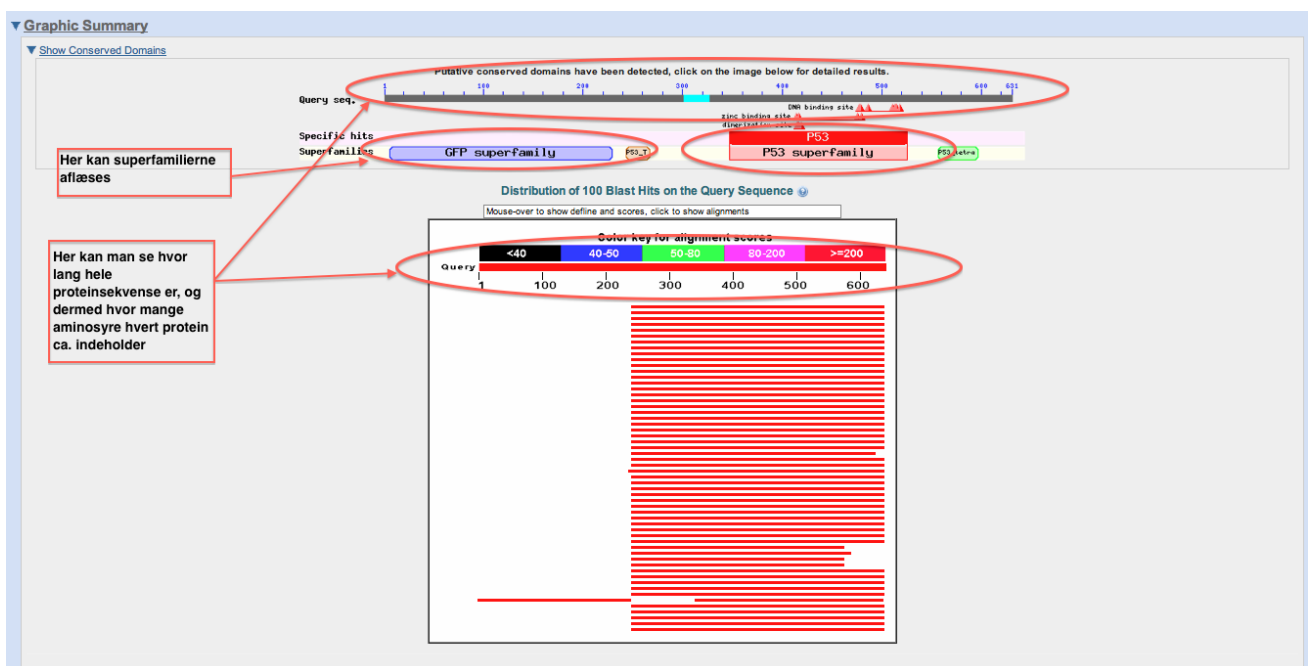


Svar til "Identifikation og visualisering af ukendt protein"

Svar til 4.A BLAST

1. Første del af sekvensen tilhører GFP (*Green Fluorescent Protein*) superfamilien og anden del af sekvensen tilhører P53 superfamilien.
2. Delen der hører til GFP superfamilien er ca. 230 AA lang og delen der hører til P53 superfamilien er ca. 400 AA lang.



3. Ja, det kan godt antages at det er de korrekte proteiner der er fundet. Starter man med at kigge på den del der tilhører P53 superfamilien skal man se på alle de øverste resultatsekvenser. Kigger man på deres e-værdi ses, at næsten alle har en e-værdi på 0. Da e-værdien siger noget om antal forventede tilfældige hits med den score er en lav e-værdi ensbetydende med, at resultatsekvenserne ikke er blevet fundet tilfældigt, men faktisk matcher query sekvensen. Ser man på sekvensdækningen (eng. *Query coverage*) er denne kun på 62%. Når man ser på denne dækningsprocent skal det huskes, at proteinet fra P53 superfamilien kun udgør omkring halvdelen af hele query sekvensen, og dermed er en dækning på 62% af hele sekvensen helt fint.
Ser man på den del der tilhører GFP superfamilien skal man se på de to nederste resultatsekvenser. Ved disse er e-værdien steget til $1e-174$ og $2e-174$, hvilke dog stadig er små nok til at sekvenserne kan antages ikke at være tilstede tilfældigt og dermed er korrekte fund. Sekvensdækningen er på 37%, hvilke stemmer overens med de 62% som P53 proteinet dækkede, så alt i alt antages det at begge proteiner er korrekt fundet gennem BLAST

Svar til 4.B UniProt

1. Accession nummeret for dette GFP er: P42212
2. Gennavnet er *GFP* og det stammer fra organismen *Aequorea victoria*, der er en vandmand.
3. Den er 238AA lang.
4. Da GFP udsender grønt lys vil man kunne bruge proteinet til at skelne mellem mutanter, da man kan lave DNA sekvenser der indeholder både GFP og det gen man har muteret (ændret) og dermed skal undersøge. Når det muterede gen bliver udtrykt, bliver GFP også udtrykt, og ved bestemme tilstedeværelsen af GFP og dennes intensitet vil man da kunne bestemme om det muteret gen er blevet udtrykt og hvor meget det bliver udtrykt.
GFP kan desuden bruges til at bestemme hvor i en organisme bestemte proteiner har deres funktion da GFP sammen med protein vil transproteres til dette sted og dermed lyse op, se fig ref.
5. Dette varierer med svaret for 4.
6. Accession nummeret for dette P53 er: P04637
7. Gennavnet er *TP53* og det stammer fra *Homo Sapiens* (mennesker).
8. Det er 393AA langt.
9. Under afsnittet function under sektionen *General annotation (Comments)* kan det læses, at P53 er en tumor suppresser og dermed undertrykker tumorer vækst. Dette sker, da P53 er bevirker til apoptose (programmeret celledød) samt inducere vækst standsning og dermed stopper cellevækst.
10. Det er forbundet med kræft. I omkring 60% af kræfttilfælde er P53 muteret eller på en anden måde forhindret i at udføre sin væksthæmmende egenskaber. Dette er ensbetydende med at kræftcellen kan få lov til at vokse og dermed sprede sig. Se eventuelt afsnittet *Involvement in disease* under sektionen *General annotation (Comments)*.

Svar til 4.C PDB

1. Se under *Summary* fanen. Id'et koder for et kompleks med P53 fra et menneske og DNA.
2. Se under *Sequence*. Id'et indeholder 6 kæder, hvor af fire er identiske. Kæde A, B, C og D koder alle for P53 molekylet og kæde K og L koder for hver en 26 nukleotider lang DNA streng, hvilke er komplementære.
3. Der er 234 aminosyrer i kæde A, B, C og D, disse er fra position 94 til 356. Dette stemmer ikke overens med længderne fundet tidligere, og det skal derfor huskes, at der mangler aminosyrer i proteinet hvis man arbejder videre med det. Dette kunne eksempelvis i PyMol, hvor man dermed ikke vil kunne finde og markere alle domænerne på proteinet.

4. Zink ionen er bundet ved aminosyrene: cystein på position 176, histidin på position 179, cystein på position 238 og cystein på position 242.

Svar til 4.D PyMol

1. Benyt kommandolinjen til dette. For at gemme alt, benyttes kommandoen: *hide everything* og for at vis det hele i tegnings struktur benyttes kommandoen: *show cartoon*. Nedenfor ses et billede af strukturen i PyMol:

For Educational Use Only



2. Der er flere måder dette kan gøres på:

a. Benyt kommandolinjen med:

color tv_red, chain A

color raspberry, chain B

color salmon, chain C

color ruby, chain D

b. Benyt kommandolinjen med:

sele A, chain A

sele B, chain B

sele C, chain C

sele D, chain D

Dette gøres for først vælge de to kæder og gemme dem som separate domæner i højre margin. Herefter benyttes ASHLC tasterne til at farve de valgte domæner, A, B, C og D, hver deres farve (*color sele* → *vælg en farve*).

Nedenfor ses et billede af strukturen i PyMol:

For Educational Use Only



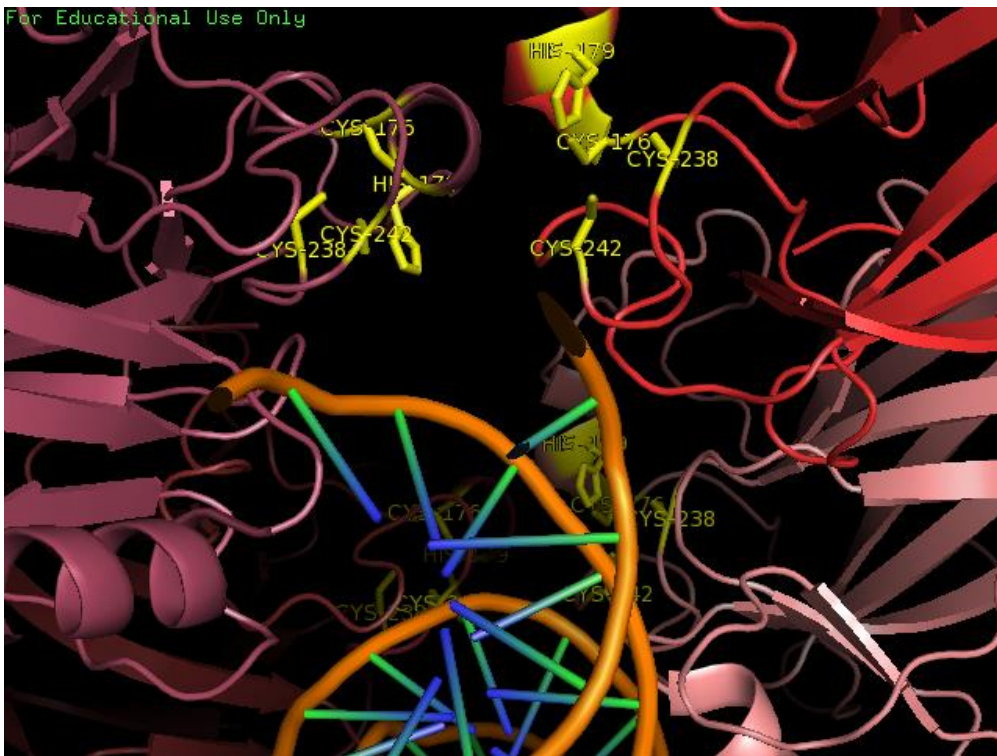
3. Benyt kommandolinien til at vælge zink fingeren:

```
sele zink_finger, resi 176+179+238+242
```

Denne kommando vil automatisk vælge de givne positionerne på alle kæderne.

Benyt herefter ASHLC tasterne til at vise sidekæderne (*show* → *sidechains* → *sticks*), farve aminosyrerne og sætte mærkat dem (*label* → *residue*). Til slut kan der zoomes ind på interaktionen ved at benytte *zoom zink_finger, -8*.

Nedenfor ses et billede af strukturen i PyMol:



4. Zink fingeren er i tæt forbindelse med DNA strengen, hvilket er at forvente da zink fingeren er med til at skabe stabilitet omkring protein-DNA interaktionen. Endvidere er zink ionen positivt ladet og DNA-strengen negativt ladet, så de modsatte ladninger er med til at skabe stabilitet, hvormed zink fingeren bliver nødt til at være placeret relativt tæt på DNA strengen.
5. Baggrundsfarven kan ændres i menulinien *Display* → *Background*, skyggerne under *Setting* → *Rendering* → *Shadows* og måden hvorpå den sekundære struktur bliver vist på kan ændres i *Setting* → *Cartoon*.