



MAX-PLANCK-INSTITUTE FOR DEVELOPMENTAL BIOLOGY
Department of Protein Evolution
Prof. Dr. Andrei Lupas
Director

Freiwilliges Übungsblatt
für die Vorlesung

Einführung in die Bioinformatik

11. Die Sprache der Proteine – Evolution konservierter Proteinstrukturen

elektronische Abgabe an: andrei.lupas@tuebingen.mpg.de

Aufgabe 1 Homologie

- (a) Was bedeutet der Begriff 'Homologie' in der Biologie? Wie unterscheidet er sich von der 'Analogie'?
- (b) Extrahieren Sie die Sequenz des Proteins Oskar von *Drosophila melanogaster* aus der NCBI Datenbank [<http://www.ncbi.nlm.nih.gov/protein/>]. Sie finden zwei Isoformen, A und C. Was sind Isoformen und wie unterscheiden sich die beiden Isoformen von Oskar?
- (c) Führen Sie eine BLAST Suche mit Oskar durch [<http://blast.ncbi.nlm.nih.gov/Blast.cgi>]; Sie können Isoform A oder C benutzen. Es sind zwei Proteindomänen in dem Protein erkennbar, welche in anderen Proteinen ebenfalls konserviert sind - benennen Sie sie und kennzeichnen Sie sie in der Sequenz (zB durch farbliche Hervorhebung). Proteine, welche über statistisch signifikante Ähnlichkeit in Sequenzsuchen identifiziert werden, werden gängigerweise als Homologe bezeichnet. Warum?
- (d) Führen Sie nun eine PSI-BLAST Suche mit mehreren Iterationen durch [Option auf derselben Seite wie BLAST]. Wie unterscheiden sich PSI-BLAST und BLAST? Weshalb findet PSI-BLAST mehr Homologe? Sind Proteine bekannter Struktur unter den Homologen (Hinweis: Einträge mit bekannter Struktur tragen das Kürzel "pdb" in der Annotationszeile)?
- (e) Führen Sie nun eine HHpred Suche durch [<http://toolkit.tuebingen.mpg.de/hhpred>]. HHpred verwendet Vergleiche zwischen Hidden Markov Modellen um Sequenzähnlichkeit zu berechnen. Dies ist deutlich empfindlicher als Sequenz-Sequenz Vergleiche oder Sequenzprofil-Sequenz Vergleiche. Finden Sie nun signifikant verwandte Proteine bekannter Struktur? Sind diese aufgrund der vorhergegangenen Suchen als Homologe plausibel?

Aufgabe 2 Strukturähnlichkeit

- (a) Klicken Sie auf den signifikantesten Hit in der HHpred Suche. Dies führt Sie zu der Alignierung von Oskar mit der ähnlichsten Sequenz in der Strukturdatenbank. Klicken Sie nun auf den blau gekennzeichneten Hyperlink; dies führt Sie zu einem Eintrag in der PDB Strukturdatenbank [<http://pdb.rcsb.org/pdb/explore.do?structureId=4HF7>]. Laden Sie die Koordinaten herunter und betrachten Sie diese in einem Visualisierungsprogramm (zB PyMol, SwissPDB Viewer, BALLView). Wie würden Sie die Topologie des Kettenverlaufs beschreiben?
- (b) Der signifikanteste Hit in der HHpred Suche hat keine Verbindung zu der SCOP Datenbank (Structural Classification of Proteins), aber weitere Hits in der Liste zeigen diese Verbindung. Klicken Sie auf einen dieser Hyperlinks. Ist die Superfamilie, die in der SCOP Datenbank eingetragen ist, als Homologes zu der Oskar Domäne plausibel? Entspricht die beschriebene Topologie derjenigen, die Sie unter Punkt (a) ermittelt haben?