**Moord op het vliegveld**



Bron: Reizende DNA labs

Aangepast door: Marjon van Liessum en Marinke van der Velde

Datum: 8-1-2017

**INHOUD**

[Moleculaire biologie, een herhaling 3](#_Toc471650043)

[Genomics 5](#_Toc471650044)

[Bioinformatica 6](#_Toc471650045)

[Inleiding in het bioinformatica practicum 9](#_Toc471650046)

[Over het practicum 9](#_Toc471650047)

[Moord op het vliegveld - de hoofdrolspelers 10](#_Toc471650048)

[Moord op het vliegveld 11](#_Toc471650049)

[Identificatie van de 4 verdachte eiwitten 11](#_Toc471650050)

[Moord op het vliegveld, verdieping 15](#_Toc471650051)

[Namen van organismen: Latijn, Nederlands en Engels 17](#_Toc471650052)

[CSI – GENOMICS FORMULIER SCREENING VERDACHTE EIWITTEN 18](#_Toc471650053)

[Lijst aminozuursequenties van de 4 verdachte eiwitten 20](#_Toc471650054)

[Tabel met codons van het mRNA en de bijbehorende aminozuren 21](#_Toc471650055)

# Moleculaire biologie, een herhaling

**DNA**

De erfelijke informatie van een organisme ligt in het DNA opgeslagen in de vorm van een code. De code is opgebouwd uit vier bouwstenen of basen (A voor Adenine, C voor Cytosine, G voor Guanine, T voor Thymine). Deze basen volgen elkaar op in een bepaalde volgorde, bijvoorbeeld ...AGTCGTAATTGGCCCCAATTGCAAAAA.

Deze volgorde bepaalt de informatie in het DNA. Heel vaak zullen we ook gewoon over deze volgorde praten (zeker in de bioinformatica). Maar de DNA moleculen hebben natuurlijk ook een structuur:

De DNA-backbone bestaat uit suikers (deoxyribose) en fosfaten. Aan elke suiker is een van de volgende basen gebonden: adenine, guanine, cytosine en thymine. DNA komt steeds voor als een dubbele helix, de twee ketens in het dubbelmolecuul worden bijeen gehouden door waterstofbruggen tussen de basen. Adenine is steeds gebonden met thymine, en guanine met cytosine.

**Opdracht 1**

Vul onderstaande figuur aan met de namen: adenine, thymine, cytosine, guanine, uracil, fosfaat, deoxyribose, waterstofbrug, backbone, stikstofbasen.

****

**Opdracht2**

Een stuk DNA bevat 33% guanine. Welk percentage adenine, cytosine en thymine vind je in dat stuk?

**RNA**

Sommige stukken van het DNA (niet alles) coderen voor een eiwit. Een dergelijke opeenvolging of sequentie die codeert voor de aanmaak van één eiwit, noemt men een gen. Het eiwit wordt niet direct van het DNA gemaakt. Eerst wordt er een afschrift gemaakt van het DNA: messengerRNA. Het maken van een mRNA molecuul van een gen heet transcriptie. RNA is enkelstrengs en heeft dezelfde basen als DNA, met één uitzondering: Thymine (T) is vervangen door Uracil (U).



Zoals reeds vermeld werd, is DNA een code die leidt tot de vorming van eiwitten. Met behulp van enzymen wordt een RNA-kopie van één van de ketens van de dubbele helix gemaakt. Dit boodschapper-RNA bevat de code, waarmee de aminozuursequentie van eiwitten wordt gedicteerd. De eiwitsynthese vindt plaats op de ribosomen.

De vorming van boodschapper-RNA wordt transcriptie genoemd, de vorming van eiwitten translatie. Dit staat hieronder schematisch weergegeven.

Het is belangrijk te beseffen dat één stuk DNA, een gen, zal coderen voor één welbepaald eiwit. Grote stukken van het DNA van de mens (95%) coderen niet ergens voor. Deze stukken DNA zullen dus nooit afgelezen worden tot RNA.

We gaan ons in dit practicum richten op coderend DNA (of de eiwitten die hier

gecodeerd liggen).

****

Op blz. 21 vind je de code van het RNA: daarin kan je opzoeken welk aminozuur wordt gecodeerd door een bepaald codon.

**Opdracht 3**

Een stukje DNA heeft de volgende basenvolgorde: TACCATCCGTTCATGATC

De streng van het DNA die hier wordt weergegeven is de non-coding streng, hierop wordt het mRNA gemaakt.

Welk peptide zal hierdoor gecodeerd worden? Denk eraan dat je het DNA eerst moet omschrijven tot RNA. De volgorde is gewoon van links naar rechts.

# Genomics

Sinds een paar jaar is de volgorde bekend van de 3 miljard basen (A, C, T, en G) waaruit het menselijk DNA bestaat. We weten ook dat de mens zo'n 20.000 genen heeft. Hoe dat mega-project precies uitgevoerd is en het DNA van welke persoon als model genomen is laten we even in het midden.

Waar we het nu over moeten hebben is dat je eigenlijk nog geen definitieve antwoorden hebt wanneer je de volledige volgorde van het DNA hebt. Je hebt alleen maar veel nieuwe vragen en nieuwe uitdagingen. het eerste megaproject (bepalen van de volgorde) levert eigenlijk alleen maar een startpunt van een nieuwe uitdaging: wat moeten we / kunnen we met deze gegevens.

Het aantal van 3 miljard is in dit verband iets om even bij stil te staan: als we het volledige DNA van een mens zouden uitschrijven op de volgende manier:

CATGCAGTCAGTAGTACGTAGCCAATTGACTAGCATGACGATGCAGTAGCGCACGTACGATACATGCAGTCAGTAGTACGTAGCCAATTGACTAGCATGACGATGCAGTAGCGCACGTACGATACATGCAGTCAGTAGTACGTAGCCAATTGACTAGCATGACGATGCAGTAGCGCACGTACGATACATGCAGTCAGTAGTACGTAGCCAATTGACTAGCATGACGATGCAGTAGCGCACGTACGATACATGCAGTCAGTAGTACGTAGCCAATTGACTAGCATGACGATGCAGTAGCGCACGTACGATACATGCAGT (enzovoort)

**Opdracht 4**

Hoeveel pagina’s zou het boek dan hebben waar de hele DNA volgorde van een mens, op bovenstaande wijze staat afgedrukt? Ga uit van 50 regels per pagina.

**Opdracht 5**

Als je weet dat een pak van 500 A4 pagina’s 5,4 cm hoog is, hoe hoog wordt dan de stapel papier waarop je het DNA van één persoon kunt afdrukken?

Als je in het echte DNA boek met je vinger langs de regel zou gaan zoeken zou je aan een mensenleven niet genoeg hebben om één gen terug te vinden.

Het hele DNA van een organisme noemen we het genoom. Van heel veel soorten eukaryoten, archea en bacteriën is inmiddels ook het hele genoom bekend. Van een aantal virussen en bacteriën was het klusje al eerder geklaard. Dit heeft een hele nieuwe tak van genetica doen ontstaan. Genetica (genetics) heet het als we ons bezig houden met losse genen, genomics (geen Nederlandse term voor) heet het als we met hele genomen bezig zijn.

De enorme hoeveelheid gegevens maakt dat onderzoekers die met DNA, RNA en eiwit sequenties werken volledig afhankelijk zijn van computers. Genomics kan dus alleen maar bestaan dankzij de computers, dit heet bio-informatica.

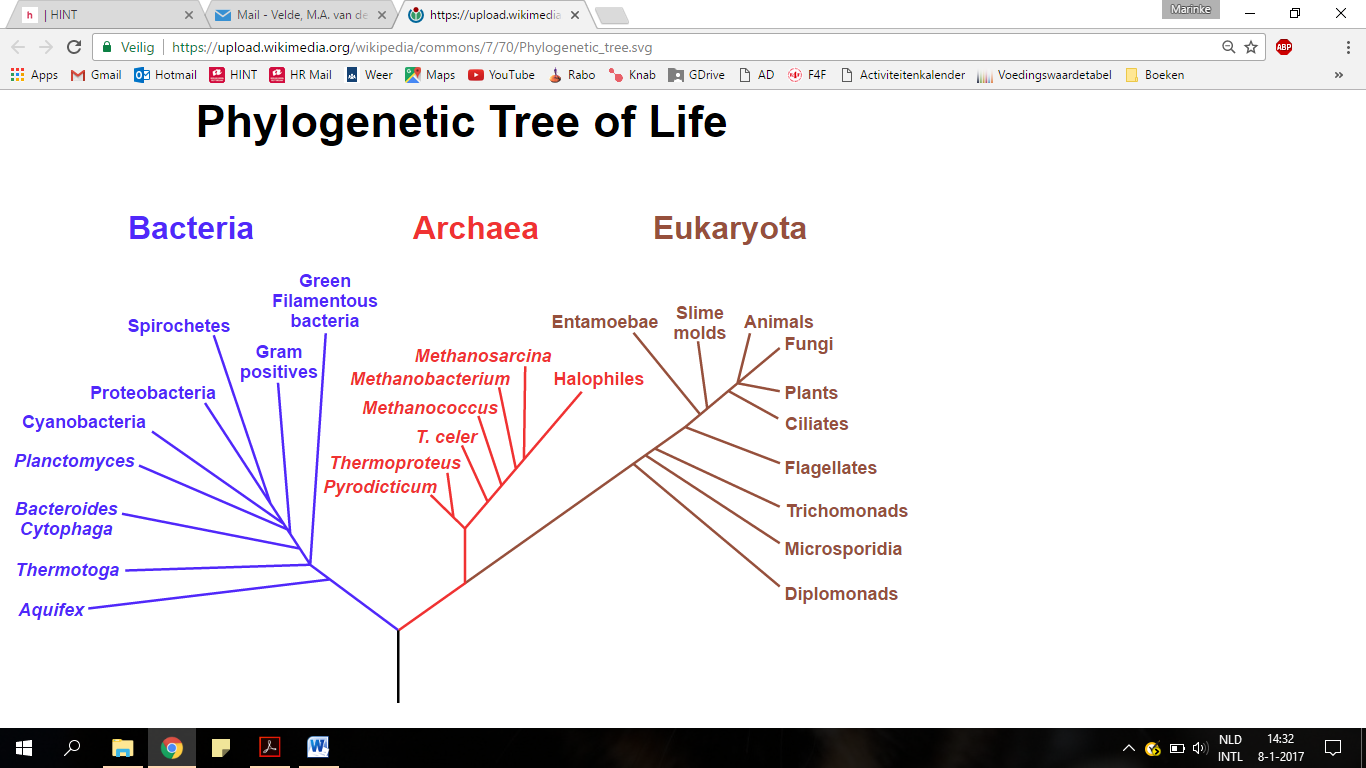
# Bioinformatica

Bioinformatica is ouder dan genomics. Ook toen nog niet het hele DNA van organismen bekend was maar alleen nog losse genen was er al lang teveel informatie om met een potloodje en een papiertje je gegevens te kunnen analyseren.

Voordat ik het type proef ga beschrijven wat we WEL gaan doen wil ik jullie even een idee geven hoe weids het landschap van de bio-informatica is. Ik beschrijf dus ook een aantal toepassingen die we NIET gaan doen. Lees het door en laat me weten als je vragen over een van deze voorbeelden hebt. Het proefje wat wij gaan doen, heeft betrekking op toepassing 6.

**Toepassing 1: Evolutionaire stambomen**

Dit komt erop neer dat je de computer de DNA-volgorde van een gen (of de aminozuurvolgorde van een eiwit) van een aantal organismen geeft. Het is wel van belang dat het van alle organismen hetzelfde gen is. Bijvoorbeeld coderend voor het enzym dat ATP maakt in de mitochondriën. De computer gaat de genen met elkaar vergelijken om dan een stamboom te maken. Als voorbeeld een evolutionaire stamboom met alle levende wezens erin. Rechts boven zitten wij (Animals).



De evolutionaire stamboom van hierboven bevat AL HET LEVEN op aarde. Deze is dus gemaakt op basis van een gen dat slechts heel langzaam verandert in de evolutie. Als je dat zelfde gen zou gebruiken om een stamboom van alle primaten (mens, gorilla, chimpansees en oerang-oetan) te maken loop je het risico dat ze gewoon allemaal hetzelfde zijn. Voor een grote stamboom neem je dus een gen wat weinig veranderd is (sterk geconserveerd is), voor een gedetailleerde stamboom van naaste verwanten een gen wat minder sterk geconserveerd is. je neemt hiervoor nooit de niet-coderende gebieden omdat daar de verschillen BINNEN EEN SOORT veel te groot zijn.

**Toepassing 2 Het vinden van de functie van een onbekend gen**

Een onderzoeker weet dat een bepaald stukje DNA ervoor zorgt dat een bacterie nu wel op een bepaald voedingsmedium kan overleven, terwijl hij dat zonder dat gen niet kon. Hij weet echter niet welk eiwit precies door dit gen gecodeerd wordt. Hij stuurt dan de nucleotide volgorde op (via internet) naar een grote computer die ergens in de wereld staat en krijgt binnen enkele minuten te horen (te lezen) op welke bekende genen zijn volgorde erg lijkt. Hij krijgt een top-100 lijst terug waarbij de computer de bekende genen die lijken op het ingestuurde gen op volgorde heeft gezet. Het gen wat het meest erop lijkt staat op 1.

Als de uitslag aangeeft dat op positie 1t/m 33 genen staan die coderen voor een enzym wat sucrose splitst in glucose en fructose (van verschillende organismen), weet hij vrij zeker dat zijn gen voor datzelfde enzym codeert, maar in een bepaalde bacterie.

**Toepassing 3 identificeren van een ziekteverwekker.**

Misschien hebben jullie wel eens het TV-programma “medical investigations” gezien. Om onverklaarbare wijze worden mensen ziek en gaan dood, het team van super-wetenschappers vliegt erheen en kan vele mensenlevens redden.

In deze situatie kan bio-informatica ook en rol spelen. Als je de DNA-volgorde van (een stukje van) het erfelijke materiaal van het onbekende virus kunt bepalen, stuur je het weer op naar de database. Het antwoord van de database vertelt je met welk bekend virus de ziekteverwekker verwant is.

Overigens: de onderzoekers in de TV serie zijn van het NIH (National Institute of Health) dat bestaat echt en heeft een van de meest gebruikte toegangsportals tot de databases onder zijn beheer.

**Toepassing 4 RNA-chips**

RNA-chips is niet het nieuwste type snack van Pringles, mocht je dat denken.

Nu zijn alle 20000 genen van de mens wel bekend, maar we weten nog lang niet de functie van allemaal. Om dat te onderzoeken zijn er kleine plaatjes (chips) gemaakt waar alle 20000 genen op zitten. Op een plaatje van 1 bij 1 cm zitten 20000 stukjes enkelstrengs DNA opgebracht (door een robot, mensen armen trillen teveel). Als je het RNA van een cel isoleert en op de chip brengt kan een andere robot zien van welk gen er veel RNA in de cel is en van welk gen weinig.

Als je deze informatie in de computer voert kan die door te vergelijken veel te weten komen over wanneer een gen wordt afgelezen.

Bijvoorbeeld: het RNA van een gezond persoon en het RNA van een persoon met koorts. (Twee keer hetzelfde celtype uiteraard) De computer kan je nu vertellen welke genen méér worden afgelezen bij koorts en welk minder.

Of je vergelijkt het expressie patroon van een maagcel en een spiercel om te ontdekken welke genen algemeen afgelezen worden en welk specifiek in een spiercel/maagcel.

Met RNA-chips onderzoek je dus vooral REGULATIE. Iedere cel heeft alle 20.000 genen, maar regulatie zorgt ervoor dat hij alleen het RNA maakt (transcriptie) wat hij nodig heeft.

**Toepassing 5 DNA fingerprinting**

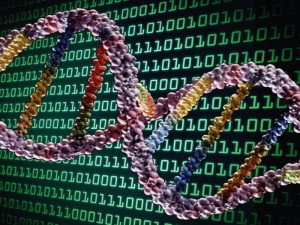
Veel mensen denken dat als je DNA fingerprint bekend is, de politie ook meteen weet of je een erfelijke ziekte hebt of iets dergelijks.

Dat is onzin. De meeste mensen hebben geen afwijkende genen. Als de politie naar de stukjes coderend DNA zou kijken krijg je een daderprofiel wat bij 99% van de bevolking past. Er worden dus vergelijkingen gemaakt van niet-coderende stukken (95% van je DNA codeert nergens voor) omdat daar de grootste verschillen zitten.

**Toepassing 6 Een eiwit identificeren**

Je vindt een eiwit in een drankje en bent benieuwd wat de herkomst daarvan is. Je bepaalt de aminozuurvolgorde van het eiwit en stuurt deze volgorde op naar een database. Zo kun je de herkomst (en eventueel de functie) van een eiwit opsporen. In ons practicum gaan we dit doen om een sterfgeval te verklaren. Sommige eiwitten van andere organismen zijn giftig en kunnen dus gebruikt worden om iemand te vermoorden.

# Inleiding in het bioinformatica practicum

Wij zijn geïnteresseerd in informatie over ons erfelijk materiaal, eiwitten en de functies van die eiwitten. Genen en eiwitten kunnen gesequenced worden: dat betekent dat men de volgorde van de basen in een gen, of de volgorde van aminozuren in een eiwit bepaalt.

De informatie moet op een intelligente manier geordend zijn, zodat onderzoekers makkelijk en snel bepaalde vragen kunnen beantwoorden aan de hand van wat ze vinden. Daarom is de informatie opgedeeld in databanken, die vaak vrij toegankelijk zijn voor iedereen op het Internet.

Er bestaat bijvoorbeeld een databank van

* eiwitstructuren (de Protein DataBank)
* aminozuursequenties van eiwitten en hun functies (SwissProt)
* enzymen en hun functies (ENZYME)
* de nucleotidesequentie van alle genen die tot vandaag gesequenced werden (EMBL)

Toch bestaan er een aantal onevenredigheden, die te maken hebben met de hedendaagse stand van de techniek. Zo bevat EMBL, de nucleotidendatabank, véél meer gensequenties dan er eiwitstructuren geregistreerd zijn in de Protein DataBank. De uitleg hiervoor is de volgende: een gen sequencen is niet erg moeilijk. Moeilijker is om daarna te bepalen voor welk eiwit het gen codeert en wat de functie van dat eiwit is. Als laatste stap moet de structuur van het eiwit bepaald worden.

De stand van zaken voor drie belangrijke databanken in 2016/2017[[1]](#footnote-1):

* EMBL: 1.996 miljard nucleotiden (waarvan het overgrote deel nog onbekende functies hebben) en 795,3 miljoen sequenties
* Swiss-Prot: 550.552 eiwitten waarvan de sequentie en functie bekend is.
* PDB: 77.487 eiwitstructuren

Swiss-Prot is eigenlijk onderdeel van een veel grotere eiwit-databank: UniProt KnowledgeBank. Uni-Prot is veel vollediger dan Swiss-Prot omdat het alle eiwitten bevat van Swiss-Prot plús de computer-gegenereerde translaties van alle genen uit EMBL (dit deel heet TrEMBL). Het nadeel is dat de informatie omtrent deze computer-gegenereerde eiwitten (nog) niet volledig is nagekeken en experimenteel bevestigd.

Met behulp van de databanken kun je allerlei vergelijkingen en zoekopdrachten uitvoeren. Heb je bijvoorbeeld een eiwit bij de mens gevonden dat een ziekte veroorzaakt, dan kun je in die databanken kijken of zo'n eiwit al eerder is gevonden en wat dat eiwit in het lichaam doet. Op die manier kun je sneller een medicijn ontwikkelen of een test om die ziekte vroegtijdig aan te tonen.

## Over het practicum

Schuilt er in jou een crime-scene-investigator? Dit practicum geeft je de mogelijkheid je voor even een forensisch onderzoeker te voelen die helpt bij het oplossen van een misdaad. In feite is het echte onderwerp van dit practicum *bioinformatica*, het gebruik van computers bij het zoeken naar en bestuderen van informatie over genen, nucleïnezuren en eiwitten. Terwijl je aan het speuren bent naar een aantal mysterieuze eiwitten, zul je in het practicum een aantal van de krachtige tools en databanken uit de bioinformatica gebruiken. Deze zijn in de inleidende les kort aan de orde geweest en hier zullen we ze in actie zien.

In het eerste deel gaan we op zoek naar het eiwit dat de dood van een Amerikaanse toerist heeft veroorzaakt. In het tweede deel gaan we de structuur van dit eiwit bekijken om hiermee een idee te krijgen over hoe belangrijk de juiste structuur van een eiwit voor zijn functie is.

## Moord op het vliegveld - de hoofdrolspelers

* Een vermoorde Amerikaanse toerist op het vliegveld.
* Een lijst met 4 verdachte eiwitsequenties.
* SwissProt, de databank met ruim vijfhonderdvijftigduizend eiwitsequenties die we tot nog toe kennen (en die experimenteel zijn bevestigd).
* BLAST, het programma om eiwitsequenties in de database SwissProt te vinden die heel veel lijken op (of gelijk zijn aan) de sequentie die je ingevoerd hebt.
* Een genomics-CSI-team. Dat ben jij samen met je teamgenoot. Je werkt dus in tweetallen.

Je weet nu genoeg om met het practicum te starten. Veel succes!

# Moord op het vliegveld

Als lid van het CSI-genomics team ben je opgeroepen om te assisteren bij een crimescene, er is een lijk van een Amerikaanse toerist gevonden op het vliegveld. Hij ligt in een soort kramp en uit autopsie blijkt dat hij inwendige bloedingen heeft. Naast hem is een flesje gevonden waaruit hij blijkbaar gedronken heeft, met daarin nog een restje vloeistof, het lijkt op melkresten. De vloeistof gaat naar het lab en je krijgt een lijstje terug van stoffen die erin zitten. Naast een aantal kleine moleculen, zoals suiker, blijkt dat er een viertal verschillende eiwitten is gevonden. Je besluit om deze verder te gaan onderzoeken om te kijken of je aanwijzingen kunt vinden voor de dood van de Amerikaan.

## Identificatie van de 4 verdachte eiwitten

De lijst met de aminozuurvolgorden (sequenties) van de 4 gevonden eiwitten (kandidaat 1 t/m kandidaat 4) is op blz. 20 te vinden. *Let op:* De aminozuursequenties worden gegeven in 1-lettercode, dit is de notatie die wetenschappers gebruiken. Als je die code niet meer weet, gebruik dan de tabel op blz. 21. Op blz. 18 en 19 staat het CSI - Genomics formulier Screening Verdachte Eiwitten. Voor elk van de 4 eiwitten moeten de volgende vragen op het formulier beantwoord worden:

1. Welk eiwit is het?

2. Uit welk organisme komt het?

3. Wat doet dit eiwit, wat is de functie van het eiwit?

4. Is dit verdacht? Is dit eiwit een potentiële dader? Waarom (niet)?

5. Eventuele andere opmerkelijke bevindingen.

Om je kennis te laten maken met het zoekprogramma BLAST zullen we eerst voor het eiwit kandidaat 1 uitleggen welke stappen uitgevoerd moeten worden. Daarna kun je kandidaat 2, 3 en 4 zelf doen en uitzoeken welk van de vier eiwitten ‘schuldig’ is. BLAST is een programma waarmee je een eiwitsequentie kunt vergelijken met alle eiwitsequenties in de database UniProt. UniProt bestaat uit twee delen: SwissProt en TrEMBL. Wij maken in dit practicum gebruik van SwissProt. Deze database bevat alle eiwitsequenties die de wetenschap op dit moment kent en die experimenteel zijn bevestigd. We zullen BLAST gebruiken vanuit de website <http://www.uniprot.org/blast/>.

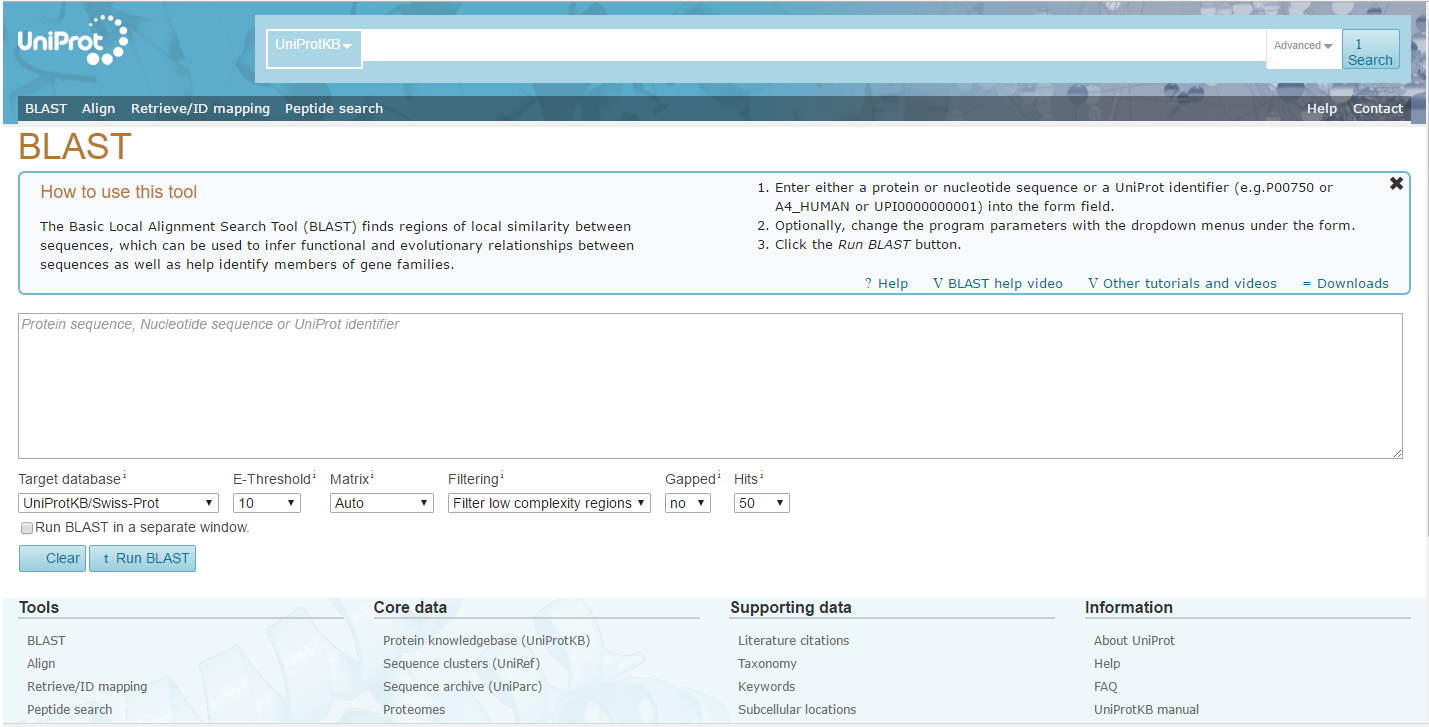
Na afloop van dit deel kun je het ingevulde formulier controleren. Als er tijd over is, dan kun je de verdiepende opdrachten voor nader onderzoek naar de 4 eiwitten maken.

In het volgende deel staat (in de vorm van opdrachten) voor kandidaat 1 uitgelegd welke stappen je moet uitvoeren en hoe je het resultaat moet bekijken.

**Opdracht 6**

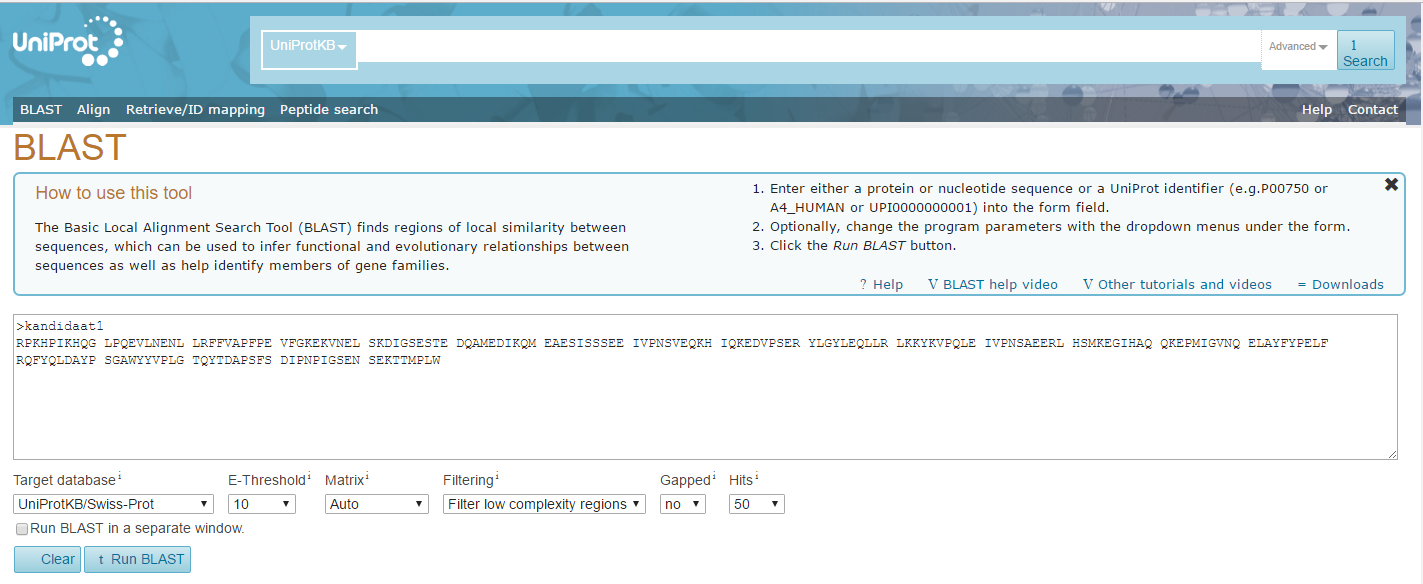
Open de website <http://www.uniprot.org/blast/>.

Het beginscherm ziet er als volgt uit:

****

**Opdracht 7**

Je ziet hieronder het voorbeeldscherm met daarin de aminozuursequentie van de eerste kandidaat. De eerste regel moet altijd beginnen met een > met daarachter de naam van het eiwit (dit is het zogenaamde FastA formaat). Hier is gekozen voor de naam kandidaat1. Op de tweede regel begint de aminozuursequentie.

****

Onder de ingevoerde aminozuursequentie staan een aantal instellingen. Controleer deze instellingen.

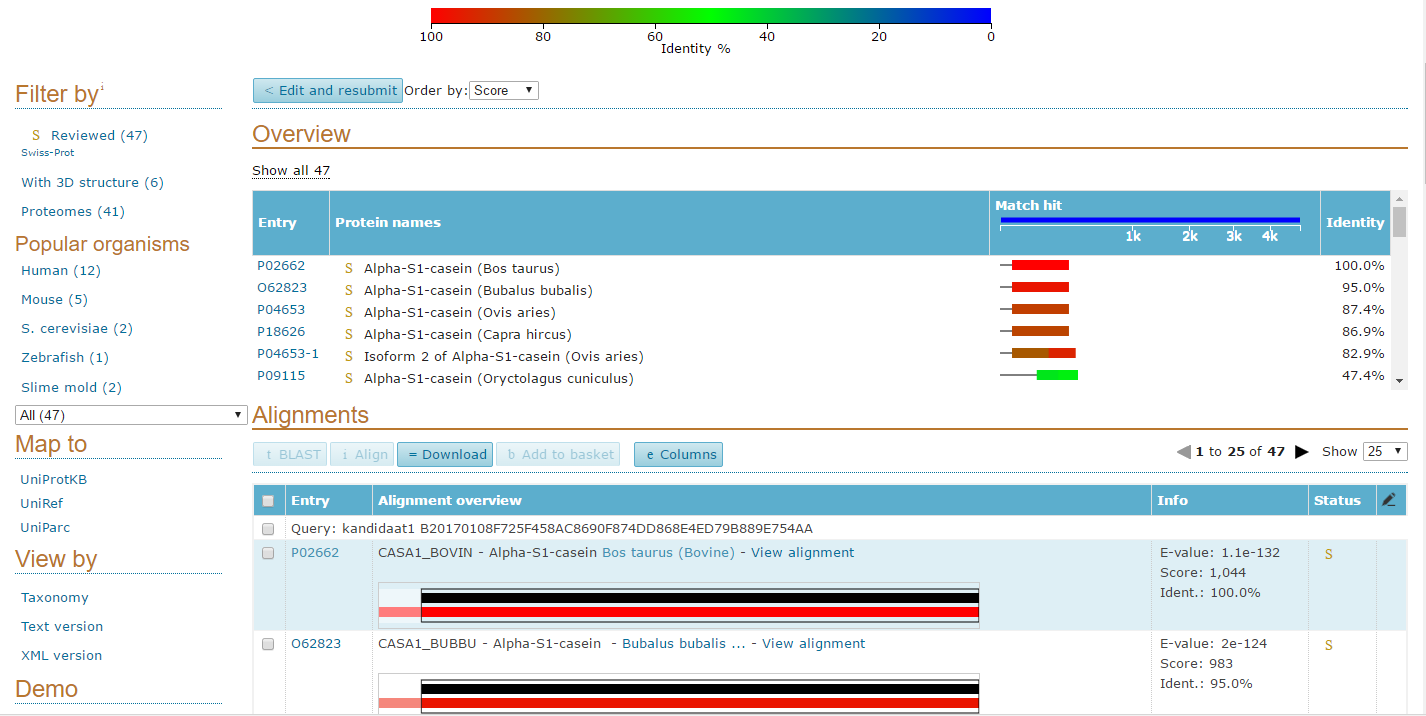
* Target database: UniprotKB/Swiss-Prot (je beperkt je tot dit deel van de database)
* E-Treshold: 10 (hoe lager dit getal, hoe nauwkeuriger de overeenkomst zal zijn)
* Matrix: Auto
* Filtering: Filter low complexity regions (de zoekmachine zoekt nu niet op aminzouursequenties die zeer algemeen zijn voor verschillende eiwitten)
* Gaps: No
* Hits: 50 (je beperkt het aantal hits en verkleint daarmee de duur van de zoekopdracht, alleen de 50 beste hits worden getoond)

**Opdracht 8**

Ga naar de lijst met verdachte eiwitten. De aminozuurvolgordes van de vier kandidaat eiwitten staan in het formaat waarin ze in BLAST gezet moeten worden. Kopieer nu de aminozuurvolgorde van kandidaat1 in het venstertje. Zet een vinkje bij ‘Run BLAST in a separate window’. Klik daarna op Run BLAST.

Je komt nu in het volgende scherm, waar ‘Job status: RUNNING’ komt te staan terwijl BLAST bezig is. Als je met veel mensen tegelijkertijd bezig bent kan dit enkele minuten duren. Je bent tenslotte een database met ruim vijfhonderdvijftigduizend eiwitsequenties aan het doorzoeken!

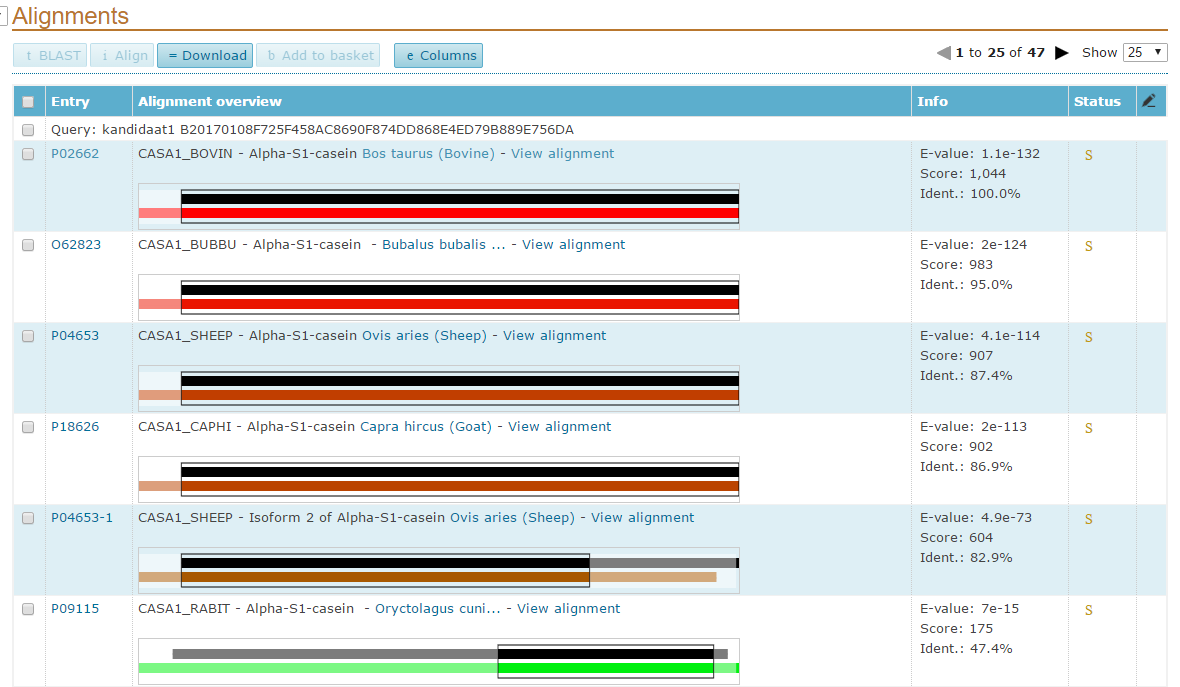
Als de zoektocht klaar is verschijnt het volgende resultaat:



Deze output betekent dat er 47 hits zijn gevonden in de database Swiss-Prot. Er zijn dus 47 eiwitten gevonden waarvan de aminozuurvolgorde lijkt op die van kandidaat 1.

**Opdracht 9**

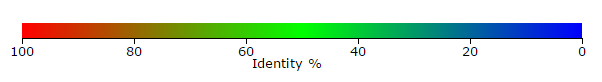
Scroll nu omlaag naar ‘Alignments’ om verder in detail naar de gevonden resultaten te kijken. Dit ziet er als volgt uit:



Dit scherm geeft je de zogenaamde hitlijst, de lijst van eiwitsequenties die gevonden zijn in de database Swiss-Prot en die lijken op kandidaat 1. Op nummer 1 staat de sequentie uit de database die het meest lijkt op (of zelfs identiek is aan) jouw ingevoerde sequentie, kandidaat 1. De code van deze sequentie is CASA1\_BOVIN.

Van links naar rechts lees je op elke regel de output van één van de hits:

* Entry: Dit is de code van de gevonden eiwitsequentie uit de database. Het eerste deel (bij de eerste hit casa1) slaat op de naam van het eiwit. Het tweede deel van de naam (bij de eerste hit bovin) wordt gevormd door een afkorting van het organisme, in dit geval bovine. Zoek op in de namenlijst op blz. 17 welk organisme dit is.
* Alignment overview: Hier zie je in kleur welk deel van jouw zoeksequentie (kandidaat 1) overeenkomt met de database sequentie (die is in zwart gegeven). Op de eerste regel zie je dat jouw zoeksequentie van begin tot eind overeenkomt met CASA1\_BOVIN. Ook geeft de kleur de mate van overeenkomst aan. Rood betekent een overeenkomst van 100%



* Info: E-value, Score en Ident. zijn getallen die het programma berekent en die jou een idee kunnen geven over de mate van overeenkomst en de betrouwbaarheid. Ook hier kun je zien dat de overeenkomst 100% is. Het getal dat je ziet bij E-value geeft aan hoe goed het beste resultaat is. Als dit getal heel klein is (en dat is in dit geval zo, het is 1,1 x 10-132), betekent dat dat het een betrouwbaar resultaat is.

**Opdracht 10**

We gaan de eerste hit, CASA1\_BOVIN, verder analyseren. Klik op het gekleurde balkje. Bekijk de gegevens.

Je ziet een zogenaamde alignment van twee aminozuursequenties. Query is de sequentie van jouw zoekvraag, in dit geval kandidaat1. Eronder staat de eiwitsequentie uit de database die het meest op kandidaat1 lijkt, in dit geval CASA1\_BOVIN. De regel tussen de 2 sequenties bevat de letters van de identieke aminozuren tussen Query en CASA1\_BOVIN. Als niet alle aminozuren hetzelfde zijn, zie je hier gaten vallen (meer hierover volgt onder de verdieping, vanaf opdracht 13).

Je kunt zien dat de volgorde van kandidaat 1 voor 100% overeenkomt met de volgorde van CASA1\_BOVIN, het alfa-S1-caseine eiwit (alle 214 aminozuren zijn identiek). Je hebt dus één van de vier eiwitten geïdentificeerd! Klik op de code P02662 om naar de database te gaan en de gegevens die bekend zijn over dit eiwit te gaan bekijken en uit te vinden of dit eiwit de ‘moordenaar’ kan zijn.

**Opdracht 11**

Probeer de informatie die er staat te lezen. Ga na wat voor eiwit kandidaat 1 is. Beantwoord nu de volgende vragen en vul ze in op het "CSI - Genomics formulier Screening Verdachte Eiwitten" op blz. 20.

* Welke informatie staat er bij Protein name?
* Welke informatie staat er bij Organism? Zoek ook de Nederlandse naam op in de lijst met organismen op blz. 17.
* Welke voor ons onderzoek interessante informatie staat er bij Function?
* Welke voor ons onderzoek interessante informatie staat er bij Keywords – molecular function?
* Wat is je conclusie? Is eiwit kandidaat1 verdacht, ja of nee?

**Opdracht 12**

**Onderzoek nu zelf de drie andere verdachte eiwitten!**

Doe nu zelf een BLAST search met sequenties van de andere drie gevonden eiwitten en vul de resultaten in op het "CSI - Genomics Formulier Screening Verdachte Eiwitten".

Wat is, rekening houdend met de autopsie, nu je eindconclusie over de moord? Waaraan is het slachtoffer overleden? Vul ook de eindconclusie in op het formulier.

**Nader onderzoek**

Je hebt in het eerste practicumdeel veel geleerd over een bepaald onderdeel van de bioinformatica, namelijk over de analyse van eiwitsequenties met behulp van bioinformaticatools en -databanken.

Heb je nog tijd over, dan kun je nog een stukje dieper gaan! In de verdieping van het eerste practicumdeel staan een paar opdrachten die wat verder ingaan op eiwitsequenties, belangrijke aminozuren, homologie tussen verwante organismen etc.

# Moord op het vliegveld, verdieping

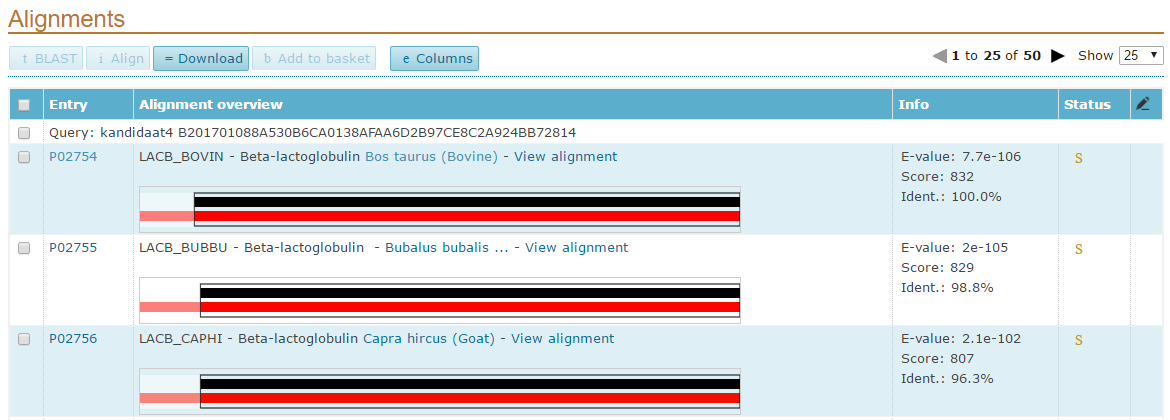
**Nader onderzoek van de verdachte eiwitten**

Voor dit onderdeel hoef je geen nieuwe BLAST searches te doen. Je kunt teruggaan naar de BLAST resultaten die je al in deel 1 hebt gevonden; je hoeft ze niet opnieuw te doen!

In de inleidende les is besproken dat eiwitten uit verschillende organismen waarvan de aminozuurvolgordes erg op elkaar lijken (homoloog zijn) vaak eenzelfde soort functie vervullen. We zullen dit nu eens nader bekijken aan de hand van de vier bovengenoemde eiwitten die je al aan een eerste inspectie hebt onderworpen.

**Opdracht 13**

Ga terug naar de BLAST resultaten van kandidaat 4:



De eerste hit was LACB\_BOVIN, kandidaat 4 is gelijk aan beta-lactoglobuline uit het rund. Kijk nu ook eens naar de derde hit in de hitlist. Dit is beta-lactoglobuline van de geit. Je ziet dat de aminozuurvolgorde van beide eiwitten voor 96,3% hetzelfde is. Er zijn maar een paar aminozuren verschillend.

Bepaal hoeveel aminozuren er verschillend zijn tussen geiten- en runder-beta-lactoglobuline.

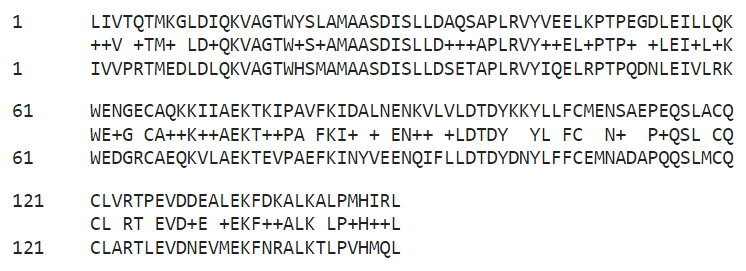
Je ziet wederom dat de regel tussen de twee sequenties de letters bevat van de identieke aminozuren tussen de Query en gevonden sequentie. Je ziet nu echter dat de regel tussen de twee sequenties ook lege plekken en plus-tekens bevat. Bijvoorbeeld, op de eerste regel zie je als eerste een L (leucine) staan in de rundersequentie en een I (isoleucine) in de geitensequentie. De aminozuren leucine en isoleucine lijken erg op elkaar en daarom staat er een plus-teken. Er zijn ook plekken waar een gat staat op de laatste regel, bijvoorbeeld vlak bij het eind. In het rund zit er een E (glutaminezuur) en in geit zit er een G (glycine). Glycine en glutaminezuur zijn twee aminozuren die helemaal niet op elkaar lijken. Zo kun je als je twee eiwitsequenties met elkaar vergelijkt door visuele inspectie heel snel de gebieden zien waar de twee eiwitten erg op elkaar lijken en ook de plekken waar ze wat meer van elkaar verschillen.

**Opdracht 14**

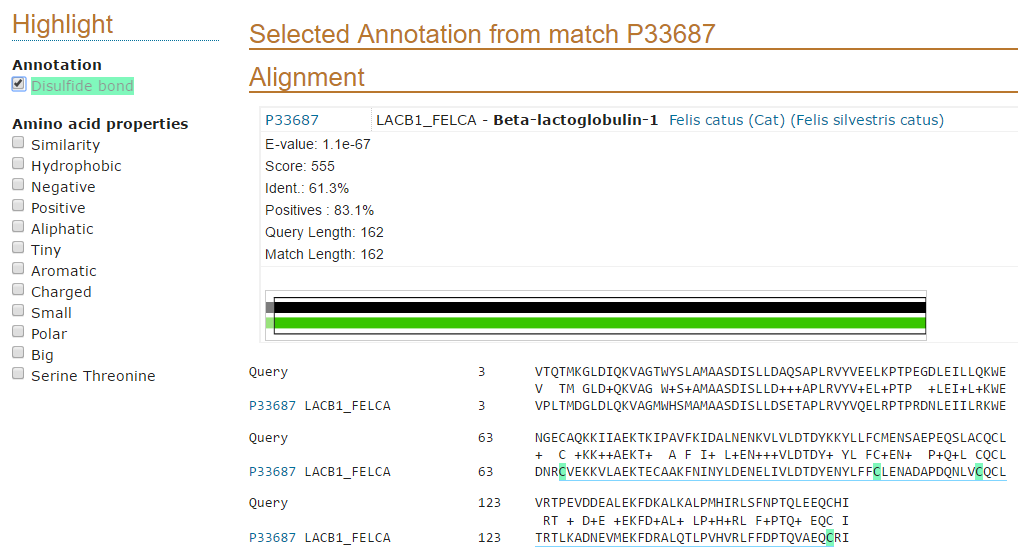
Zoek op in de BLAST results lijst hoeveel aminozuren gelijk zijn tussen het runder en het honden beta-lactoglobuline 1.

**Opdracht 15**

Bekijk hieronder de BLAST output van de vergelijking van runder- met honden-beta-lactoglobuline 1.



Alle belangrijke aminozuren voor beta-lactoglobuline zijn hetzelfde tussen deze twee dieren. Zij zijn behouden (geconserveerd) gebleven tijdens de evolutie. Zo zit er bijvoorbeeld een zwavelbrug tussen Cys66 en Cys160 in de sequenties. Wanneer je links in het scherm ´Disulfide bond´ aan vinkt, worden alle cysteines die zwavelbruggen vormen in de BLAST output weergegeven. Dit ziet er als volgt uit:



Je kunt zien dat zowel in de runder- als in de hondensequentie op vier overeenkomstige plekken cysteines zitten.

Je hebt nu veel geleerd over een bepaald onderdeel van de bioinformatica, namelijk over de analyse van eiwitsequenties met behulp van bioinformatica tools en databases.

|  |  |  |
| --- | --- | --- |
| Namen van organismen: Latijn, Nederlands en Engels | | |
| **Latijn** | **Nederlands** | **Engels** |
| *Bos taurus* | Rund | Bovine |
| *Bubalus bubalis* | Buffel | Domestic water buffalo |
| *Camelus dromedarius* | Dromedaris | Dromedary |
| *Canis familiaris* | Hond | Dog |
| *Capra hircus* | Geit | Goat |
| *Cavia porcellus* | Cavia | Guinea pig |
| *Crotalus atrox* | Texaanse ratelslang | Western diamondback rattlesnake |
| *Crotalus ruber* | Rode diamantratelslang | Red diamond rattlesnake |
| *Deinagkistrodon acutus* | Spitsneusgroefkopadder | Hundred-pace snake |
| *Equus asinus* | Ezel | Donkey |
| *Equus caballus* | Paard | Horse |
| *Felis silvestris catus* | Kat | Cat |
| *Gorilla gorilla* | Gorilla | Gorilla |
| *Homo sapiens* | Mens | Human |
| *Lemmus Lemmus* | Berglemming | Norway lemming |
| *Mus musculus* | Muis | Mouse |
| *Naja kaouthia* | Monocle cobra | Monocled cobra |
| *Naja naja naja* | Cobra (brilslang) | Cobra |
| *Okinawa habu* | Trimeresurus flavoviridis | Okinawa habu |
| *Oryctolagus cuniculus* | Konijn | Rabbit |
| *Ovis aries* | Schaap | Sheep |
| *Ovis orientalis musimon (Mouflon)* | Moeflon | Mouflon |
| *Oxyuranus scutellatus scutellatus* | Taipan | Taipan |
| *Pan troglodyte* | Chimpansee | Chimpanzee |
| *Rattus norvegicus* | Rat | Rat |
| *Sus scrofa* | Varken | Pig |

# CSI – GENOMICS FORMULIER SCREENING VERDACHTE EIWITTEN

**Vul hieronder de resultaten van het onderzoek in voor elk van de 4 gevonden eiwitten en je eindconclusie.**

Namen van het CSI-team

**Vragen voor elk eiwit:**

1. Welk eiwit is het?

2. Uit welk organisme komt het?

3. Wat doet deze stof, wat is de functie van het eiwit?

4. Is dit verdacht? Is dit eiwit een potentiële dader? Waarom (niet)?

5. Eventuele andere opmerkelijke bevindingen.

|  |  |
| --- | --- |
| Kandidaat 1 | 1 |
| 2 |
| 3 |
| 4 |
| 5 |
| Kandidaat 2 | 1 |
| 2 |
| 3 |
| 4 |
| 5 |
| Kandidaat 3 | 1 |
| 2 |
| 3 |
| 4 |
| 5 |
| Kandidaat 4 | 1 |
| 2 |
| 3 |
| 4 |
| 5 |
| Kandidaat 5 | 1 |
| 2 |
| 3 |
| 4 |
| 5 |
| **Eindconclusie: waaraan is het slachtoffer overleden?** |  |

# Lijst aminozuursequenties van de 4 verdachte eiwitten

*Let op:* De aminozuursequenties worden gegeven in 1-lettercode, dit is de notatie die wetenschappers gebruiken. Als je die code niet weet, kijk dan op blz. 21. Je kunt de eiwitsequenties van de kandidaten naar het MRS BLAST venster kopiëren. Zie ook de uitleg in de practicumtekst.

>kandidaat1 RPKHPIKHQG LPQEVLNENL LRFFVAPFPE VFGKEKVNEL SKDIGSESTE DQAMEDIKQM EAESISSSEE IVPNSVEQKH IQKEDVPSER YLGYLEQLLR LKKYKVPQLE IVPNSAEERL HSMKEGIHAQ QKEPMIGVNQ ELAYFYPELF RQFYQLDAYP SGAWYYVPLG TQYTDAPSFS DIPNPIGSEN SEKTTMPLW

>kandidaat2 QYSSNTQQGR TSIVHLFEWR WVDIALECER YLAPKGFGGV QVSPPNENVA IHNPFRPWWE RYQPVSYKLC TRSGNEDEFR NMVTRCNNVG VRIYVDAVIN HMCGNAVSAG TSSTCGSYFN PGSRDFPAVP YSGWDFNDGK CKTGSGDIEN YNDATQVRDC RLSGLLDLAL GKDYVRSKIA EYMNHLIDIG VAGFRIDASK HMWPGDIKAI LDKLHNLNSN WFPEGSKPFI YQEVIDLGGE PIKSSDYFGN GRVTEFKYGA KLGTVIRKWN GEKMSYLKNW GEGWGFMPSD RALVFVDNHD NQRGHGAGGA SILTFWDARL YKMAVGFMLA HPYGFTRVMS SYRWPRYFEN GKDVNDWVGP PNDNGVTKEV TINPDTTCGN DWVCEHRWRQ IRNMVNFRNV VDGQPFTNWY DNGSNQVAFG RGNRGFIVFN NDDWTFSLTL QTGLPAGTYC DVISGDKING NCTGIKIYVS DDGKAHFSIS NSAEDPFIAI HAESKL

>kandidaat3 QQNLPQRYIE LVVVADHRVF MKYNSDLNTI RTRVHEIVNF INGFYRSLNI HVSLTDLEIW SNEDQINIQS ASSDTLNAFA EWRETDLLNR KSHDNAQLLT AIELDEETLG LAPLGTMCDP KLSIGIVQDH SPINLLMGVT MAHELGHNLG MEHDGKDCLR GASLCIMRPG LTKGRSYEFS DDSMHYYERF LKQYKPQCIL NKP

>kandidaat4 LIVTQTMKGL DIQKVAGTWY SLAMAASDIS LLDAQSAPLR VYVEELKPTP EGDLEILLQK WENGECAQKK IIAEKTKIPA VFKIDALNEN KVLVLDTDYK KYLLFCMENS AEPEQSLACQ CLVRTPEVDD EALEKFDKAL KALPMHIRLS FNPTQLEEQC HI

# Tabel met codons van het mRNA en de bijbehorende aminozuren



1. Bronnen, geraadpleegd op 8-1-2017:

   <http://www.ebi.ac.uk/ena/about/statistics>

   <http://www.sib.swiss/sp30/sp-down-the-years>

   <http://www.rcsb.org/pdb/home/home.do> [↑](#footnote-ref-1)