

Übung: Die PDB und Proteindomänen

In dieser Übung werden wir die Internetseite der PDB ('Protein Data Bank') kennenlernen und die im Browser zugänglichen Funktionalitäten nutzen um verschiedene Methoden zur Bestimmung von Proteindomänen vergleichen.

1 Eine Proteinstruktur suchen und finden

Besuche die Seite www.rcsb.org. Auf der Startseite befindet sich ein Eingabefeld über das man Strukturen finden kann. Kennt man den Identifikationscode ist es einfach. Gib z.B. 1CXL ein. Nun wollen wir die erweiterte Suchfunktion nutzen, um ein Protein zu finden, das bestimmte Kriterien erfüllt. Auf der Startseite findet man einen Link zur erweiterten Suche ('advanced search') unter dem Eingabefeld für die einfache Suche. Dort kann man beliebig viele Suchkriterien auswählen.

Wir wollen uns mit Proteindomänen beschäftigen und suchen dementsprechend nach der Struktur eines Proteins. Wähle unter 'Structure Features' die Option 'Macromolecule Type' und gib an, dass du nur Proteine und keine Nukleoteinsäuren wünschst. Füge ein weiteres Suchkriterium hinzu und wähle unter 'Methods' die Option 'Experimental Method' um dort 'X-RAY' auszuwählen. Außerdem suchen wir eine Struktur die groß genug ist, um aus mehr als einer Domäne zu bestehen, aber gleichzeitig noch einigermaßen überschaubar ist. Suche deshalb mit der Option 'Chain Length' unter 'Sequence Features' nach Strukturen mit 250 bis 300 Aminosäuren Länge.

Aktiviere außerdem das Feld 'Retrieve only representatives at 70% sequence identity' um nicht zu viele ähnliche Strukturen zu finden. Das Feld befindet sich unter den Suchoptionen.

Jetzt kann die Suche gestartet werden. Es dauert einen kleinen Moment. Auf der Ergebnisseite findest du nun einige Proteine. Nimm nicht das erste und nicht das gleiche wie dein Sitznachbar, sondern such dir das schönste oder interessanteste Protein aus.

2 Proteindomänen in der PDB

Die PDB zeigt die Ergebnisse verschiedener Werkzeuge zum Bestimmen von Proteindomänen. Tabelle 2 beinhaltet eine Zusammenfassung. Die beiden Klassiker sind SCOP und CATH. Beide versuchen die Welt der Proteine in eine hierarchische Struktur aus evolutionär verwandten Domänen zu ordnen. Erstere basiert weitgehend auf Handarbeit von Experten während letztere automatisiert ist. Andere Methoden wie DomainParser und ProteinDomainParser implementieren Algorithmen, welche versuchen die Proteinstruktur geschickt zu zerteilen. Pfam

und InerPro klassifizieren Proteinsequenzen. Hat man nur eine Proteinsequenz aber keine Struktur, kann man diese Werkzeuge nutzen um Domänen vorherzusagen.

Methode	Kurzbeschreibung	Server/Download & Dokumentation
SCOP	Handgemachte Klassifizierung	http://scop2.mrc-lmb.cam.ac.uk/
CATH	Automatisierte Klassifizierung	http://www.cathdb.info/
DomainParser2	Erkennung von Domänen in Strukturen	http://csbl.bmb.uga.edu/downloads
ProteinDomainParser	Erkennung von Domänen in Strukturen	
Pfam	Klassifizierung der Sequenzen	http://pfam.xfam.org/
InterPro	Klassifizierung der Sequenzen	http://www.ebi.ac.uk/interpro/

Es sind nicht immer Ergebnisse aller Methoden vorhanden. Die Werkzeuge sind jedoch normalerweise auch anderweitig verfügbar - entweder als Server oder als Programme, die man auf dem eigenen Rechner installieren kann. Informationen über die Methoden findet man auf den entsprechenden Internetseiten, in den Originalpublikationen oder im Web (z.B. in Wikipedia).

2.1 Visualisierung von Struktur und Domänen

Auf der PDB-Seite einer Proteinstruktur befinden sich oben mehrere Tabs. Zunächst befindet man sich im Bereich 'Structure Summary'. Daneben finden wir '3D View'. Dort kann man die 3D Struktur im Jmol-Viewer (oder einer Alternative) betrachten. Unter der Strukturansicht kann man auf Domain klicken um ein Fenster zu öffnen, in dem Domänen angezeigt werden, die auf verschiedene Art bestimmt wurden. Neben der SCOP-Klassifizierung befinden sich dort meistens mit DomainParser (DP) und ProteinDomainParser (PDB) bestimmten Domänen. Klicke auf 'select all' um die verschiedenen Domänen zu betrachten. Vergleiche die Ergebnisse der verschiedenen Methoden. Speicher das Bild mit einem Rechtsklick in den Bildbereich und anschließender Wahl von 'File → Export'.

2.2 Das Sequenzfenster

Ein weiterer Tab oben auf der Seite ist 'Sequence'. Hier wird die Aminosäuresequenz mit einer Gruppe von Annotationen angezeigt. Die Liste mit den Annotationen befindet sich links daneben. Durch 'hide' lassen sich welche entfernen und unter 'Add an Annotation' neue hinzufügen. Im besten Fall sind alle in Tabelle 2 gelisteten Methoden vorhanden. Klickt man auf die Domänenabschnitte werden sie in der 3D-Struktur gezeigt. Wähle die Annotationen aus und vergleiche die Domänen.

2.3 Domänen-Bestimmung aufgrund der Aminosäuresequenz

Wie bereits erwähnt, kann man auch mit der Aminosäuresequenz Information über die Domänenstruktur eines Proteins erhalten. Pfam und InterPro sind Datenbanken, die klassifizierte Sequenzen von Proteinen enthalten. Man kann die Sequenz eines Proteins mit der Datenbank vergleichen und dadurch etwas über die Domänen eines beliebigen Proteins erfahren. Strukturinformation braucht man dafür keine.

Häufig kann man die Annotationen in der zuvor beschriebenen Sequenzbetrachtung sehen. Ist das nicht der Fall, kann man <http://pfam.xfam.org/> bzw. <http://www.ebi.ac.uk/interpro/> besuchen. Um die Sequenz eines Proteins aus der PDB zu bekommen, nutze den Link 'Display Files' in der PDB Webseite und wähle die FASTA-Datei. Kopiere den Inhalt in das dafür vorgesehene Feld der Pfam bzw. InterPro Webseite.

3 Aufgaben

Illustriere die Ergebnisse mit Abbildungen der Strukturen und Sequenzannotationen.

1. Das Programm DomainParser2 ist unter <http://csbl.bmb.uga.edu/downloads> beschrieben. Erkläre kurz wie die Methode funktioniert.
2. Beschreibe deine Proteinstruktur. Wie groß ist sie? Von welchem Organismus stammt sie? Kannst du bestimmen was ihre Funktion ist? Aus welchen Sekundärstrukturelementen ist dein Protein aufgebaut?
3. Wie unterscheiden sich die Domänen, die von den unterschiedlichen Methoden/Klassifizierungen stammen? Wo unterscheiden sie sich? Wann sind sie gleich? Kannst du die Unterschiede erklären?
4. Falls eine oder mehrere der Methoden für dein Protein nicht in der PDB vorhanden sind, kannst du die Domänen über einen in Tabelle 2 genannten Server bekommen?
5. Nutze die Internetseite von Pfam und/oder InterPro, um die Domänenstruktur über die Sequenzen vorherzusagen. Was passiert wenn du einen Teil der Sequenz weglässt? (z.B. eine ganze annotierte Domäne, oder eine halbe) Was passiert wenn du einen Teil vervielfachst oder einzelne Reste entfernst oder ersetzt?
6. Wenn du Zeit und Lust hast, probiere noch eine grössere Struktur aus.