

Opgave 5

Undersøgelse af spytamylase ved hjælp af bioinformatik

Denne øvelse i bioinformatik er sat sammen af tre deløvelser (A-C).

I øvelse A skal du finde læserammerne for en kendt mRNA-streng af spytamylase ved hjælp af programmet 'Virtual Ribosome'.

I øvelse B skal du finde oplysninger om det protein som mRNA-strengen translateres til ved hjælp af databasen 'UniProt'.

I øvelse C skal du finde oplysninger om proteinets struktur ved hjælp af databasen 'PDB' og derefter visualisere strukturen ved hjælp af programmet 'Protein Workshop'.

Du har fået følgende mRNA streng til rådighed:

```
>gi|56549659|ref|NM_001008221.1| Homo sapiens amylase, alpha 1A (salivary) (AMY1A),  
transcript variant 2, mRNA  
CTTCTCAATATCAGCACTGGATTGTAGAAGCTTGTGCTGATTTTGGCCTGGCATTCAAGTAACTCTTCC  
CCTTGGTATCTGTACATACCTTTGATGTCAGTGTGTTAGTACACGTGGCTTGGTCACTTCATGGCTAAAAA  
CGTGCTTGTGGAAGACAAGTCTGGCTTGAATATAAATAGTTTCTGGAAAGGACACTGACAACCTCAAAG  
CAAAATGAAGCTCTTTTGGTTGCTTTTACCATTGGGTTCTGCTGGGCTCAGTATTCCTCAAATACACAA  
CAAGGACGAACATCTATTGTTTCATCTGTTTGAATGGCGATGGGTTGATATTGCTCTTGAATGTGAGCGAT  
ATTTAGCTCCCAAGGGATTTGGAGGGGTTTCAGGTCTCTCCACCAAATGAAAATGTTGCCATTCACAACCC  
TTTCAGACCTTGGTGGGAAAGATACCAACCAGTTAGCTATAAATATGCACAAGATCTGGAAATGAAGAT  
GAATTTAGAAACATGGTGACTAGATGCAACAATGTTGGGGTTCGTATTTATGTGGATGCTGTAATTAATC  
ATATGTGTGGTAATGCTGTGAGTGCAGGAACAAGCAGTACCTGTGGAAGTTACTTCAACCCTGGAAGTAG  
GGACTTTCCAGCAGTCCCATATTCTGGATGGGATTTTAATGATGGTAAATGTAAAACCTGGAAGTGGAGAT  
ATCGAGAACTATAATGATGCTACTCAGGTCAGAGATTGTCGCTGTCTGGTCTTCTCGATCTTGCAGTGG  
GGAAGGATTATGTGCGTCTAAGATTGCCGAATATATGAACCATCTCATTGACATTGGTGTGTCAGGGTT  
CAGAATTGATGCTTCCAAGCACATGTGGCCTGGAGACATAAAGGCAATTTTGGACAACTGCATAATCTA  
AACAGTAACTGGTTCGCGGAAGGTAGTAAACCTTTTCAATTTACCAGGAGGTAATTGATCTGGGTGGTGAGC  
CAATTTAAAAGCAGTGACTACTTTGGTAATGGCCGGGTGACAGAATTCAGTATGGTGCAAACTCGGCAC  
AGTTATTCGCAAGTGAATGGAGAGAAGATGTCTTACTTAAAGAACTGGGGAGAAGGTTGGGGTTTCATG  
CCTTCTGACAGAGCGCTTGTCTTTGTGGATAACCATGACAATCAACGAGGACATGGCGCTGGAGGAGCCT  
CTATACTTACCTTCTGGGATGCTAGGCTGTACAAAATGGCAGTTGGATTTATGCTTGCTCATCCTTATGG  
ATTTACACGAGTAATGTCAAGCTACCGTTGGCCAAGATATTTTGAAAATGGAAAAGATGTTAATGATTGG  
GTTGGGCCACCAAATGATAATGGAGTAACTAAAGAAGTTACTATTAATCCAGACACTACTTGTGGCAATG  
ACTGGGTCTGTGAACATCGATGGCGCCAAATAAGGAACATGGTTAATTTCCGCAATGTAGTGGATGGCCA  
GCCTTTTACAACTGGTATGATAATGGGAGCAACCAAGTGGCTTTTGGGAGAGGAAACAGAGGATTCATT  
GTTTTCAACAATGATGACTGGACATTTTCTTTAACTTTGCAAACCTGGTCTTCTGCTGGCACATACTGTG  
ATGTCATTTCTGGAGATAAAATTAATGGCAACTGCACAGGCATTAATAATCTACGTTTCTGATGATGGCAA  
AGCTCATTTTTCTATTAGTAACTCTGCTGAAGATCCATTTATTGCAATTCATGCTGAATCTAAATTGTAA  
AATTTAAAATTAATGCAATCCGCAAAGCA
```

mRNA-strengen er 'transkriberet' tilbage til en DNA-streng.

A. Virtual ribosome

Formålet med øvelsen er at blive bekendt med codons og læserammer på et DNA-molekyle.

Læs evt. om dette på følgende link:

<http://www.biotechacademy.dk/Undervisningsprojekter/Gymnasiale-projekter/Bioinfo/1-Teori2/1-Codons>

1. Du skal bruge programmet Virtual Ribosome, som kan lave en online translation af en mRNA-streng.

- Gå til Virtual Ribosome: <http://www.cbs.dtu.dk/services/VirtualRibosome/>
- Kopier mRNA-sekvensen af amylase ind i feltet i Virtual Ribosome.
- Translater sekvensen i Virtual Ribosome med standardindstillingerne ved at klikke på 'submit query'-knappen.

Programmet viser nu en 'translation report', hvor du kan se følgende:

- mRNA sekvensen hvor baserne tre og tre er blevet oversat til en aminosyre.
- Nogle tegn '>>>' og ')))' og '***' under sekvensen. '>>>' betyder at her kan være en startcodon. ')))' betyder at her kan være en alternativ startcodon. '***' betyder at her kan være en stopcodon.

Spørgsmålet er om aminosyresekvensen er den rigtige, for en dobbeltstrenget DNA-streng kan som nævnt i linket om codons og læserammer aflæses i seks forskellige læserammer.

2. Prøv nu at translaterere sekvensen med hensyn til den mest almindelige startcodon ATG. Dette vil ikke ændre læserammen, dvs. translationen vil stadig starte på position 1 på den positive DNA-streng. Derimod vil translationen give det længste protein, der har ATG som startcodon, og ikke translaterere hele sekvensen som før.

- Du går tilbage med pilen øverst til venstre.
- Til denne søgning skal feltet 'OFR finder' ændres til 'Start codon: strict', klik derefter på 'submit query'.
- Hvor langt er dette protein?
- Vil du antage at dette er det korrekte protein?

Ved at vælge 'Start codon: strict' kan man nemt visualisere den længste aminosyresekvens (mellem start- og stopcodon) fra en given læseramme. Som regel vil man således hurtigt kunne danne sig et overblik over hvorvidt det er en korrekt translation, da en peptidsekvens skal have en vis længde for at udgøre et funktionelt protein.

3. Prøv nu at translaterere sekvensen med strict startcodon på alle seks læserammer. Ved at kombinere alle seks læserammer og have en strict startcodon vil Virtual Ribosome finde den længste translaterede sekvens, der har ATG som startcodon på tværs af alle læserammerne.

- Til denne søgning skal feltet 'Reading frame' ændres til 'All (6 reading frames)', klik derefter på 'submit query'.
- Hvor langt er dette protein, og hvilken læseramme er blevet brugt til translationen?
- Som nævnt finder Virtual Ribosome den længst translaterede sekvens, men kan det altid antages at denne translation giver det korrekte protein? Hvorfor/hvorfor ikke?

B. UniProt

Formålet med denne øvelse er at lære at søge efter og gøre sig bekendt med proteiner ud fra den information, der er tilgængelig i UniProt. Dette gøres ved at benytte almene fritekst søgekriterier eller accession-numre og derefter undersøge UniProt-siderne for de proteiner der findes gennem søgningerne.

Åbn programmet her: <http://www.uniprot.org/>

1. Brug det almindelig søgefelt i UniProt til at søge efter det ønskede protein når du ved at:
 - a. Det er spytamylase (salivary amylase)
 - b. Det skal komme fra mennesker (human)
 - c. Det sandsynligvis er 511 aminosyrer langt (information fra Virtual Ribosome)

Hvert protein har et accession-nummer, og du er sandsynligvis nået frem til at spytamylase fra mennesker har nummeret P04746.

2. Klik på dette nummer og du kommer til UniProt-siden for det givne protein.
3. Dan dig et overblik over UniProt-siden for proteinet og find ud af hvilke informationer man kan få om proteinet ved at besvare følgende spørgsmål:
 - a. Hvad er proteinets anbefalede navn (eng. recommended name) og dets gennavn (navnet på det gen der koder for proteinet)?
 - b. Hvor lang er sekvensen angivet i aminosyrer?
 - c. Hvad er dets funktion og subcellulære lokation (det område i cellen hvor proteinet fungerer)?

C. PDB (protein databank) og Protein Workshop

Formålet med øvelsen er at lære at finde relevant information om et protein gennem dets PDB-side så de rette dele af det kan observeres gennem tredimensionelle visualiseringer i Protein Workshop.

Gå til PDB databasen her: <http://www.pdb.org/pdb/home/home.do>

1. Søg efter proteinet med id 1SMD som gerne skulle være det spytamylase der minder mest om det protein, du har undersøgt hidtil.
 - a. Hvilke oplysninger taler for, at det er det rigtige protein du har fat i?
2. Undersøg nu proteinets 3-D struktur ved hjælp af Protein Workshop.
 - a. Find linket til 'Protein Workshop' under billedet af proteinet til højre på siden. Måske har du brug for at opdatere din version af programmet Java, men det skal du så bare gøre.
 - b. Tag fat med musen i proteinet og drej det rundet. Hvilken form har molekylet? (tertiær struktur).
 - c. Gå ind under 'Shortcuts' og vælg 'chain color ramp' og tryk på knappen 'Enact'. Identificer Den N-terminale og C-terminale del af proteinet.
 - d. Undersøg dets sekundære struktur ved at skifte til 'Conformation Type'. Husk at trykke 'Enact'. Hvor mange α -helixer og hvor mange β -foldeblade? Gå evt. tilbage i PDB under 'Molecular Description' og tjek om dit antal stemmer overens med opgørelsen der.
 - e. Undersøg proteinets polaritetsforhold ved at skifte til 'Hydrophobicity'. Husk at trykke 'Enact'. Passer fordelingen af hydrofile og hydrofobe aminosyrer med at proteinet udskilles i mundhulen?
 - f. Undersøg proteinets aminosyresammensætning ved at skifte til 'By compound'. Husk at trykke 'Enact'. Kom med forslag til hvor der kan være disulfidbindinger. Hvor mange disulfidbindinger vil du foreslå at proteinet har?