Andrew Torda
Timur Olzhabaev

Zentrum für Bioinformatik
Übung zur Vorlesung
Angewandte Bioinformatik: Strukturen
Wintersemester 2017/2018

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

12. Dezember 2017

Übung 6: Homology Modelling

Zielsequenz und Template Suche

Laden Sie die Struktur des Ferredoxins ***1DUR*** in ***Chimera***. Im Folgenden gehen wir davon aus, dass die Struktur des Proteins nicht bekannt ist, schalten Sie daher dessen Sichtbarkeit aus. Es wird jedoch die Sequenz für das *Homology Modelling* und die Struktur später zur Bewertung der Ergebnisse benötigt.

Führen Sie mit Hilfe des ***Blast Protein*** Tools eine ***BLAST***-Suche durch. Das beste Ergebnis wird offensichtlich die offene Struktur sein – ignorieren Sie diese. Laden Sie die Struktur des nächstbesten Ergebnisses ***1FCA*** (Bei ***1CLF*** handelt es sich um eine mit NMR ermittelte Struktur, die wir ebenfalls ignorieren). Diese Struktur wird im Folgenden als Template verwendet. Führen Sie ein Sequenzalignment beider Strukturen durch und berechnen Sie die Sequenzidentität (im Alignmentfenster: ***Info -> Percent Identity***). Bewerten Sie das Sequenzalignment. Reicht die Sequenzidentität aus um ***1FCA*** als Template zu verwenden?

MODELLER

Starten Sie aus dem Alignmentfenster das ***Modeller*** Tool:

Structure -> Modeller (homology)

Wählen Sie als Zielsequenz ***1DUR*** und als Template ***1FCA*** (Modell muss in der Template-Liste angeklickt werden) aus. Wählen Sie die Option ***Run Modeller locally*** aus und geben Sie für ***Location of Modeller executable*** folgenden Pfad an:

/home/torda/bin/modeller9.14/bin/mod9.14

Schauen Sie sich auch die Optionen unter ***Advanced Options*** an, lassen Sie aber die Standardeinstellungen. Klicken Sie auf ***Apply*** um eine Modellierung auszuführen.

Klicken Sie auf ***Apply*** um eine Modellierung auszuführen. Die erstellten Vorhersagemodelle werden in ***Chimera*** geladen und es erscheint ein Fenster in dem die Ergebnisse durch zwei Metriken bewertet werden, welche auf statistischem Potential beruhen. Ein ***GA341*** Wert von > 0.7 deutet dabei ein zuverlässiges Modell an, während für ***zDOPE*** (*normalized Discrete Optimized Protein Energy*) negative Werte bessere Modelle bedeuten. Beides bewertet dabei nur die Qualität der erstellten Strukturen, nicht die Ähnlichkeit zum Zielmodell.

Überlagern Sie die erstellten Modelle und das Template mit der Zielstruktur und vergleichen Sie diese. Wie bewerten Sie die Qualität des Rückgrats? Wählen Sie ein Vorhersagemodell aus und berechnen Sie die Rückgrat-RMSD Werte zwischen diesem und dem Zielmodell (***rmsd #<Modell Nr>:@n@ca@c #<Modell Nr>:@n@ca@c***). Berechnen Sie auch die Rückgrat-RMSD Werte zwischen der Vorhersage und dem Template und zwischen dem Template und dem Zielmodell. Lassen Sie sich die Seitenketten anzeigen. Wo gibt es größere Unterschiede? Gibt es Auffälligkeiten im Sequenzalignment an diesen Stellen?

Größere Strukturen

Schließen Sie alle Modelle und wiederholen Sie die vorherigen Anweisungen für die Struktur ***5L43*** (entfernen Sie eine der Ketten nach Öffnen der Struktur). Verwenden Sie dabei ***1Y79*** als Template (bei ***5L44*** handelt es sich um das gleiche Molekül). Beobachten Sie dabei wieder die Qualität des Sequenzalignments, die Sequenzidentität und die strukturelle Ähnlichkeit zum Zielmodell. Lassen Sie aus Zeitgründen dieses Mal nur ein Vorhersagemodell erstellen.

Vergleichen Sie anschließend das Vorhersagemodell mit dem Zielmodell. Bleiben die strukturellen Elemente gut erhalten? Gibt es interessante Auffälligkeiten und können Sie diese erklären? Wie steht es um die interagierenden Seitenketten am aktiven Zentrum? Würde dieses anhand der lokalen Konformationen korrekt erkannt werden?

Mehrere Templates

Laden Sie die Struktur ***4OZQ*** und entfernen Sie wieder eine der Ketten. Führen Sie eine ***BLAST*** Suche durch, lassen Sie sich das multiple Sequenzalignment anzeigen und verschaffen Sie sich einen Überblick. Was fällt Ihnen bzgl. Sequenzlücken auf (achten Sie auf den Bereich um Position 800)? Schauen Sie sich die Struktur und die Residue Nummerierung genau an. Was vermuten Sie?

Führen Sie eine Modellierung mit den Strukturen ***4FE8*** und ***3GBJ*** (löschen Sie jeweils 2 der 3 Ketten) als Templates durch (auch dieses Mal nur mit einem Vorhersagemodell). Für die Modellierung ist die relative Position der Templates relevant. Bewegen Sie daher vorher die Strukturen und platzieren Sie den N-Terminus von ***3GBJ*** nahe des C-Terminus von ***4FE8*** nach Augenmaß.

Bewerten Sie die Qualität des erstellten Modells. Ist die grobe Struktur dem Zielmodell ähnlich? Überlagern Sie die Templates auf das Vorhersagemodell und schauen Sie, ob diese wiederzuerkennen sind. Schauen Sie sich nur das Vorhersagemodell an. Müssen Sie Qualitätskriterien (Clashes, Winkel-Outlier, etc.) berechnen um einzuschätzen, ob es sich um eine realistische Struktur handelt?

Schauen Sie sich das Sequenzalginment der Zielstruktur und der Templates nochmal an. Im Abschnitt 374-403 gibt es eine interessante Überlagerung. Ist diese sinnvoll? Wählen Sie den entsprechenden Abschnitt in der Struktur ***4FE8*** aus. Um welchen Teil handelt es sich dabei? Löschen Sie diesen Teil der Struktur und führen Sie ein neues Sequenzalignment und Modelling durch (vergessen Sie nicht die Strukturen wieder an den entsprechenden Enden zu arrangieren).

Welche Ergebnisse beobachten Sie jetzt? Schauen Sie sich die Zielstruktur an und schätzen Sie ein, ob die Anordnung der Domänen mit *Homology Modelling* ohne weiteres korrekt vorhergesagt werden kann.

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 18.12.2017, 08:00 Uhr. Fügen Sie dabei bitte das Stichwort [AST] dem Betreff zu.

a) Nennen Sie einige Gründe, aus denen man gezwungen ist eine Struktur mit Hilfe von *Homology Modelling* vorherzusagen.

b) Viele Schritte des Modelling Protokolls wurden von ***MODELLER*** übernommen. Um welche handelt es sich dabei? Beschreiben Sie diese kurz.

c) Warum sollte nach der Templatesuche mit Hilfe von ***BLAST*** ein eigenes Sequenzalignment zwischen Zielstruktur und Templates durchgeführt werden und nicht der entsprechende Teil des erstellten multiplen Sequenzalignments verwendet werden?

d) Nennen Sie einige Gründe dafür mehrere Templates bei der Modellierung zu verwenden.