

Vejledning til brug af AMPept 1.0 serveren

Formål

Formålet med denne øvelse er at finde et antimikrobielt peptid (AMP) i en ukendt DNA sekvens. Derudover er det mere overordnede formål, at få en indsigt i hvordan man som forsker kan bruge bioinformatiske redskaber til at finde fremtidens medicin.

Indledning

1. Gå ind på www.biotechacademy.dk og find AMP projektet.
2. Læs indholdet om bioinformatik og teorien.
3. Gå ind under "Opgaven" og hent den DNA sekvens der svare til det nummer du har fået fra din lærer.

Analyse

1. Kopier din DNA sekvens fra hjemmesiden og over i et tekstdokument. Dette er dit udgangspunkt. Gem dokumentet på computeren under et navn du kan huske. Du skal i den nedenstående øvelse kopiere en række informationer over i dette dokument, og bruge det når du skal skrive din rapport. Så vær omhyggelig med at skriv hvad det er du kopiere ind, så du kan huske det.

Når man skal kopier tekst (f.eks. lange sekvenser af ATCG...) mellem dokumenter så kan man med fordel markere det man vil kopiere og trykke Ctrl+c. Så klikker man der hvor man vil have sin tekst over i og trykker Ctrl+v. Denne genvej vil kunne spare dig meget tid i den efterfølgende analyse.

2. GeneFinder serveren

Dette program finder mulige gener i DNA sekvenser. Det vil sige områder i DNA'et der potentielt kan oversættes til protein, idet de starter med et start-codon og slutter med et stop-codon disse områder kaldes åbne læserammer, da det er disse områder ribosomerne senere kan translaterer til protein. Det er de områder der kan "læses" af ribosomerne.

- a. Kopier hele din DNA sekvens over i vinduet på GeneFinder serveren og klik på "søg". Der er mulighed for at indstille hvor mange baser man mindst ønsker ens gen har. Som standard skal den stå på 100, men hvis du har tid og lyst, så prøv at se hvad der sker hvis du ændre dette tal til 50 eller 150. Kan du forklare dette?
- b. Resultatet kopieres over i dit tekstdokument. Et eksempel på et resultat er givet nedenfor. Læg mærke til at din DNA sekvens nu er oversat til en aminosyresekvens, hvor aminosyrerne er repræsenteret med en et bog-

stavskode. I videnskabelige artikler skrives proteinsekvenser altid med denne et bogstavskode. (se artiklen om proteinstruktur)

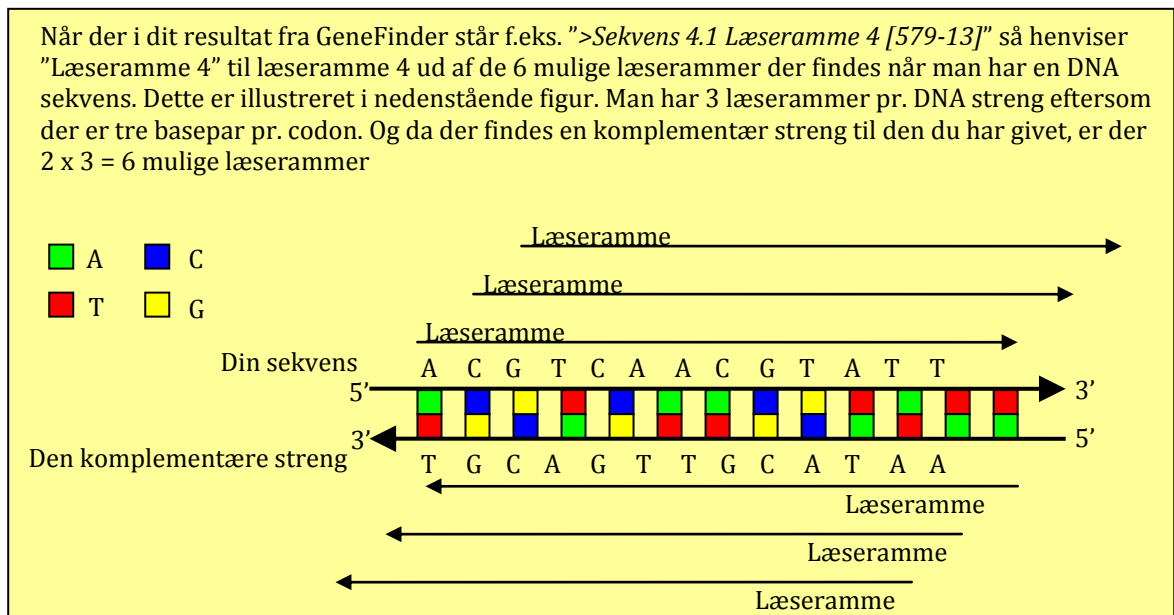
>Sekvens 4.1 Læseramme 4 [579-13]

```
LGTSERPRRRASWVSTVTLNRLAVIVALHIILIPRTIASKSTFLKLVRIHTSKIVRMSLG  
IRNSVSLGNRLGCKKSGVEKSEGDGTDGEDGGKLCLEAMTRTEPFSLKIKCVHVESLP  
CFLTTTLEIGLVIFTWAYSTSSHVAPEVAAIRVTFSASTEGGVSHNGDSCCRCRVRSPAS  
ACYIKRAVV
```

>Sekvens 2.1 Læseramme 2 [155-601]

```
LLVLYAHVKITSPISNVVVKKHGRLSTWTHLIFKLGSVRVIASRQC�LPSSPSVSPSS  
DFSTPEPLQHPSLFPRLLTFLIPRLILTILLVWMRTNFRNVLDAMVLGMRMICSATITA  
SLLRVTVDTQLARLLGRSEVPSYCCRGHL
```

Osv.



3. Klik videre til BLAST ude til venstre

4. BLAST serveren

BLAST serveren leder efter om der er gener som andre forskere har fundet som ligner dine mulige gener. På den måde kan man enten finde ud af at det gen man har fundet, allerede er kendt, eller man kan finde ud af hvilke kendte gener der ligner det gen man har fundet. Det er trods alt lettere at vide at ens gen er 90 % ligesom et kendt gen, da man så har et udgangspunkt for sin videre forskning.

- Kopier hele dit resultat fra GeneFinder over i vinduet på BLAST serveren og klik på "søg".
- Da BLAST serveren leder blandt 100 millioner kendte gener i en enorm database, tager det mellem 30 sekunder og 2 minutter for at færdiggøre dette trin. Så hav tålmodighed.

- c. Resultatet kopieres over i dit tekstdokument. Et eksempel på et resultat er givet nedenfor

Query id	% identity	alignment length	q. start	q. end	e-value
Sekvens_1.5	32.88	73	68	138	0.287
	SLKHRLATWTRHPGLTSRSPTRVFLDPKASMRFRPRTLIGFHIHKFRNRIDRGTKKEAVP RYSRPMSLMRRFIQTYLFVYALPLSASRANVEGVLRRLRYLRMTPQLVYAPADPSYDVALLT KIFHIDSNDLSLDIHEGKER				
Sekvens_5.8	Ingen hits blev fundet!				
Sekvens_3.4	80.00	135	1	135	4.5e-55
	METQRASLCLGRWSLWLLLLALVPSASAQALSREAVLRAVDRLNEQSSEANLYRLEL DQPPKADEDPGTPKPVSTFKETVCPRPTRQPELCDFKENGVRVKQCVGTVTLQIKDPL DITCNEVQGVRRGLCYCRRRRCVCGRG				

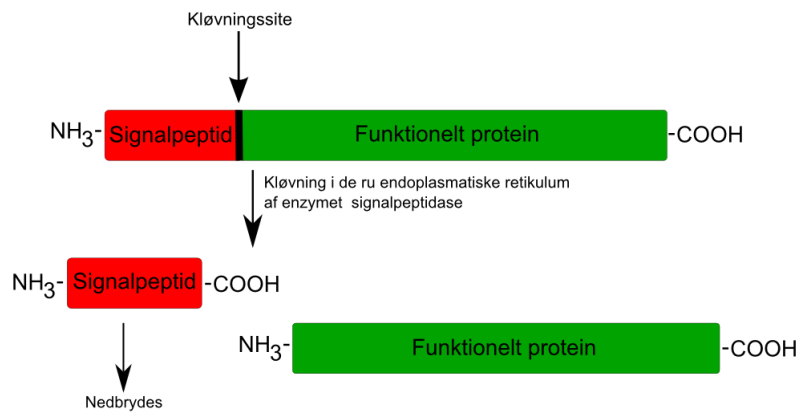
Query id	Dette er navnet på sekvensen. Det er det samme som resultatet i GeneFinder. (f.eks. Sekvens_1.5)
% Identity	Dette beskriver hvor mange % overensstemmelse i aminosyresekvensen der er mellem dit mulige gen og det gen der bedst ligner dette bedst. Jo større den er, desto bedre match er der mellem de to gener.
	Gen fra databasen KMIQTRANRLYR $\frac{7}{12} * 100 = \underline{\underline{58,3\%}}$ Output EAVPRYSRPMS LMILTSANRYLR
Alignment length	Beskriver hvor langt det gen BLAST har fundet der ligner det mulige gen der er i din sekvens er.
q. start	Svarer til den første sorte aminosyre i dit output. Det er den sorte sekvens der passer til det gen som BLAST har fundet i databasen. Den røde sekvens er "overflødig" og matcher ikke noget, så det røde stykke skal klippes væk på det hit i arbejder videre med i SignalP! HUSK DET.
q. end	Slutningen af det gen der ligner.
	<div style="text-align: center;"> <p>q. start = 11 q. end = 49</p> <p>↓ ↓</p> <p>Gen fra databasen LMRRFIQTYLFVYALPLSASRANVEGVLRRLRYLRMTPQL</p> <p>Output EAVPRYSRPMSLMRRFIQTYLFVYALPLSASRANVEGVLRRLRYLRMTPQL</p> <p style="text-align: center;"> ----- </p> <p style="text-align: center;">Alignment length = 38</p> </div>
e-value	E-værdien beskriver sandsynligheden for at det match BLAST har fundet er en tilfældighed og dermed forkert. Med andre ord kan man aldrig være 100% sikker på at det BLAST finder er 100% rigtigt, men det kan være så vanvittigt usandsynligt at det er forkert (e-value f.eks. = $4.5 * 10^{-55}$) at man godt sætte sine penge på at det er rigtigt (ville du spille lotto hvis der var $4.5 * 10^{-55}$ % chance for at du IKKE vandt?)

- d. På baggrund af "% Identity" og "e-value" vælges det bedste hit. Det antages at dette er det eneste AMP i din sekvens.

5. Klik videre til SignalP ude til venstre

6. SignalP serveren

Inden man kan forudsige strukturen af et protein skal man have klippet signalpeptidet af proteinsekvensen. Dette sker når proteinet translateres ind i ruER. Se "Signalpeptid" artiklen for mere. SignalP er et program der kan bruges til at finde signalpeptider, og det kløvningsite hvor signalpeptidet klippes fra resten af det antimikrobielle peptid som du har fundet ovenfor.






- Klip den røde del fra den åbne læseramme som du fik størst tillid til var et gen efter BLAST søgningen. Gør dette i dit tekstdokument i et nyt afsnit (kopiere genet fra BLAST resultatet uden af ændre på dette). Se forklaring under "q. start" i tabellen ovenfor. Det er denne "rest" som du skal arbejde videre med.
- Du har nu med stor sandsynlighed fundet frem til aminosyresekvensen af dit antimikrobielle peptid. Det handler nu om at identificere det signalpeptid der sidder på dit AMP. Dette skal klippes af inden du kan forudsige strukturen af dit funktionelle protein (dit AMP).
- Kopier dit AMP ind i vinduet på SignalP serveren og noter dig med hvilke parametre (organisme og ende af peptid hvor der skal søges) du har indstillet SignalP inden på klikker på søg. Der er kun en kombination af indstillingerne der er rigtig. Du kan finde svarene i teorimaterialet samt figuren ovenfor. Skriv i din rapport hvilke indstillinger du valgte, og forklar hvorfor du valgte disse.
- I dit resultat vil du se to grafer. Den øverste er udregnet med et neuralt netværk (SignalP-NN) og den nederste med en statisk model (SignalP-HMM). To forskellige metoder der tilsammen skal gøre dig i stand til at udvælge det mest sandsynlige kløvningsite for signal peptidase enzymet.

I. SignalP-NN



På denne graf er der en kombination af to grafer (grøn & blå) og en masse søjler (røde). Se forklaringer til disse nedenfor. Højreklik på billedet af grafen og vælg "kopier". Gå over i dit tekstdo-

kument og klik Ctrl+v hvorefter grafen gerne skulle komme frem. Se forklaring til grafen nedenfor.

Forklaring til SignalP-NN	
	Muligt kløvningssite før aminosyren. Jo højere, desto større sandsynlighed for et kløvningssite
	Muligt kløvningssite før aminosyren. Jo højere, desto større sandsynlighed for et kløvningssite
	Mulig signalpeptid. Jo højere, desto større sandsynlighed for at aminosyren er en del af signalpeptidet der skal klippes fra.

II. SignalP-HMM

På denne graf er der en kombination af tre grafer (grøn, blå & turkis) og en masse søjler (røde). Se forklaringer til disse nedenfor. Kopier grafen over i dit tekstdokument ligesom ovenfor.

Forklaring til SignalP-HMM	
	Muligt kløvningssite før aminosyren. Jo højere, desto større sandsynlighed for et kløvningssite
	Forskellige regioner af signalpeptidet. Jo højere graferne er jo større sandsynlighed for at der er et signalpeptid.

- e. Benyt de to grafer og den tilsvarende tekst på resultatsiden (om hvor SignalP-NN og SignalP-HMM forudsiger signalpeptidet og kløvningssitet) til at udvælge dig et kløvningssite. Alt **FØR** kløvningssitet er altså dit signalpeptid. Dette signalpeptid skal slettes inden du går videre til CPH Models.

7. Klik videre til CPH Models ude til venstre

8. CHP Models serveren

CPH Models er en program der kan forudsige strukturen af proteiner, alene ud fra en proteinsekvens. Da signalpeptidet bliver klippet af det funktionelle protein i det ruER, er det vigtigt du også har gjort dette inden du går i gang med CPH Models.

- a. Kopier resultatet (uden signalpeptid) fra SignalP over i vinduet på CPH Models serveren og klik på "søg".
- b. Husk at kopier billedet over i dit tekstdokument.
- c. Nedenfor billedet af strukturen på dit peptid er der er 3D model af peptider. Her kan du dreje det rundt og kigge på det fra alle vinkler.