

“Fasit” for spørsmål 1-4, KJB492 H00

Oppgave 1

Figuren viser at proteinet har et repetert område fra aa 200-400 og aa400-600, samt fire repetisjoner i området aa700-800. Det ideelle besvarelse ville kanskje nevnt kort hvordan slike diagrammer fremstilles (vindu osv.).



Oppgave 2

a) Heuristisk: For problemer som er store eller umulige regnemessig sett benyttes andre typer kunnskap for å redusere omfanget. En heuristisk metode tar ”snarveier” og vil ikke nødvendigvis finne det matematisk sett optimale resultat.

b) For flersekvenssammenstillinger er tidsforbruket for en matematisk eksakt algoritme proporsjonalt med M^N hvor N er antall sekvenser og M er lengden. Med økende antall sekvenser blir det fort nærmest umulig å benytte eksakte algoritmer

c) Redusert gapstraff i strekk ≥ 5 hydrofile aa, Reusert gapstraff i områder med gap i allerede sammenstilte sekvenser, høyere gapstraff i områder nær eksisterende gap, justering av gapstraff ut fra naboaminosyren. Mange nevner høyere GOP enn GEP, men dette har ClustalW til felles med de fleste andre programmer. Kan allikevel telle litt positivt.

Oppgave 3

Lokale gap: Matrisen initieres ved å fylle 0-er i første rad og kolonne (ikke straff for endegap) og markering av poeng for det aminosyrepar hver rute tilsvarer:

		A	C	D	E	D
	0	0	0	0	0	0
D	0	-0.5	-0.5	1	-0.5	1
C	0	-0.5	1	-0.5	-0.5	-0.5
E	0	-0.5	-0.5	-0.5	1	-0.5
F	0	-0.5	-0.5	-0.5	-0.5	-0.5

Matrisen fylles så ut ved å starte med øverste venstre rute og fylle ut kolonne etter kolonne. For en lokal sammenstilling er poengsummen som utfylles det høyeste alternativet av:

- Verdien i ruten diagonalt over til venstre + initieringstallet i ruten det gjelder

- b. Verdien i ruten over ruten det gjelder + gappoeng (-0.8)
- c. Verdien i ruten til venstre for ruten det gjelder + gappoeng (-0.8)
- d. 0

Ruten den høyeste verdien kommer fra markeres med en pil

		A	C	D	E	D
	0	0	0	0	0	0
D	0	-0.5 0	-0.5 0	1 1	-0.5 0.2	1 1
C	0	-0.5 0	1 1	-0.5 0.2	-0.5 0.5	-0.5 0.2
E	0	-0.5 0	-0.5 0.2	-0.5 0.5	1 1.2	-0.5 0.4
F	0	-0.5 0	-0.5 0	-0.5 0	-0.5 0.4	-0.5 0.7

Backtrackingen starter fra ruten med den høyeste poengsum (her 1.2) og følger pilene til en rute med verdien 0 nås. For en gitt rute gjelder:

Dersom pilen går diagonalt ut av ruten sammenstilles de to aminosyrene ruten representerer.

Dersom pilen går vertikalt ut av ruten det gjelder sammenstilles aminosyren i den vertikale sekvens med et gap i den horisontale

Dersom pilen går horisontalt ut av ruten det gjelder sammenstilles aminosyren i den horisontale sekvens med et gap i den vertikale

		A	C	D	E	D
	0	0	0	0	0	0
D	0	-0.5 0	-0.5 0	1 1	-0.5 0.2	1 1
C	0	-0.5 0	1 1	-0.5 0.2	-0.5 0.5	-0.5 0.2
E	0	-0.5 0	-0.5 0.2	-0.5 0.5	1 1.2	-0.5 0.4
F	0	-0.5 0	-0.5 0	-0.5 0	-0.5 0.4	-0.5 0.7

Altså:

D-E
DCE

Og

CDE
C-E

b) Sammenstilling av nært beslektede globinsekvenser, som kunne sammenstilles manuelt uten innføring av gap. Beregning av frekvens av hver aminosyre og hver overgang $aa_1 \leftrightarrow aa_2$. Ut fra dette fremstilling av log odds tabeller for 1 PAM og tabeller for høyere PAM-verdier ved ekstrapolering. Se besvarelsen fra kandidat 5 for en mer fyldig beskrivelse enn jeg hadde ventet meg.

Oppgave 4

blastp	compares an amino acid query sequence against a protein sequence database
blastn	compares a nucleotide query sequence against a nucleotide sequence database
blastx	compares a nucleotide query sequence translated in all reading frames against a protein sequence database. You could use this option to find potential translation products of an unknown nucleotide sequence.
tblastn	compares a protein query sequence against a nucleotide sequence database dynamically translated in all reading frames. For finding specific genes in a genome etc.
tblastx	compares the six-frame translations of a nucleotide query sequence against the six-frame translations of a nucleotide sequence database. For gene finding in genomes etc.
PSI-BLAST	provides an automated, easy-to-use version of a "profile" search, which is a sensitive way to look for sequence homologues. The program first performs a gapped BLAST database search. The PSI-BLAST program uses the information from any significant alignments returned to construct a position-specific score matrix, which replaces the query sequence for the next round of database searching. PSI-BLAST may be iterated until no new significant alignments are found. For finding distant homologues