne nun ein Alignment mit allen drei Sequenzen aus der Datei. Wie sieht das Alignment nun aus? Fertige ebenfalls eine Skizze an. Betrachte dazu die nicht alignierten Regionen der dritten Sequenz genau. Warum kann sie nicht problemlos an die anderen beiden aligniert werden?

## Aufgabe 2 (*DIALIGN 2 im Vergleich*)

1. Berechne mit *DIALIGN* ein multiples Alignment der „unbekannten“ Beta-Lactamase-Sequenzen von der Veranstaltungshomepage.
2. Betrachte nun die Zahlen, die du unter jeder Alignmentsspalte im „Custom Format“ findest. (Achtung: Es kann sein, dass die Alignmentsspalten nicht richtig übereinander stehen, wenn das Format verschoben ist.) Wofür sind die Zahlen da? Was bedeutet ein Wert von 9, was einer von 0? Löse diese Aufgabe alleine mit diesem Alignment, du musst die Information nicht im Internet heraus suchen. Gibt es ähnliche Informationen über die Alignmentsspalten auch bei *Clustal W2* und *T-Coffee*. Wenn ja, welche Symbole wurden hier benutzt? Welche Darstellung findest du intuitiver und warum?
3. Erstelle für die Sequenzen multiple Alignments mit *Clustal W2* und *T-Coffee*. Benutze für beide Tools Standardeinstellungen und führe *Clustal W2* von der EMBL-Seite und *T-Coffee* von der *T-Coffee*-Plattform aus aus. Vergleiche die Alignments mit dem *DIALIGN*-Alignment. Diskutiere kurz die Ähnlichkeiten und Unterschiede (ca. sechs Sätze).
4. Betrachte das Alignment von *DIALIGN* nun genauer. Finde den Block über alle Sequenzen, der am besten konserviert ist. Woran erkennst du ihn? Gib an, über welche Positionen der Query-Sequenz sich dieser Bereich erstreckt. Füge zudem diesen Teil der Sequenz in dein Protokoll ein.
5. Ist dieser Bereich auch im Alignment von *Clustal W2* und *T-Coffee* konserviert? Woran erkennst du das.
6. Vergleiche den konservierten Bereich der Query-Sequenz mithilfe von *blastp* gegen die *SwissProt*-Datenbank mit Standardeinstellungen. Wird die Vermutung bestätigt, dass es sich bei dem Bereich um eine konservierte Domäne handelt?

### Aufgabe 3 (DIALIGN-Pfam at GOBICS)

1. Welche Versionen von *DIALIGN* findest du auf dem „Göttingen Bioinformatics Compute Server“?
2. Beschreibe kurz, wie *DIALIGN-Pfam* funktioniert.
3. Lasse *DIALIGN-Pfam* mit den auf der Seite gespeicherten Beispielsequenzen laufen. Betrachte die Proteindomänen und das Alignment zu ihnen genauer. Welche Domänen werden gefunden. An welchen Positionen in den Sequenzen startet der Match zur Domäne in *Pfam*? Welche Sequenz macht beim Alignment zum *Pfam*-Match Probleme und warum?
4. Beschreibe das fertige Alignment kurz.