

27611 Eksamen Sommer 2008

Dette sæt indeholder 10 opgaver.

En online version af opgavesættet vil være tilgængeligt fra kursets lektionsplan under selve eksamen (2. juni 2008 klokken 15:00-19:00). DNA/Protein sekvenser kan kopieres direkte herfra - det er *ikke* meningen at sekvenserne skal tages ind i hånden.

Lektionsplan: <http://www.cbs.dtu.dk/dtucourse/27611spring2008/lektionsplan.php>

Svar til opgavesættet skal skrives enten i rå tekst (fx i Notepad/Wordpad) eller i Microsoft Word (.doc) format.

Svaret skal uploades på CampusNet under kursus 27611 (under "Afleveringer -> Eksamen 2008"). **Husk at gemme seneste version af dokumentet inden du uploader svaret.** Når du afleverer får du en kode som skal skrives i feltet "Afleveringskode" nedenfor.

Dit studienummer skal fremgå af filnavnet (fx. s022717.doc eller s022717.txt) og skal også stå i starten af dokumentet (fx: "Studienummer: s022717")

Udfyld denne forside og aflever den til eksamensvagten. Lokalnummer og computernummer skal udfyldes med henblik på kontrol af netværkstrafikken.

Navn: _____

Studienummer: _____

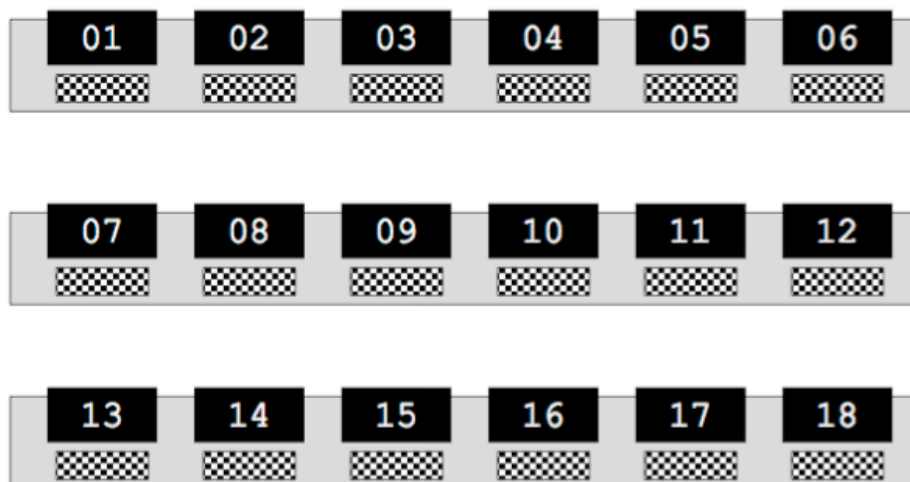
Afleveringskode: _____

Lokalnummer: _____

Computernummer: _____

(For eksaminander i lokale 042-048, byg. 210 - skriv nummeret på låget af den bærbare computer.
For eksaminander i lokale 052 og 152 i byg. 210, brug oversigten herunder):

Tavle



Indgang

Hvad gør man hvis en web-server ikke virker:

1. Verificer at input-data er i korrekt format. Forkert inputdata er i næsten alle tilfælde årsagen til problemet.
2. Prøv evt. at finde en alternativ server med samme funktion (Google).
3. Rapportér fejlen til eksamensvagten - den kursusansvarlige vil så blive tilkaldt.

Opgave 1 (15%):

Efter en farefuld færd i Sydamerikas jungler er det lykkedes dig at hjembringe knoglerne fra en hidtil ukendt dinosaur-art. Fra knoglerne har du isoleret et lille stykke DNA som du er interesseret i at undersøge nærmere, dels for at finde ud af hvad DNAets funktion er, dels for at lære mere om den mystiske dinosaur (og derved på sigt blive rig og berømt og få råd til et hus på Bahamas).

Du skal i dit svar forklare hvilke værktøjer du har benyttet for at komme frem til svaret, samt hvorfor du benyttede netop dem.

Du vil først og fremmest gerne vide noget om DNAets funktion.

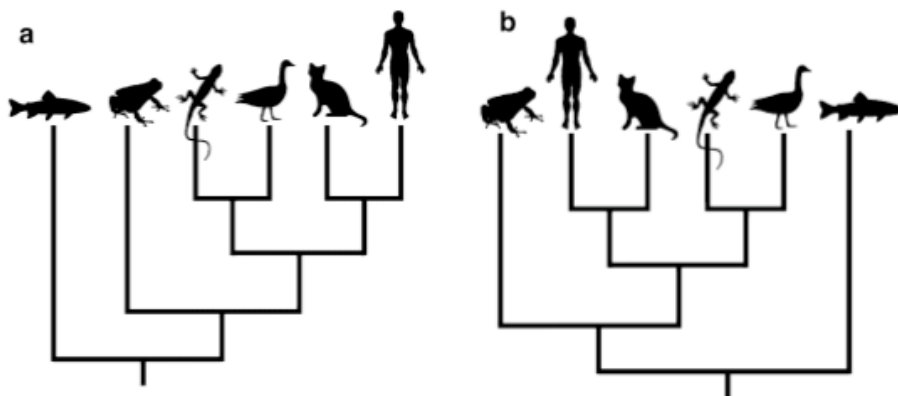
1. Er DNA-fragmentet protein-kodende?
2. Hvad er den sandsynlige funktion af dette stykke DNA (eller evt. det protein det koder for)?
3. Du vil også gerne vide noget om hvilken slags nulevende dyr den afdøde dinosaur mindede om. Det er ret velunderbygget at dinosaurer er tæt beslægtede med fugle, og man ville umiddelbart forvente at i et fylogenetisk træ ville dinosaurer ligge et sted tæt på fugle og krokodiller.

Hvilken database-sekvens er tættest beslægtet med dit ukendte DNA-fragment, og hvilken organisme stammer den fra? (Den laborant som ekstraherede DNA-fragmentet for dig er meget, meget glad for skinke-sandwich. Med den information in mente: Tror du der kan være sket en fejl under arbejdet med dit dinosaur-DNA?)

```
>ukendt_sekvens
GAAGGATGGCCAGTACCTTCTCAAATTACTAACTGGCCGCAATGCTTCTTTGTTATTTGCCACCTTGGG
CACCGGTGCCCTGACCACCGGTTACTTGATGAATCAGCGGAGCGTGTGTGCTGAGGCCCGGGAGCAGCAC
AGGCTGTTCCTCCGCAAGCGCAGACTACCCTGATCTACGCAAGCACAACAACACTGCATGGCCGAGTGC
```

Opgave 2 (5%):

På basis af to forskellige DNA-datasæt, fra to forskellige gener, har du konstrueret de følgende fylogenetiske træer. Viser træerne den samme evolutionære historie eller er der forskelle?



Opgave 3 (10%):

Du har indsamlet en række relaterede enzym-sekvenser fra forskellige organismer (se herunder), og er nu interesseret i at undersøge i hvor høj grad de ligner hinanden.

1. Lav et multiple alignment af sekvenserne. Find den længste, helt konserverede blok af aminosyrer og angiv hvilke positioner, målt ud fra alignment-nummereringen, denne blok dækker.
2. Hvis man istedet nummererer ud fra sekvensen med navnet NP_067263, hvad er så positionen

af den konserverede blok?

>NP_067263
MLVVGSELQSDAQQLSAEAPRHGELQYLQVEHILRCGFKKEDRTGTGTLSVFGMQARYSLRDEFPLLT
KRVFWKGVLEELLWFIKGSTNAKELSSKGVRIWDANGSRDFLDSLGF SARQEGDLGPVYGFQWRHFGAEY
KMDSDYSGQGVQDQKVIDTIKTNPDDRRRIIMCAWNPKDLPLMALPPCHALCQFYVNVNSELSCQLYQRS
GDMGLGVFPFNIAASYALLTYMIAHITGLQPGDFVHTLGDHAHYLNHIEPLKIQLQREPRPFPKLIKLRKVE
TIDDFKVEDFQIEGYNPHPTIKMEMAV

>NP_001062
MPVÅGSELPRRPLPPAAQERDAEPRPPHGELQYLGQIQHILRCGVRKDDRTGTGTLSVFGMQARYSLRDE
FPLLTTRKRVFWKGVLEELLWFIKGSTNAKELSSKGVRIWDANGSRDFLDSLGFSTREEDLGPVYGFQWR
HFGAEYRDMESDYSGQGVQDQKVIDTIKTNPDDRRRIIMCAWNPRLPLMALPPCHALCQFYVNVNSELSC
QLYQRSGDMGLGVFPFNIAASYALLTYMIAHITGLKPGDFIHTLGDHAHYLNHIEPLKIQLQREPRPFPKLR
ILRKVEKIDDFKAEDFQIEGYNPHPTIKMEMAV

>NP_571835
MPDPTAVEVTNGHCVNGESKKEETNGGKKEFSLFCDERGYLSLVEFIIQNGAKKGDRTGTGVI SVFGTQAR
YSLRDQFPLLTTRKRVFWKGVLEELLWFIKGSTNAKDLSEKGVRIWDANGSREFLDKNGFTDREEDLGPV
YGFQWRHFGAEYKDMHTDYSGEQVQDQKVIDTIKSNPEDRRRIIMCAWNPKDLPLMALPPCHALCQFYV
NGELSCQLYQRSGDIGLVFPFNIAASYALLTYMIAHITGLKPGDFVHTLGDHAHYTNHIEPLKEQIQREPR
PFPKLRIRKRVQINDFCAEDFEIYDYDPHPTIKMQMAV

>NP_001096659
MPVISESTAASSCEPQAGKAENRDELQYLDQIKYILEHGRKEDRTGTGTVSVFGMQARYSLRDQFPLLT
TTRKRVFWKGVLEELLWFVKGCTNSKELSAKGVKIWDANGSREFLDKQGFSTREEDLGPVYGFQWRHFGA
EYKDTHTDYSQGVQDQKVIDTIKNNPDDRRRIIMCSWNPKDISLMALPPCHALCQFYVVDKELSCQLYQ
RSGDMGLGVFPFNIAASYALLTYMIAHVTGLKPGDFVHTLGDHAHYLNHMEPLKIQLQRSRPPFPKLIKLRQ
VENIDDFTADDFLIEGYDPHPTIKMEMAV

>NP_491532
MNKENIADAPSDVVKTVQQQVHLNQDEYKYLKQVEQILREGTRRDDRTGTGTISIFGMQSKYCLRNGTI
PLLTTRKRVYWKGVLEELLWFISGSTDGKLLMEKNVKIWEKNGDRAFLDNLGFTSREEDLGPVYGFQWRH
FGAKYVDCHTDYSQGVQDQAEVIRQIKEQPDSTRRIIMSAWNPDLGQMVLPCHTMCQFYVDNSELSCQ
LYQRSGDMGLGVFPFNIAASYGLLTHMIAKVCGLKPGTLVHTLGDHAHVSNHVDALKIQLDREPYAFP KIRF
TRDVASIDDFTSDMIALLDYYKCHPKIPMDMAV

>NP_477367
MVLTPTKDGPDOESMPLPADNGESPQKQAPVNRDEMHYLDLLRHI IANGEQRMDRTEVGTLSVFGSQMR
FDMRNSFPLLTTRKRVFRAVAEELLWFVAGKTDAKLLQAKNVHIWDGNSREFLDKMGFTGRAVGDLDGPV
YGFQWRHFGAQYGTCDDDYSKGIQDQKVIDTIRNNPDDRRRIIMSAWNPDLIPKMLPCHCLAQFYVVS
EKRELSQLYQRSADMGLGVFPFNIAASYALLTHMIAHVTGLKPGDFVHTMGDTHVYLNHVEPLKEQLERT
PRPFPKLIKRVQDIEDFRFEDFQIVDYNPHPKIQMDMAV

>NP_179230
MATTTLNDSVTTTLASEPQRTYQVVVAATKEMGIGKDGKLPWNLPDCLKFFKIDITLTTSDSSKKNVVMG
RKTWESIPIKYRPLSGRLNVVLTSSGGFDIANTENVVTCSSVDSALDLLAAPPYCLSIERVFVIGGGDIL
REALNRPCDAIHLTEIDTSDVCDTFIPADITSVYQWSSSFPVTENGLRFCFTTFVRVKSSADESSDES
NGSOSLQFDGKFLFLPKMVFDQHEEFLYLNVEDIISNGNVKNDRTGTGTLSKFGCOMKFNLRSPFPLL
TTRKRVFWRGVVEELLWFISGSTNAKVLQEKGIHIWDGNASREYLDGIGLTEREEDLGPVYGFQWRHFGA
KYTDMHADYTGQGFQDQKVIDTIKNNPDDRRRIIMSAWNPDLKMLALPPCHMFAQFYVAEGELSCQMYQ
RSADMGLGVFPFNIAASYLLTCMLAHVCDLVPDFIHLVGDHAHVYKTHVRPLQEQLLNLPKPFVPMKINPE
KKQIDSFVASDFDLTGYDPHKKIEMKMAV

>NP_014717
MTMDGKNKEEQYLDLCKRIIDEGEFRPDRGTGTLSLFAAPPQLRFLSRDDTFPLLTTKKVFTRGIILEL
LWFLAGDTDANLLSEQGVKIWDGNSREYLDKMGFKDRKVGDLGPVYGFQWRHFGAKYKTCDDDYTGQGI
DQLKQVIHKLKTNPYDRRIIMSAWNPADFDKMLALPPCHIFSQFYVVSFPKEGEGSGKPRLSCLLYQRS
GLGVFPFNIAASYALLTRMIAKVVDMEPEGFIHTLGDHAHVYKDHIDALKEQITRNPRPFPKLIKIRVDKID
DFKLTDFEIEDYNPHPRIQMKMSV

Opgave 4 (10%):

Du har to sekvenser du gerne vil have alignet, og overvejer nu hvilket af to følgende alignments der er det bedste.

1. Udregn alignment score for alignment 1 og 2 givet det nedenstående scoringssystem (gap scoring system + substitutionsmatrix)
2. Hvilket alignment er bedst givet dette scoringssystem?

Alignment 1:	Alignment 2:
KLMNRVATQPTWPR	KLMNRVATQP-TWP-R
: : : : : :	: : : : : :
KVMNRVGTWP--VR	K----VMNRVGTWPVR

Affint gap-scoringssystem:
Gap opening: -4
Gap elongation: -1

2. På position 5 i de 15 peptider er der udelukkende N. Forklar hvorfor logo'et ranker D højere end E?
3. Ud over de 15 peptider ovenfor, har I også fået et sæt data med blandede peptider. Hvilket af nedenstående peptider vil sandsynligvis IKKE kunne binde til jeres MHC molekyle?

```
AAFFNKTEF  
AELSLFTTE  
YAPVSPIVI
```

Opgave 7 (10%)

Du skal brug BLAST og PSI-BLAST til at lære mere om funktion og struktur af følgende protein. (Hvis PSI-BLAST kommer op med en fejlmeddelelse i rødt, skal du bare fortsætte. Det betyder ingenting.)

Du skal i dit svar forklare hvilke værktøjer du har benyttet for at komme frem til svaret, samt hvorfor du benyttede netop dem.

```
>gi|21495178|ref|NP_659802.1| hypothetical protein LOC221786 [Homo sapiens]  
MTPESRDTTDLSPGGTQEMEGIVIVKVEEEDHDFQKERNKVESSPQVLSRSTTMNERALLSSYLVA  
YR  
VAKEKMAHTAAEKIILPACMDMVRTIFDDKSADKLRTIPLSDNTISRRICTIAKHLEAMLITR  
LQSGIDF  
AIQLDESTDIASCPTLLVYRVYVWQDDFVEDLLCCLNLSHITGLDLFTELENCLLGOYKLN  
WKHKCGIS  
SDGTANMTGKHSRLTEKLEATHNNAVWNHCFIHREALVSKEISPSLMDVLKNAVKT  
VNFVFKGSSLSNRL  
LEIFCSEIGVNHHTLLFHTFEVRWLSQGVLSRVYELRNEIYIFLVEKQSHLANIFEDDI  
WVTKLAYLSDI  
FGILNELSLKMQGKNNDIFOYLEHILGFQKTLWQARLKSNRPSYMPPTLLQHI  
EENIINEDCLKEIK  
LEILLHLTSLSQTFNYFPPEEKFESLKENIWMKDPFQNPESIELNLEPEEENEL  
LQLSSTLKNYY  
KILSLSAFWIKDKDDFPLLSRKSILLLLPFTTYLCELGFSILTRLKTKRNLNSAPDMR  
VALSSCVDP  
WKELMNRQAHPHSH
```

1. Hvad er funktionen af dette protein?
2. Hvor mange PSI-BLAST iterationer skal der køres for at finde et signifikant PDB hit (E-value < 0.005)?
3. Hvad er E-værdien af det signifikante PDB hit, og hvad er dets PDB identifikator?
4. Hvad er det mest brugte sekundærstrukturelement i denne PDB struktur?

Opgave 8 (5%):

Du har kørt en BLAST søgning med et protein, og har fået 3 hits. Hvilket af de følgende tre hits vil du vælge og hvorfor?

1. %ID=30, E-value=0.1
2. %ID=10, E-value=2
3. %ID=10, E-value=0.001

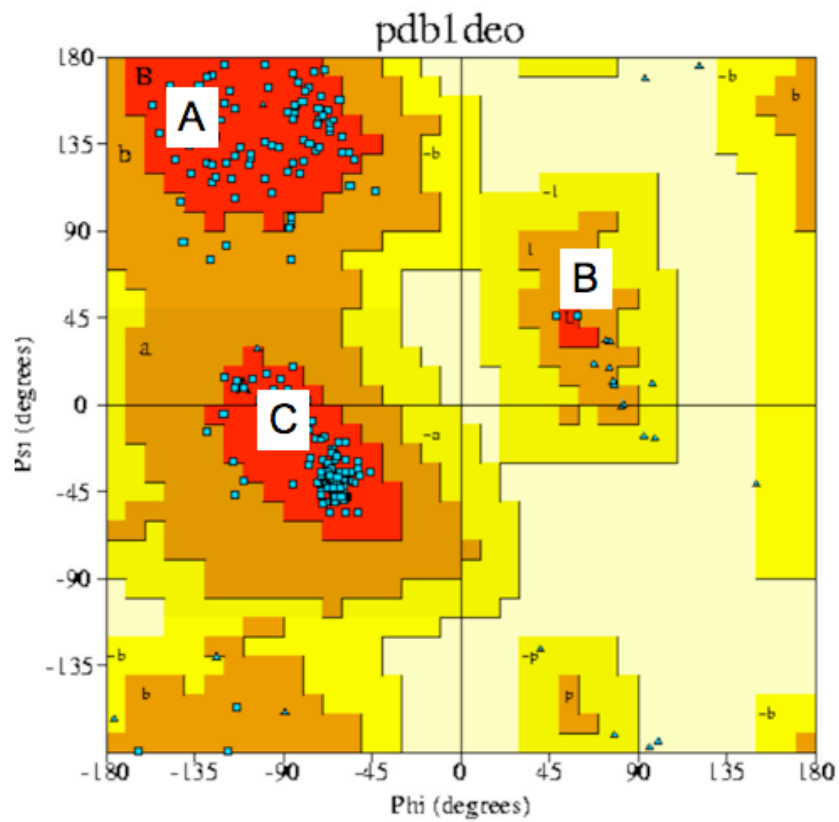
Opgave 9 (10%)

En patient er blevet indlagt på Hvidovre Hospitals afdeling for infektionssygdomme med en farlig viral infektion. I har fået virusen sekventeret, og skal nu finde ud af hvilken infektion det handler om. Du skal for hvert spørgsmål forklare hvordan du fandt frem til svaret.

1. Genomet er blevet indlagt i Genbank med følgende Genbank ID: NC_001722. Hvilken organisme stammer dette genom fra?
 2. Hvor mange baser indeholder genomet?
 3. Hvor mange proteiner koder genomet for?
 4. Hvor langt er det korteste protein og hvad hedder det?
-

Opgave 10 (10%)

Et Ramachandran plot er en metode der kan bruges til at visualisere kvaliteten af en proteinstruktur. Nedenfor er vist et Ramachandran plot for proteinet 1deo



1. Hvilke protein sekundærstrukturelementer findes primært i de tre områder A), B) og C) i figuren?
2. Beskriv kort hvordan du ville bruge Ramachandran plots til at validere kvaliteten af en protein struktur
3. Hvilke af følgende udsagn er rigtige:
 1. Beta-sheets dannes af lokale interaktioner
 2. Beta-sheets dannes af ikke-lokale interaktioner
 3. alpha-helices dannes af lokale interaktioner
 4. alpha-helices dannes af ikke-lokale interaktioner
 5. Et proteins kerne (core) er fyldt med ladede aminosyrer