

## INF4350 Deleksamen 1

Dette er den norske utgaven av oppgavene. Skulle det være uoverensstemmelser mellom den norske og den engelske utgaven av oppgavene er det den engelske utgaven som teller.

### Generell informasjon

- Denne deleksamen skal løses individuelt.
- Vedlagte erklæring "Krav til innleverte oppgaver ved Institutt for informatikk (Ifi)" skal leses og aksepteres.
- Programmene kan skrives i et valgfritt programmeringsspråk som du finner passende.
- Skriv en kort rapport med beskrivelse av din fremgangsmåte for å løse oppgavene, et sammendrag av resultatene underveis, svar på spørsmålene, samt kildekoden til alle programmene du har skrevet.
- Rapporten må leveres senest onsdag 15. oktober 2008 kl 2359 per epost til torognes@ifi.uio.no i form av ett dokument (helst PDF).

### Oppgave 1: Gensøk

- a) Skriv et program som leser inn en lang DNA-sekvens i FASTA-format fra en fil og finner alle åpne leserammer som kan kode for proteiner på minst 100 aminosyrer. Ignorer kortere leserammer. Bruk den universelle genetiske koden. Programmet skal skrive ut følgende for hver av leserammene: start-posisjon, stopp-posisjon, retning, lengde og tilhørende translaterede aminosyresekvens. Det skal også skrive ut totalt antall leserammer funnet.
- b) Hent hele genomsekvensen til bakterien *Helicobacter pylori* J99 fra GenBank. Sekvensen har accession number AE001439. Last den ned i FASTA-format. Kjør programmet på sekvensen som du hentet ned. Vis et utdrag av resultatene (f.eks. informasjonen om de 20 første leserammene du finner).
- c) Hvor mange leserammer fant du totalt?
- d) Hvor mange proteiner er kjent fra denne bakterien? Sammenlign dette med antall leserammer programmet fant og forklar bakgrunnen for eventuelle forskjeller.

### Oppgave 2: Databasesøk

- a) Et av proteinene til bakterien *H. pylori* som du kanskje fant i forrige oppgave finnes i NCBI's databaser med accession number NP\_223766. Hent denne sekvensen og lagre den i en FASTA-formatert fil. Vis denne sekvensen.
- b) Bruk NCBI BLAST på nettet på <http://blast.ncbi.nlm.nih.gov/Blast.cgi> til å søke med dette proteinet i SWISSPROT-databasen. Under "Algorithm parameters" skal du slå av "Compositional adjustment" av scoringsmatrisen ("No adjustment") og slå på filtrering av "Low complexity regions". Behold ellers alle standard parametre, dvs

scoringmatrise BLOSUM62, gapåpningsstraff på 11, og gaputvidelsesstraff på 1. Vis et utdrag av resultatene av søket.

- c) Hvor mange statistisk signifikante treff fikk du?
- d) Basert på resultatene av BLAST-søket, beskriv kort den mest sannsynlige funksjonen til proteinet.
- e) Du bør ha fått minst ett godt treff mot en sekvens fra menneske. Hva er gensymbolet og navnet til dette humane genet? Lagre sekvensen til det beste treffet fra menneske i FASTA-format i en fil.
- f) Vis sammenstillingen mellom proteinet fra bakterien og fra mennesket som du fikk fra BLAST.
- g) Hvilken score (direkte sammenstillingsscore, ikke normalisert bitscore) og E-verdi fikk denne sammenstillingen?
- h) På hvilket kromosom hos menneske er genet som koder for dette proteinet plassert?

### **Oppgave 3: Sammenstillinger**

- a) Skriv et program for å beregne optimal lokal sammenstillingsscore for to aminosyresekvenser ved hjelp av Smith-Waterman algoritmen. Programmet skal lese de to sekvensene i FASTA-format og skrive ut optimal sammenstillingsscore. Programmet skal bruke samme scoringsparametre som BLAST (scoringmatrise BLOSUM62, gapåpningsstraff på 11, og gaputvidelsesstraff på 1). Det er ikke nødvendig å lagre eller vise selve sammenstillingen, kun score.
- b) Kjør programmet ditt på proteinet fra bakterien og proteinet fra menneske fra forrige oppgave. Hvilken score får du?
- c) Forklar bakgrunnen for eventuelle forskjeller fra scoren du fikk med BLAST i forrige oppgave.



# Institutt for informatikk

## Krav til innleverte oppgaver ved Institutt for informatikk

Ved alle pålagte innleveringer av oppgaver ved Ifi – enten det dreier seg om obligatoriske oppgaver, hjemmeeksamen eller annet – forventes det at arbeidet er et resultat av studentens egen innsats. Å utgi andres arbeid for sitt eget er uetisk og kan medføre sterke reaksjoner fra Ifis side.

Derfor gjelder følgende:

1. Hvis du tar med tekst, programkode, illustrasjoner og annet som andre har laget, må du tydelig merke det og angi hvor det kommer fra.
2. Det er greit å få hint om hvorledes en oppgave kan løses, men dette skal eventuelt brukes som grunnlag for egen løsning og ikke kopieres uendret inn.
3. Kursledelsen kan innkalle studenter til samtale om deres innlevering.

## Gruppearbeid

I noen kurs skal det leveres gruppearbeid. Ifi krever da at alle medlemmer av gruppen kan gjøre rede for hovedtrekkene i det innleverte arbeidet. Dessuten må alle ha utført en rimelig del av det hele, og kunne identifisere og svare i detalj for sin del.

## Samarbeid

Reglene om kopiering betyr ikke at Ifi fraråder samarbeid – tvert imot, Ifi oppfordrer studentene til å utveksle faglige erfaringer om det meste. Men det kreves som nevnt at man kan stå inne for det som leveres.

Hvis du er i tvil om hva som er lovlig samarbeid, kan du kontakte gruppelærer eller faglærer.

[www.ifi.uio.no/studinf/skjemaer/erklaring.pdf](http://www.ifi.uio.no/studinf/skjemaer/erklaring.pdf)

27. jan. 2004