

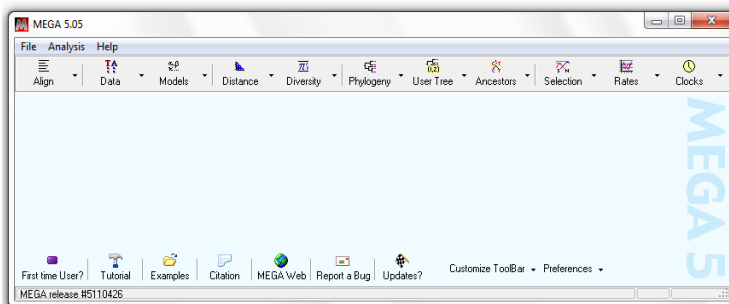


# Sekvens-alignment

Denne vejledning indeholder først en introduktion til softwareprogrammet MEGA og dernæst en vejledning der undervejs indeholder en række små øvelser og opgaver der illustrerer, hvordan man kan bruge softwareprogrammet MEGA til at lave alignments af DNA- og protein-sekvenser, både i hånden og via algoritmer.

## Introduktion til MEGA

Denne vejledning er skrevet til udgave 5.05 af programmet MEGA. Nogle elementer er en smule anderledes i beta-udgaven 5.1, så hent den gennemtestede 5.05 udgave af programmet. Programmet kan hentes på følgende link: <http://www.megasoftware.net/>. Programmet er gratis at hente og bruge til undervisning, men hver person skal hente sin egen personlige licens. Efter programmet er downloadet og installeret, åbner det til følgende startbillede:

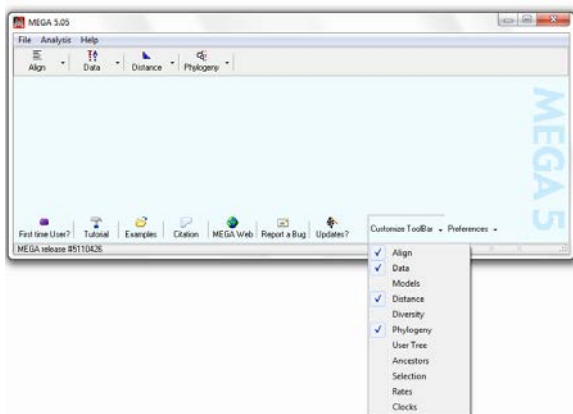


Tamura K, Peterson D, Peterson N, Stecher G, Nei M, and Kumar S (2011) **MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods.** Molecular Biology and Evolution 28: 2731-2739.

Under 'Help'-menuen findes en meget udførlig dokumentation af programmets mange muligheder, så hvis du løber ind i problemer der ikke er dækket af denne øvelse, kan du sikkert finde svar der.

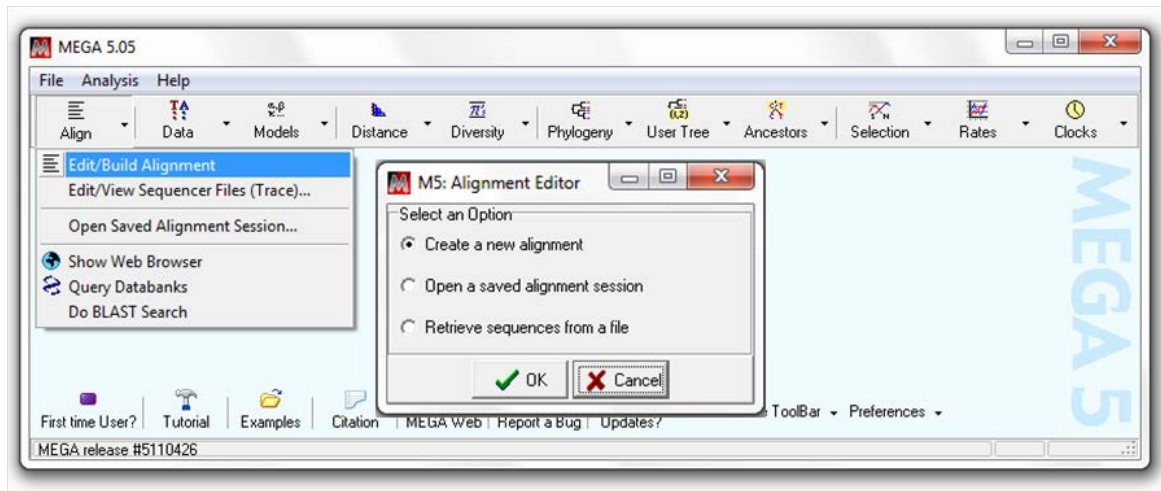
### *Fjern avancerede funktioner*

Under knappen 'Customize Toolbar' fjernes først alle de funktioner vi ikke får brug for, i forhold til stof der er gennemgået i Bioteknologi 6. De fire knapper vi får brug for er Align, Data, Distance og Phylogeny.

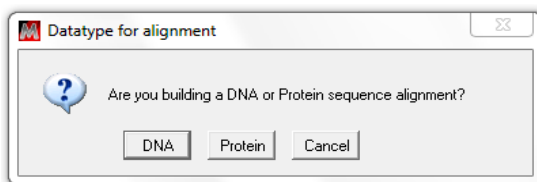


## Opbygning og alignment af sekvenser

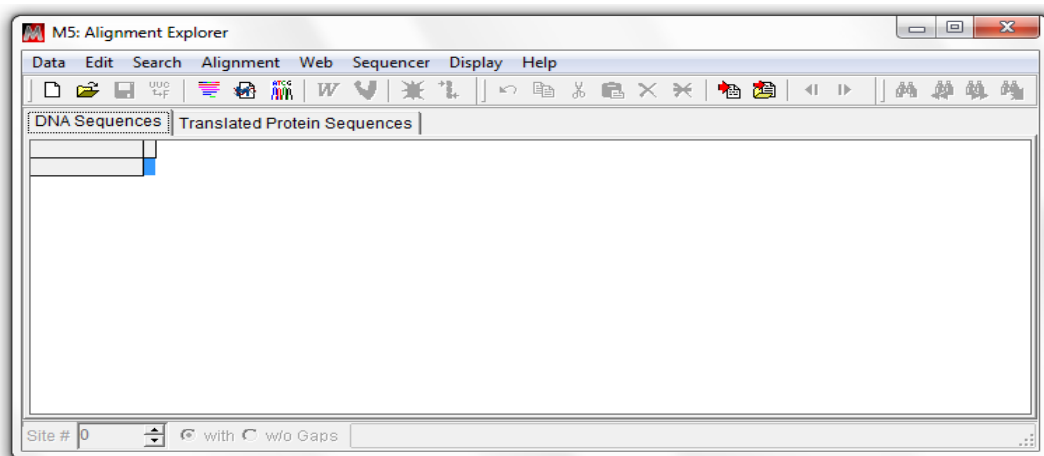
Vælg knappen 'Align' og derefter 'Edit/Build alignment'. Herefter dukker et nyt vindue op som vist herunder. Her er der tre muligheder hvoraf vi her vælger den første og opbygger en ny alignment fra bunden.



Efter at have valgt at vi vil bygge en ny alignment, skal vi vælge hvilken type data vi vil arbejde med. Vælg DNA som vist herunder:



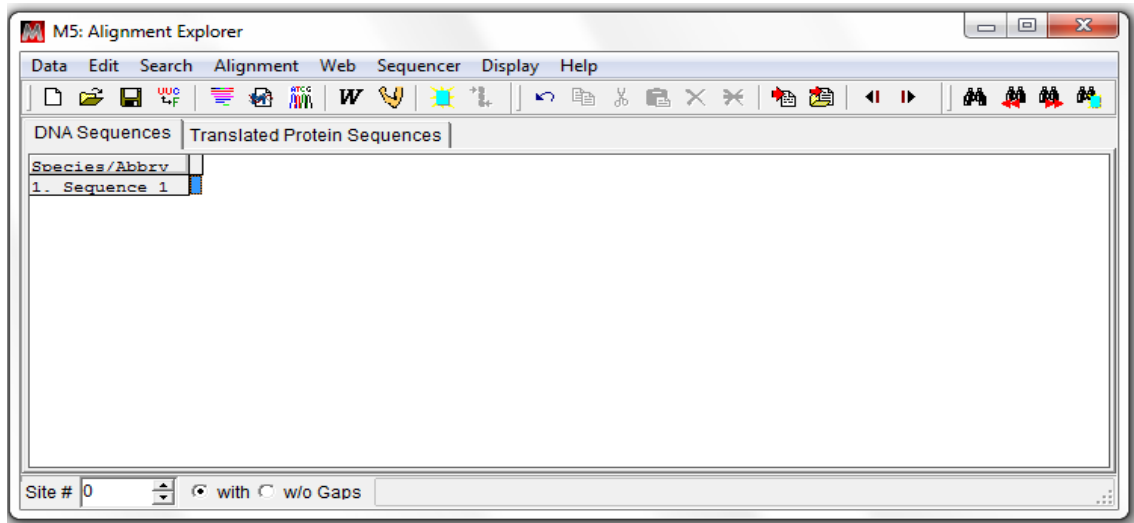
Nu åbnes et nyt vindue, den såkaldte Alignment Explorer som ser ud som vist herunder:



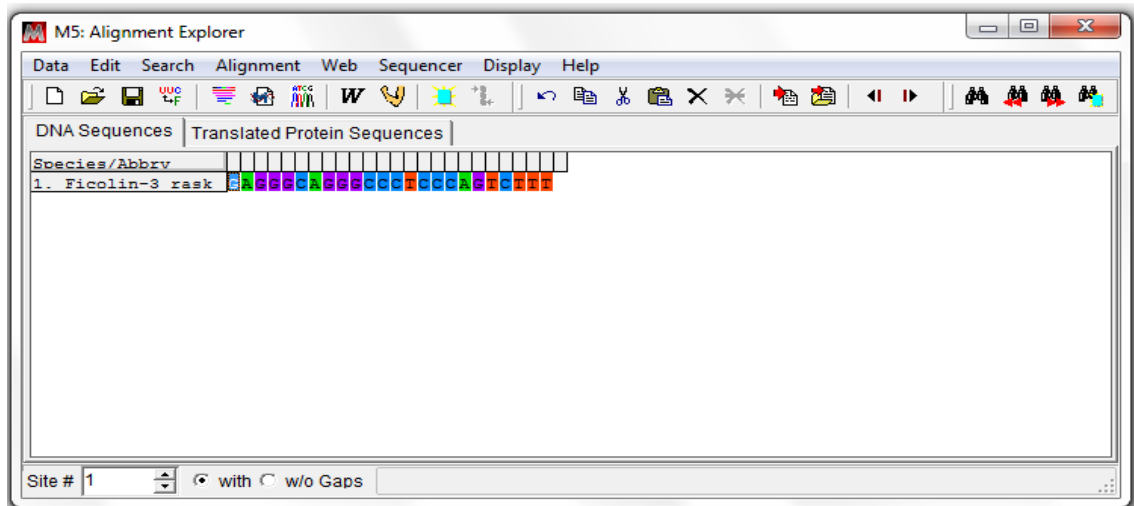
På nuværende tidspunkt har vi en hel tom side da vi ikke har indlæst eller indtastet nogle sekvenser. Her bliver vist to forskellige måder at oprette sekvenser på: 1) Manuel indtastning af data, 2) Indlæsning af sekvenser fra en fil. Et par yderligere metoder er gennemgået i link nummer 4 i Bioteknologi 6.

### Manuel indtastning af sekvenser

1. Vælg 'Edit' og 'Insert Blank Sequence' hvorefter vinduet vil se ud som vist herunder:

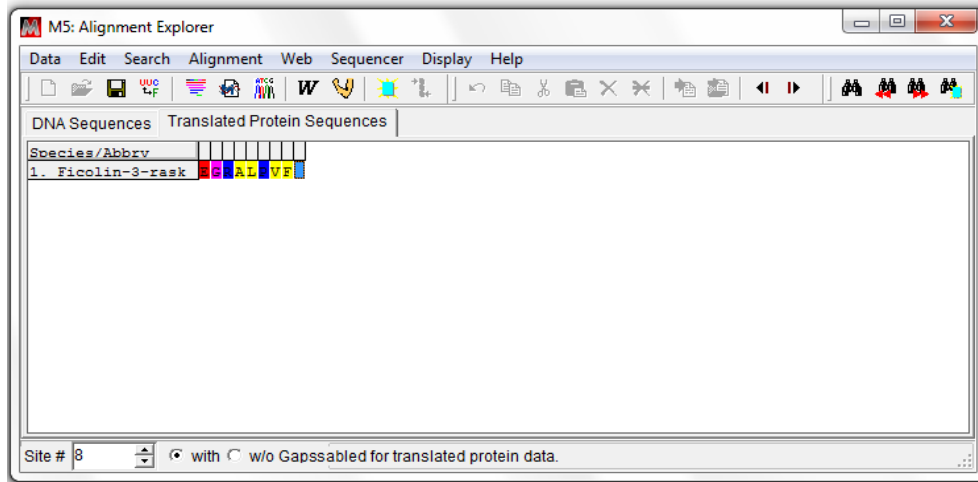


2. Højreklik på feltet hvor der står '1. Sequence 1' og vælg 'Edit Sequence Name'. Her kan man omdøbe sekvensens navn til noget mere fornuftigt. Kald denne sekvens for 'Ficolin-3-rask'.
3. Indtast følgende sekvens i det blå cursor-felt: GAGGGCAGGGCCCTCCCAGTCTTT. Nu bør dit vindue se ud som vist herunder. Læg mærke til at de fire forskellige nucleotider har hver sin farve.

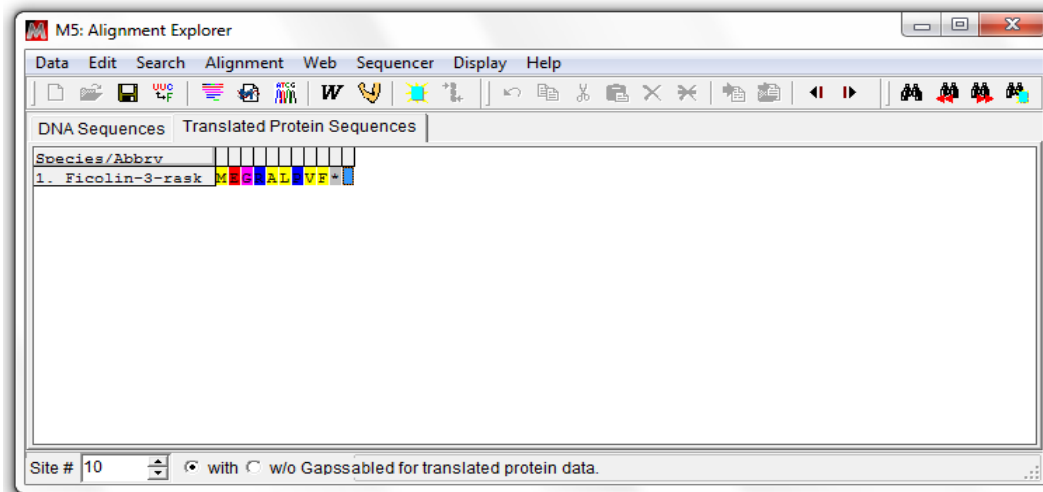


### Translation af DNA til protein

En af mange brugbare funktioner i Alignment Explorer er funktionen til at oversætte mellem DNA og protein. Før vi kan gøre dette, skal vi definere hvilken genetisk kode vi vil benytte til oversættelsen. Dette gøres under 'Data' og 'Select Genetic Code Table'. Vælg den øverste 'Standard'. Nu kan vi oversætte vores DNA-sekvens til protein ved at trykke på knappen 'Translated Protein Sequences'. Dit vindue bør nu se ud som vist herunder. Læg mærke til at der bruges et-bogstavskoder for proteinsekvenser, kan du ikke huske disse kan de findes på figur 2 på side 10 i Bioteknologi 6.



1. Opskriv det viste proteins primære struktur med korrekte aminosyrenavne.
2. Oversæt tilbage til DNA og tilføj en methionin-codon før resten af sekvensen og et stopcodon i slutningen af sekvensen.
3. Oversæt herefter til protein og tjek at du får et resultat som vist herunder:



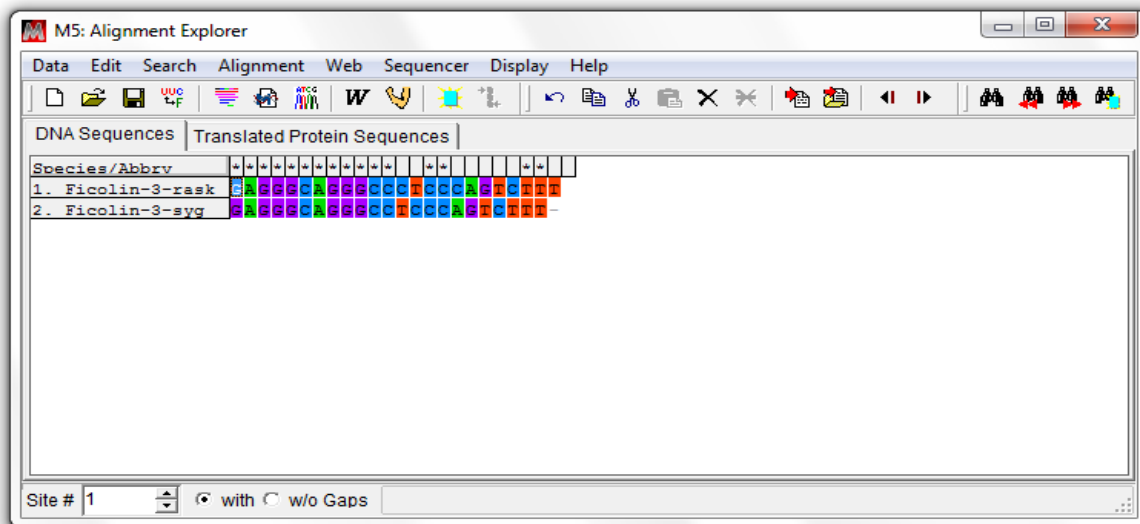
4. Oversæt tilbage til DNA og fjern de to indsatte codons igen inden du går videre med øvelsen. Bemærk at du kun kan redigere i sekvensen (ud over at indsatte mellemrum) når du ser på DNA-sekvensen. Dette er både biologisk fornuftigt da mutationer altid opstår på DNA-niveau, og teknisk eneste løsning da mange aminosyrer er kodet af flere codons, så hvordan skal computeren gætte den rigtige codon?

### *Indlæsning af sekvens fra tekstfil*

Sørg for at alignment explorer viser DNA. Den anden sekvens der skal bruges, skal indlæses ved først at downloade filen med sekvensen fra følgende link:

[www.bioteknologibogen.dk/bioteknologi-6/data/Ficolin-3-syg.txt](http://www.bioteknologibogen.dk/bioteknologi-6/data/Ficolin-3-syg.txt)

1. Sekvensen i filen tilføjes alignment ved at vælge 'Edit' og derefter 'Insert Sequence From File'. Vælg den downloadede fil (husk at ændre filtypen til '.txt' ellers kan filen ikke findes). Nu skulle Alignment Explorer-vinduet gerne se ud som vist herunder:

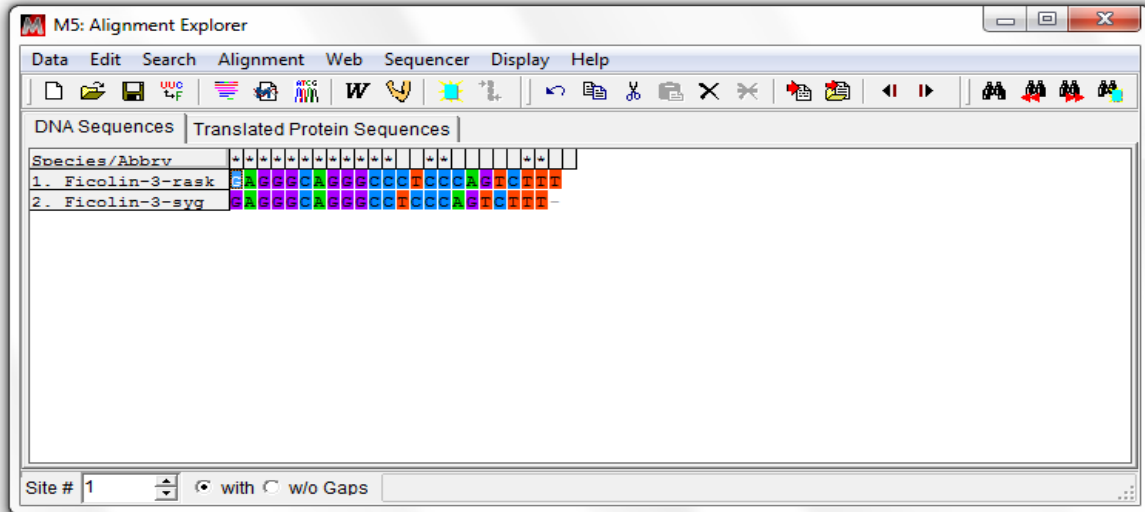


Læg mærke til stjernerne "\*" over sekvenserne. En stjerne viser at alle nucleotiderne i alle sekvenserne i den pågældende alignment-position er identiske. I dette tilfælde er de første 12 nucleotider identiske i de to sekvenser, hvorefter der kun er få tilsyneladende tilfældige identiteter i den viste alignment. Nederste venstre hjørne af skærbilledet viser alignment-positionen af din cursor, i det viste vindue er cursoren i alignment-position 1.

2. Indlæs samme fil en gang mere og se at man får to identiske sekvenser.
3. Kopier den øverste sekvens ved at markere den og vælge 'Edit' og 'Copy' og herefter 'Edit' og 'Paste'.
4. Slet de ekstra sekvenser ved at markere dem og trykke 'delete'-knappen på dit tastatur.
5. Ret det første G i sekvensen fra den raske person til et T. Hvad vil konsekvensen af en sådan 'mutation' være på proteinets primærstruktur? Sammenlign med figur 2, side 10 i Bioteknologi 6. Ret sekvensen tilbage når du er sikker på, at programmet gør som du forventer.

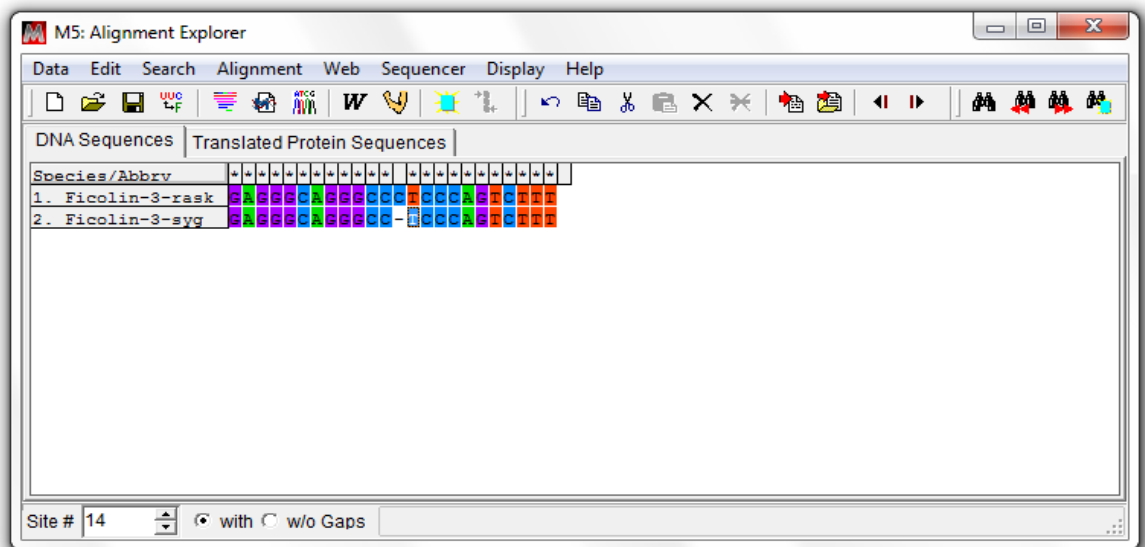
### Dannelse af manuel alignment

De to sekvenser du arbejder med, er identiske med sekvenser brugt i en tidligere eksamensopgave i Biologi A. Opgave 1 fra sættet den 25. maj 2010.



Ved hjælp af MEGA kan delspørgsmål 2 i ovennævnte eksamensopgave let besvares.

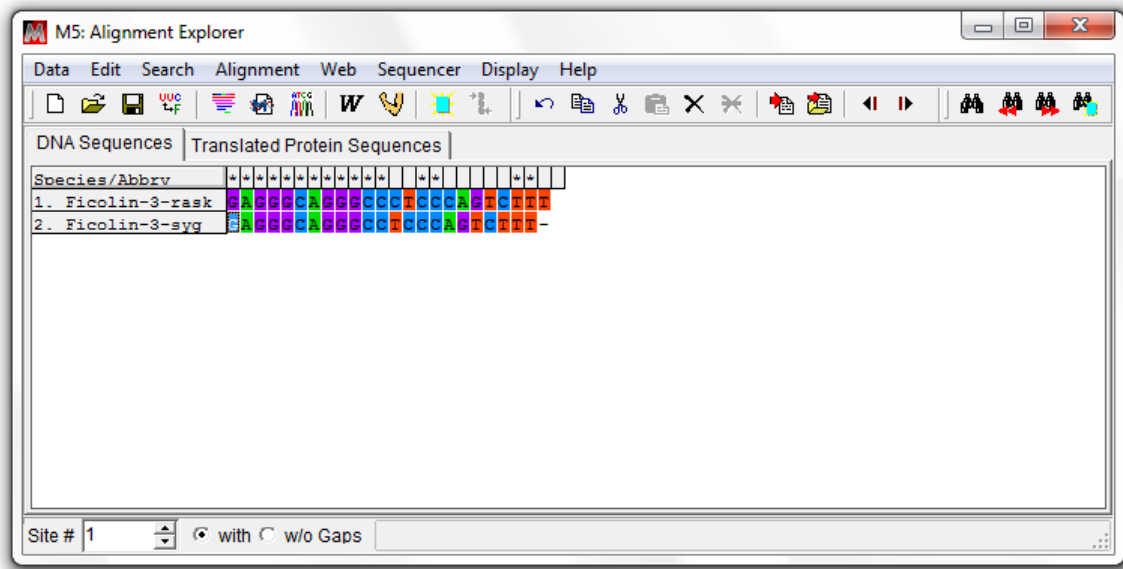
1. Placer cursoren i den nederste sekvens i position 13, og tryk på mellemrumsknappen. Hvad sker der med sekvensen? Hvad sker der med alignmenten? Hvis du har gjort det rigtigt, så bør dit vindue nu se ud som vist herunder:



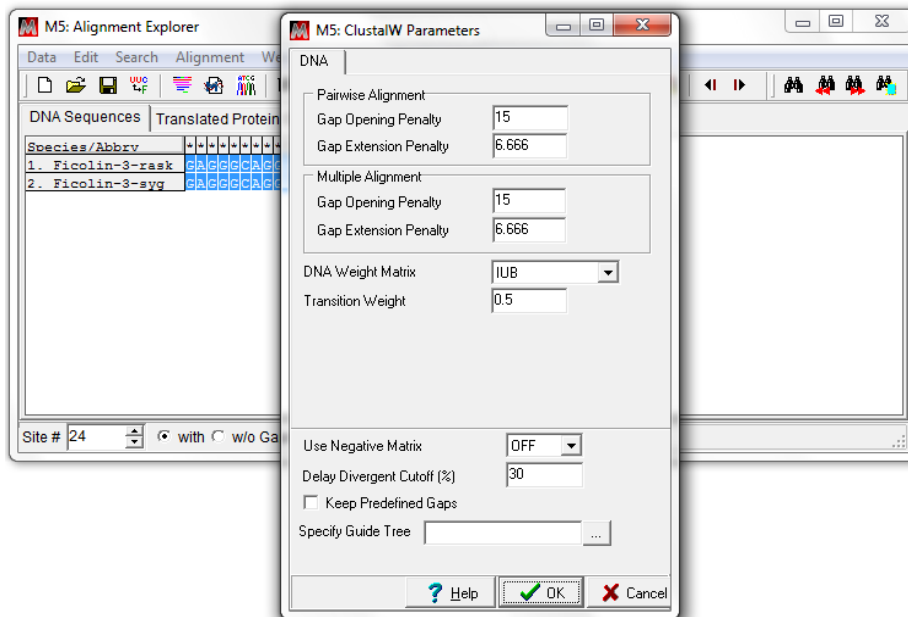
2. Angiv ud fra ovenstående figur, hvilken type mutation der er tale om hos den syge person. Begrund dit svar.
3. Fjern det indsatte mellemrum med 'delete'-knappen, og oversæt dernæst begge sekvenser til protein. Hvad er konsekvensen af mutationen på proteinets primærstruktur i den syge allel?

**Dannelse af computergenereret alignment**

1. Start med at slette alle indsatte mellemrum så de to sekvenser ikke længere er alignet.



2. Vælg nu 'Alignment' og 'Align by ClustalW' hvorefter følgende vindue dukker op:



3. De viste muligheder kan ændres efter behov afhængigt af hvor ensartede sekvenserne er, men i langt de fleste tilfælde er standardparametrene fornuftige. Vælg OK og undersøg den beregnede alignment.
4. Sammenlign den computergenererede alignment med den du selv lavede i hånden, er de ens?

## Alignment af proteinkodende sekvenser (codons)

I mange situationer ved man at de DNA-/RNA-sekvenser man ønsker at aligne, er proteinkodende og derfor bør alignment-metoden tage højde for dette. I denne øvelse skal to homologe sekvenser fra menneske (*Homo sapiens*) og ko (*Bos taurus*) alignes. Sekvenserne er begge et udsnit af mRNA for præinsulingenet.

1. Download og indlæs nedenstående to sekvenser eller indskriv dem manuelt:

### Sekvens 1

Insulin-udsnit-ko

GGAGGCCCGGGCGCGGGCGGCCUGGAGGG

www.bioteknologibogen.dk/bioteknologi-6/data/Proinsulin-udsnit-menneske.txt

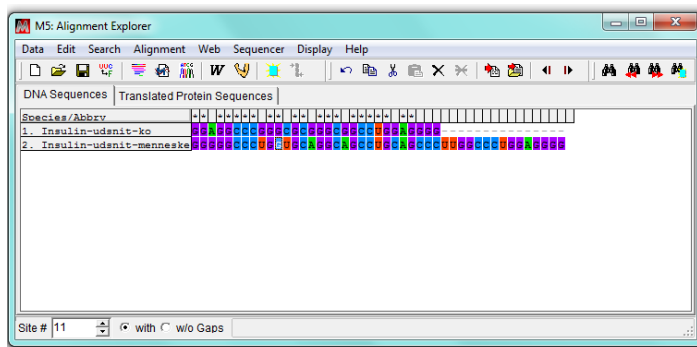
### Sekvens 2

Insulin-udsnit-menneske

GGGGCCUGGUGCAGGCAGCCUGCAGCCCUUGGCCUGGAGGGG

www.bioteknologibogen.dk/bioteknologi-6/data/Proinsulin-udsnit-ko.txt

Skærbilledet bør nu se ud som vist herunder:



De to sekvenser er ikke lige lange, og dette skyldes enten at der er sket en insertion i menneskets udviklingslinje, eller en deletion i koens udviklingslinje. Uden yderligere sekvenser kan vi ikke sige noget om hvilken af de to mulige forklaringer der er mest sandsynlig. Læg ydermere mærke til at det er RNA-sekvenser der er indlæst, hvordan kan man se det?

Scoringssystem	
Identitet:	2 point
Mismatch:	0 point
Indel	-2 point

Udvalgt scoringssystem, for flere detaljer se Bioteknologi 6 side 21.

2. Beregn alignment-scoren manuelt ved hjælp af det viste scoringssystem.
3. Lav en manuel alignment så du får den bedst mulige score, og udregn scoren.
4. Oversæt din alignment til protein, ser det fornuftigt ud?
5. Lav en computergenereret alignment ved hjælp af funktionen 'Align by ClustalW (Codons)' eller 'Align by Muscle (Codons)'.
6. Er den computergenererede alignment den samme som du fandt manuelt? Hvad er scoren af den computergenererede alignment i forhold til din manuelle alignment? Oversæt den computergenererede alignment til protein, ser det fornuftigt ud?
7. Sammenlign den fundne alignment med figur 12, side 19 i Bioteknologi 6.



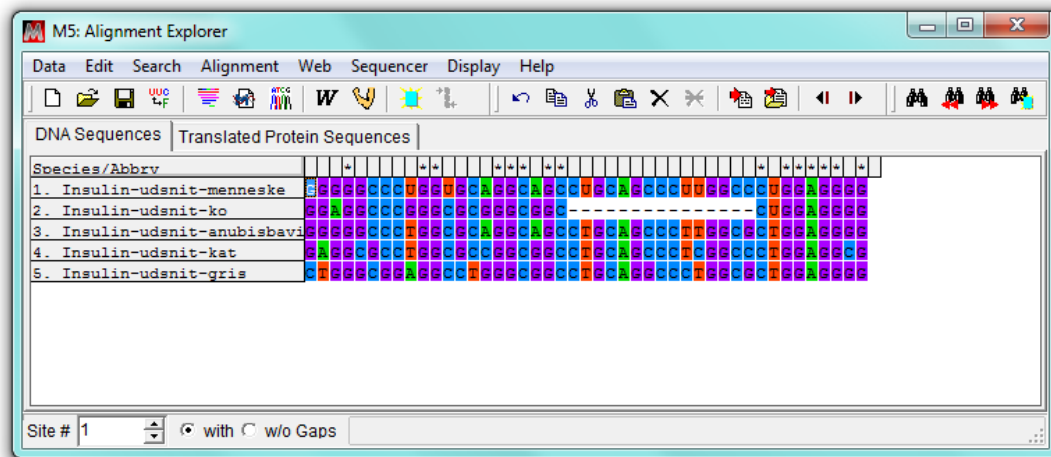
## Multiple alignments

Det er ikke altid man kun ønsker at sammenligne to sekvenser. Manuelt bliver det hurtigt meget komplekst at lave multiple alignments, men computeren kan heldigvis hjælpe med dette. I denne opgave tages udgangspunkt i det samme udsnit af insulingenet som blev brugt i den sidste opgave, men fra flere forskellige organismer.

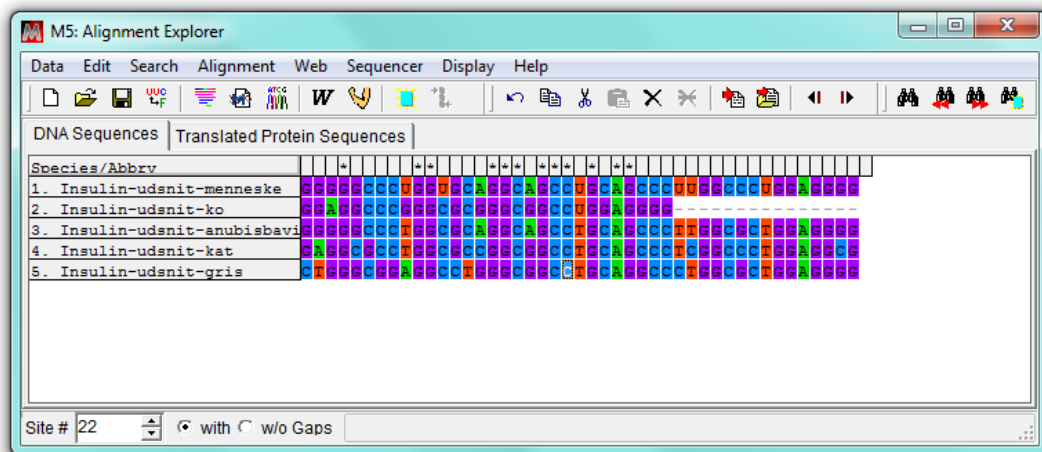
1. Download sekvenserne fra nedenstående link og indlæs dem i MEGA:

[www.bioteknologibogen.dk/bioteknologi-6/data/Proinsulin-udsnit-gris-bavian-kat.txt](http://www.bioteknologibogen.dk/bioteknologi-6/data/Proinsulin-udsnit-gris-bavian-kat.txt)

Skærbilledet skulle nu se ud som vist herunder:



2. Slet alle 'gaps' i ko-sekvensen. Placer cursoren på et gap og tryk delete. Man kan også markere alle '-' i sekvensen på en gang og trykke på delete. Skærbilledet skulle nu se ud på følgende måde:



3. Align de fem sekvenser med en af codon-alignment-metoderne.
4. Vurdér hvorvidt den viste indel-mutation er en insertion eller en deletion.