

Beste docent/leerling

Voor u ligt NAVIGENE. Dit is een navigatie-instrument ontwikkeld voor biologiedocenten in het voorgezet onderwijs door het Netherlands Bioinformatics Centre en het Freudenthal Instituut voor Didactiek van Wiskunde en Natuurwetenschappen. Bioinformatica is een onderzoeksveld waar onderzoek wordt gedaan aan DNA en eiwitten. U kunt bioinformatica-tools en databanken gebruiken om authentiek onderzoek in uw klas te halen. Daarnaast maakt het genetica-onderwijs tastbaarder en visueler, waarmee het begrip van DNA en eiwitten van uw leerlingen zal worden vergroot.

In ontwikkeling

Graag wijzen wij u erop dat NAVIGENE altijd in ontwikkeling is: bioinformatica is een dynamisch onderzoeksveld, dus links naar websites en de lay-out van tools kunnen veranderen. Graag ontwikkelen wij NAVIGENE met behulp van u. Als u een dode link, veranderde lay-out van een website of verkeerde instructies tegenkomt, horen we dat graag. Ook wensen en andere opmerkingen zijn welkom. Samen komen we verder.

Gebruik van NAVIGENE

NAVIGENE bestaat uit een voorblad en een instructiegids. Op het voorblad treft u een schema met vragen. Het idee is dat u begint in het driehoek fenomeen-eiwit-gen, doordat u een bepaald celproces nader wilt toelichten of een gen of eiwit bent tegengekomen in het biologieboek. Vanuit dit startpunt kunt u pijlen (met daarbij

behorende instructies) volgen. Onderaan het schema staan vragen die ook beantwoord kunnen worden met bioinformatica.

Bij iedere pijl staat een vraag, achter iedere vraag staat een nummer, bijvoorbeeld P.4. Dit betekent dat u deze vraag kunt beantwoorden met de instructie die staat beschreven op pagina 4. De instructie is opgebouwd zoals aangegeven in het figuur.



Om vertrouwd te raken met NAVIGENE kunt u zelfstandig de instructiegids doornemen. De opdracht-blokjes aan het einde van iedere instructie helpen u hierbij. U kunt ook contact opnemen om informatie te winnen over de eerstvolgende workshop die over de NAVIGENE gegeven wordt. Een workshop op maat voor uw school behoort ook tot de mogelijkheden.

Wij wensen u veel nuttige ontdekkingen en waardevolle verrassingen bij het gebruik van NAVIGENE. Suggesties, vragen, informatie over workshops en andere opmerkingen kunt u richten aan onderwijs@nbic.nl.

De NAVIGENE is gemaakt door Hienke Sminia in nauwe samenwerking met Dirk Jan Boerwinkel.

→ Welk eiwit of gen is betrokken bij dit fenomeen?

1. Zoek via Google (www.google.nl) naar het verschijnsel
2. Scan de beschikbare informatie naar genen
3. Gebruik eventueel Wikipedia om je verder op weg te helpen

1. Bij eigenschappen (zoals haarkleur, oogkleur, een zesde teen of oorlel-vorm), ziektes en aandoeningen (zoals klompvoet, ziekte van Huntington, sikkelcelziekte of kleurenblindheid) en processen (zoals spijsvertering, zien, fotosynthese en insulineproductie) spelen eiwitten en genen een grote rol. Er bestaat geen databank waarmee je eenvoudig kunt achterhalen welke eiwitten en genen een specifieke rol spelen in een bepaald verschijnsel. Wel staat heel veel informatie op het internet op websites van bijvoorbeeld onderzoeksinstituten, patiëntenverenigingen en online encyclopedieën. Met een zoekmachine zoals Google kun je hierin zoeken.

Ga naar <http://www.google.nl>

2. Type het verschijnsel in het tekstvak. Vul dit eventueel aan met specifieke zoektermen als 'eiwit' of 'gen'. Als een verschijnsel uit meerdere woorden bestaat kun je de woorden tussen dubbele aanhalingstekens zetten. Bijvoorbeeld: "ziekte van Huntington". Klik op 'Google zoeken' en bekijk diverse pagina's om te achterhalen welke eiwitten en/of genen betrokken zijn bij het verschijnsel. Je kunt de functie van het gevonden eiwit en/of gen controleren door het op te zoeken in een databank: zie pagina 5 *Waarbij is dit eiwit betrokken?* of zie pagina 6 *Waarbij is dit gen betrokken?*
3. Je kunt ook direct zoeken in Wikipedia naar de betrokken eiwitten en genen van een verschijnsel. Vooral de Engelse Wikipedia kun je heel eenvoudig inzetten om verder te zoeken naar gegevens over eiwitten in databanken.

Voer de Engelse naam van het verschijnsel waar je meer over wilt weten in het tekstvak in Google en voeg hieraan de zoekterm 'wiki' toe.


De eerste hit, of eventueel een van de bovenste hits, geeft de juiste wiki-pagina. Klik hierop.

Je kunt natuurlijk ook direct in Wikipedia zoeken met behulp van de zoekfunctie op <http://en.wikipedia.org>

4. De hoofdtekst van de wikipagina over een eiwit beschrijft in welke processen het eiwit betrokken is. Aan de rechterkant staat vaak een tabel met de volgende onderdelen:

- Afbeelding van de structuur: onderaan de structuur staat een beschrijving van de getoonde afbeelding. Soms staat hierin dat de afgebeelde structuur alleen voorkomt in een bepaalde vorm van opslag, of van welk organisme deze structuur is.

Insulin



Computer-generated image of six Insulin molecules assembled in a hexamer, highlighting the threefold symmetry, the zinc ions holding it together, and the histidine residues involved in zinc binding. Insulin is stored in the body as a hexamer, while the active form is the monomer. [1]

Available structures

PDB: [Ortholog search: PDBe](#) [RCSB](#)

[List of PDB id codes](#) [show]

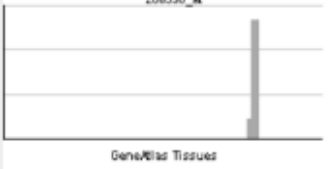
Identifiers

Symbols [INS](#); [IDDM2](#); [ILPR](#); [IRDN](#); [MODY10](#)

External IDs [OMIM: 176730](#) [MG: 96573](#) [HomoloGene: 173](#)
[ChEMBL: 5881](#) [GeneCards: INS Gene](#)

[Gene Ontology](#) [show]

RNA expression pattern



[More reference expression data](#)

Orthologs

Species	Human	Mouse
Entrez	3630	16334
Ensembl	ENSG00000254647	ENSMUSG00000000215
UniProt	P01308	P01326
RefSeq (mRNA)	NM_000207.2	NM_001185083.1
RefSeq (protein)	NP_000198.1	NP_001172012.1
Location (UCSC)	Chr 11: 2.18 – 2.18 Mb	Chr 7: 142.68 – 142.7 Mb
PubMed search	[1]	[2]

- *Available structures*: (klik op *show*) PDB staat voor de Protein DataBank (zie pagina 13 – *Hoe ziet de 3D-structuur van het eiwit eruit?*), ernaast staan alle codes voor structuren van het betreffende eiwit die in de databank staan. Dit zijn onder andere structuren van eiwitten die gemuteerd zijn, verschillende verschijningsvormen hebben of bestaan uit meerdere complexen. De beste structuur voor gebruik in de klas is een kwestie van bekijken en uitproberen.

- *Identifiers*: Bij *symbols* staan de ID-codes van het eiwit zoals die in verschillende databanken worden benoemd. De eerste code is over het algemeen de meest gebruikte. Deze code kun je gebruiken als zoekterm in MRS (zie pagina 5). *External IDs* zijn externe ID-codes van databanken die een afwijkende codering hebben.

- *Gene Ontology*: (klik op *show*) Geeft een opsomming van moleculaire functies, cellulaire componenten en biologische processen waar het eiwit direct of indirect bij betrokken is.

- *RNA expression pattern*: Geeft een grafiek met balkje voor verschillende weefsels in het lichaam. Hoe hoger de balk, hoe meer expressie van het gen voor het eiwit in dit weefsel. Door te klikken op de grafiek, vergroot je de afbeelding. Als er meerdere grafieken staan, dan geven zij het expressiepatroon in verschillende proefpersonen of in verschillende organismen weer. Dit staat echter lang niet altijd duidelijk aangegeven.

- *Orthologs*: Geeft een overzicht van gen-gegevens en een vergelijking tussen mens en een ander organisme (vaak de muis). Van onder naar boven staan hier:

Species – soort organisme

Entrez – zoekmachine dat verschillende medische databanken doorzoekt

Ensembl – databank met genomen van verschillende organismen (genomebrowser) (zie pagina 4)

Uniprot – databank die eiwitdata combineert van verschillende eiwitdatabanken

RefSeq (mRNA) en *RefSeq (protein)* – Referentie sequenties van het mRNA en het eiwit

Location (UCSC) – locatie op het chromosoom (chromosoomnummer en coördinaat)

PubMed search – zoeken in de databank van artikelen in wetenschappelijke tijdschriften. (Voor veel artikelen in deze databank heb je een abonnement op het tijdschrift nodig. De meeste universiteiten hebben zo'n abonnement.)

Zoek eens op welk verteringseiwit betrokken is bij de secretie van maagzuur in de maag tijdens de spijsvertering.

Of zoek eens op welke genen waarschijnlijk betrokken zijn bij de oogkleur. Het gen met het grootste aandeel is ook betrokken bij een ander kleurig fenomeen, kun je vinden welke?

→ Voor welk eiwit codeert dit gen?

1. Ga naar <http://www.ensembl.org>
2. Zoek het gewenste gen
3. Klik door naar de Uniprot identifiers

1. Ensembl is een zogenoemde genome browser waarin je allerlei informatie kunt opzoeken over genen, zoals bijvoorbeeld voor welk eiwit het gen codeert. Op pagina 18 staan andere toepassingen van deze genome browser. Let op: in Ensembl kun je alleen zoeken in vertebraten en andere eukaryoten. Planten zijn bijvoorbeeld niet opgenomen in de databanken. Ga naar <http://www.ensembl.org>
2. Gebruik de zoekfunctie op de hoofdpagina van Ensembl om het gen van jouw interesse te vinden. Je kunt hiervoor verschillende zoektermen gebruiken, zoals de naam van het gen, het gensymbool of de coördinaten van de locatie van het gen. Klik op 'Go' om de zoekopdracht te starten.
3. Je krijgt nu een overzicht van de gevonden items. Zoek naar de naam van jouw gen, met daarachter (*Human Gene*). Als je een gen zoekt van een ander organisme dan de mens, moet je uiteraard uitkijken naar (*soortnaam Gene*). Klik op dit item voor meer informatie.
4. Het informatieblad dat je nu ziet geeft veel informatie over het gen, zoals de locatie, het aantal transcripten, etc. Om te weten waarvoor het gen codeert moet je kijken in de *Summary*, en vervolgens achter *UniprotKB*. Uniprot is een eiwitdatabank (vergelijkbaar met Swissprot) waarin veel eiwitten staan. Als jouw gen codeert voor een bekend eiwit, dan heeft het een 'identificer' in Uniprot. Je kunt rechtstreeks op de link klikken voor informatie over het eiwit, maar dit functioneert niet altijd. Je kunt ook naar www.uniprot.org en de identificer-code kopiëren in het zoekveld. Je wordt dan direct naar de juiste informatie gebracht.

Gene: BRCA2 ENSG00000139618

Description: breast cancer 2, early onset [Source:HGNC Symbol;Acc:HGNC:1101]

Synonyms: BRCC2, FACD, FAD, FAD1, FANCD, FANCD1

Location: Chromosome 13: 32,315,474-32,400,266 forward strand.

INSDC coordinates: chromosome:GRCh38:CM000675.2:32315474:32400266:1

Transcripts: This gene has 7 transcripts (splice variants) [Hide transcript table](#)

Name	Transcript ID	bp	Protein	Biotype	CCDS	RefSeq	Flags
BRCA2-001	ENST00000380152	11986	3418 aa	Protein coding	CCDS9344	-	TSL:5 GENCODE basic (P)
BRCA2-201	ENST00000544455	10984	3418 aa	Protein coding	CCDS9344	NM_000059 NP_000050	TSL:1 GENCODE basic (P)
BRCA2-002	ENST00000470094	842	186 aa	Nonsense mediated decay	-	-	CDS 5' incomplete TSL:5
BRCA2-005	ENST00000528762	495	64 aa	Nonsense mediated decay	-	-	CDS 5' incomplete TSL:4
BRCA2-004	ENST00000614259	7950	No protein	Processed transcript	-	-	TSL:2
BRCA2-003	ENST00000530893	2011	No protein	Processed transcript	-	-	TSL:1
BRCA2-006	ENST00000533776	523	No protein	Retained intron	-	-	TSL:3

Summary

Name: [BRCA2](#) (HGNC Symbol)

CCDS: This gene is a member of the Human CCDS set: [CCDS9344](#)

UniprotKB: This gene has proteins that correspond to the following Uniprot identifiers: [P51587](#)

LRG: [LRG_293](#) provides a stable genomic reference framework for describing sequence variations for this gene

Ensembl version: ENSG00000139618.12

GRCh37 assembly: This gene maps to [32,889,611-32,974,403](#) in GRCh37 coordinates. View this locus in the GRCh37 archive: [ENSG00000139618](#)

Gene type: Known protein coding

Prediction Method: Annotation for this gene includes both automatic annotation from Ensembl and [Havana](#) manual curation, see [article](#)

Alternative genes: This gene corresponds to the following database identifiers:
Havana gene: [QTHUMG0000017411](#)

Zoek naar het eiwit waar het gen BRCA2 voor codeert.

- Welk gen codeert voor dit eiwit?
- Waarbij is dit eiwit betrokken (functie)?
- Wat is de aminozuurvolgorde van dit eiwit?
- Waar zit het eiwit in de cel?

1. Ga naar de website <http://mrs.cmbi.ru.nl>
2. Voer in het zoekveld de Engelse naam van het eiwit in
3. Klik op de juiste hit en scroll naar de juiste informatie

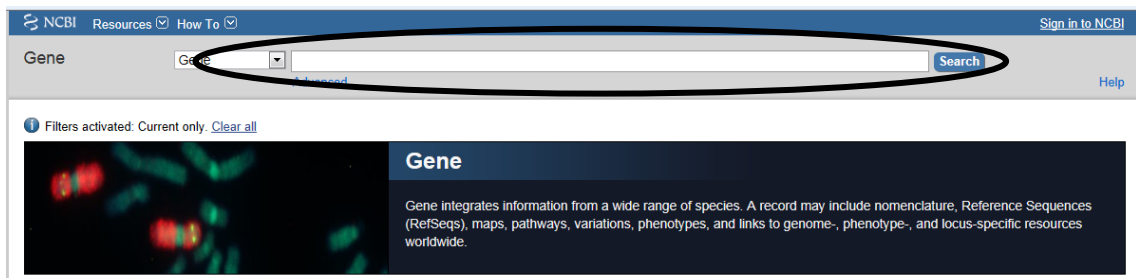
1. Via de website <http://mrs.cmbi.ru.nl> kun je in allerlei databanken zoeken naar eiwitten en genen. Als je zoekt naar een eiwit kun je het beste de databank Swiss-Prot selecteren. Typ in het zoekveld de Engelse naam van het eiwit in.
2. Je krijgt nu een lijst met hits. Deze eiwitten zijn allemaal gerelateerd aan jouw zoekopdracht. Dat kan betekenen dat het het eiwit is, maar ook dat het een eiwit is dat een interactie aangaat met het gezochte eiwit. Bekijk dus goed of de korte omschrijving klopt met jouw wensen. De ID kan hierbij helpen. De eerste letters staan voor de afkorting van het eiwit. De letters na het liggend streepje geven het organisme waarin het eiwit voorkomt aan. Als je alleen maar resultaten wilt zien van één soort organisme, voeg dan *os:human* (origin species: human) toe aan de zoekopdracht. Nu vind je alleen maar eiwitten van de mens. Hetzelfde trucje geldt natuurlijk ook voor andere organismen.
Het programma kent een soort relevantie toe aan de hits. Hoe verder het balkje gevuld is, hoe hoger de relevantie.
3. Klik op de ID-code van jouw gekozen eiwit. Je ziet nu een overzicht van gegevens die over dit eiwit bekend zijn. Controleer bij de *protein name* of je inderdaad het juiste eiwit hebt gekozen. Als dit het geval is, kun je verder. Afhankelijk van welke informatie je zoekt, moet je in een andere sectie van de pagina zoeken:
 Naam van het coderende gen: in de tweede sectie 'Name and origin of the protein' staan *Gene names*. De *Name* geeft de meest gebruikte naam van het gen. Bij *Synonyms* staan alternatieve namen van hetzelfde gen. Soms staat er alleen een gensymbool (afkorting van letters en cijfers)
 Aminozuurvolgorde: helemaal naar onder scrollen. Hier vind je de *sequence information*. De lengte in aminozuren en het gewicht in Dalton worden hier gegeven. Daaronder staat de aminozuurvolgorde in de één-lettercode.
 Functie: scroll naar onder tot het kopje *Comments*. Hier staat de functie beschreven. Als het eiwit enzymatische activiteit(en) betreft, staat dit beschreven als *catalytic activity*, (procesversnellende activiteit). Vaak kun je ook al wat trefwoorden over de functie vinden bovenaan bij *name and origin of the protein* bij *Keywords*.
 Locatie van het eiwit in de cel: staat eveneens onder het kopje *Comments*. Hier staat de *Subcellular location* beschreven.

Zoek de naam van het gen, de functie, aminozuurvolgorde en locatie in de cel van het grootste eiwit in ons lichaam maar eens op: titine (Engels titin). *Let op: bij het zoeken heeft dit eiwit niet de hoogste relevantie, dus het staat niet helemaal bovenaan bij de getoonde hits!*

→ Waarbij is dit gen betrokken (functie)?

1. Ga naar <http://www.ncbi.nlm.nih.gov/gene>
2. Zoek het gewenste gen
3. Bekijk de 'General gene information'

1. Het NCBI (National Centre for Biotechnology Information) onderhoudt verschillende databanken en biedt velerlei middelen om te zoeken hierin. "Gene" is een databank met veel informatie over genen. Je kunt hierin de functie van genen vinden, maar ook de locatie op de chromosomen, de pathways waarbij het gen betrokken is, fenotypes en variaties van het gen. Ga naar <http://www.ncbi.nlm.nih.gov/gene> om "Gene" om in de databank te zoeken.
2. Typ je zoekopdracht in het tekstveld, zoals hieronder aangegeven. Eventueel kun je je zoekopdracht verfijnen door een soortnaam toe te voegen.



3. Je krijgt nu een tabel te zien met het gensymbool (=universele code voor ieder gen, bestaande uit een combinatie van letters en cijfers), een korte omschrijving van het gen, de locatie in het genoom en eventuele alternatieve namen. Klik op het gensymbool van het gen naar jouw keuze om een volledig rapport van jouw gen te bekijken.
4. Het rapport over jouw gen bestaat uit verschillende secties. Om de functie van het gen te achterhalen kun je kijken bij de 'Summary'. Gedetailleerdere informatie over de processen waar het gen bij betrokken is, kun je vinden in de sectie 'General gene information'.

Zoek naar de functie van het gen *CONSTANS* in de *Arabidopsis thaliana* (zandraket)

→ Wat is de naam van het eiwit?

1. Ga naar het programma BLAST op de website <http://mrs.cmbi.ru.nl>
2. Kopieer de aminozuurvolgorde in het scherm en klik op Blast
3. Klik op de bovenste 'hit' en vervolgens op de ID-code van de 'hit'.

1. Op internet staan verschillende bioinformatica-tools. Om een basenvolgorde te identificeren, gebruiken we de tool Blast. Blast is een zoekmachine die zoekt in een databank vol aminozuurvolgordes waarvan de functie of herkomst bekend is. Door jouw gezochte aminozuurvolgorde te vergelijken met al die volgordes in de databank, krijg je een aantal 'hits'. De beste hit komt nagenoeg overeen met de door jou ingevoerde aminozuurvolgorde. De conclusie is dat dit gevonden eiwit hetzelfde is als het eiwit dat je hebt ingevoerd.

2. Blast-tools bestaan in verschillende vormen en kun je op verschillende locaties op het internet vinden. Zo is de Blast-tool van Ensembl handig om te gebruiken voor basenvolgordes, maar kun je beter de Blast-tool van de Radboud Universiteit gebruiken voor het identificeren van aminozuurvolgordes. Deze is namelijk makkelijk te hanteren. Je kunt hem in het menu vinden op: <http://mrs.cmbi.ru.nl>

(Afhankelijk van wat jij het prettigst vindt werken, kun je ook bijvoorbeeld die van Ensembl gebruiken: <http://www.ensembl.org/Multi/blastview>. Selecteer in Ensembl *peptide queries* en *protein database*. Automatisch zoekt Ensembl in mensen, selecteer andere soorten om hen ook in de resultaten te betrekken.)

Onderstaande instructies gaan verder vanuit de Radboud-versie.

3. Kopieer de aminozuurvolgorde in het tekstveld. Begin de volgorde met een tekstregel die er uit ziet als *>aminozuurvolgorde 1*. Druk dan op enter en voer de aminozuurvolgorde in. Het gebruik van deze eerste tekstregel heet het FastA-format. Omdat Blast een computerprogramma is, moeten commando's heel precies worden uitgevoerd, anders kan het achterliggende programma niets uitzoeken.
4. Selecteer vervolgens welke database je wilt doorzoeken. Swissprot (afkorting: Swiss Protein) is een grote en bekende eiwitdatabank. Eventueel kun je ook Uniprot (afkorting: Universal Protein) gebruiken.

BLAST-venster: met invulveld, databank-selectie en filterfunctie.

Resultaten venster: hierin staat de status van jouw zoekopdracht

Resultatenlijst: als jouw zoekopdracht voltooid is kun je erop klikken voor de resultaten lijst.

Nr	ID	Coverage	Description	E-value	E-value
1	AMY2_DROAN	100%	1	1066	0.00
2	AMY2B_HUMAN	100%	1	1041	0.00
3	AMYP_HUMAN	100%	1	1037	0.00
4	AMY1_MOUSE	100%	1	932	1.29e-270
5	AMYP_PIG	100%	1	931	2.59e-270
6	AMYP_MOUSE	100%	1	924	3.31e-268
7	AMYP_RAT	100%	1	918	2.12e-266
8	AMYP_STRCA	100%	1	910	5.41e-264
9	AMY_BECMA	100%	1	605	3.54e-172
10	AMN1_DROAN	100%	1	555	2.49e-158
11	AMY1_DROAN	100%	1	555	3.98e-157
12	AMY2_DROAN	100%	1	555	3.98e-157

5. Het 'Filter sequence' kun je aan of uit zetten. Als het filter aan staat, zal hij bepaalde veelvoorkomende aminozuurvolgordes niet meenemen in de vergelijking. Hierdoor krijg je minder hits, omdat eiwitten die alleen op dat domein hetzelfde zijn als de ingevoerde aminozuurvolgorde niet worden meegenomen. Voor het opzoeken van een bekend eiwit (en dat zijn de meeste in het biologieboek), kun je het filter uit zetten.
6. Klik op Blast. De zoekopdracht kan een paar minuten duren, omdat er zo enorm veel aminozuur-volgordes moeten worden vergeleken. Als hij klaar is (status 'finished' in de tabel aan de rechterkant van de pagina), klik je op het resultaat.
7. Blast toont nu een lijst met eiwitten die hij in meer of mindere mate vindt lijken op jouw ingevoerde eiwit. De BitScore en de E-value zijn waardes die aangeven in hoeverre de eiwitten overeen komen. Hoe hoger de BitScore en hoe lager de E-value, hoe betrouwbaarder jouw resultaat. Als je nog twijfelt, kun je op de gekleurde balk van een eiwit klikken. Je krijgt dan een extra regel met informatie. Hierin staat de Identity en de Similarity (beiden geven aan in hoeverre de aminozuren van het gezochte en gevonden eiwit overeen komen) in percentages weergegeven.
 Om de daadwerkelijke aminozuurvolgorde-vergelijking te zien kun je klikken op de zojuist verschenen informatiebalk. De balk klapt nu nog verder open en toont een zogenaamde 'alignment'. Q staat voor Query, dat is de volgorde van jouw ingevoerde aminozuurvolgorde. S staat voor Sequence, dat is de aminozuurvolgorde van het gevonden eiwit in de databank. Doorgestreepte stukken in de volgorde zijn niet meegenomen in de vergelijking. Als een aminozuur in beide eiwitten is gegeven, dan toont Blast dezelfde letter in de tussenliggende regel. Bij een verschil blijft dat een gat. In sommige gevallen zie je een + staan. Dat betekent dat het aminozuur weliswaar anders is, maar dat het verschil tussen de twee aminozuren niet zo groot is (beide aminozuren zijn bijvoorbeeld hydrofoob).
8. Klik op de 'ID' van het eiwit dat het meest lijkt op jouw eiwit. Meestal is dat gewoon het eiwit dat bovenaan de lijst staat. Nu krijg je alle informatie over dat eiwit wat in de databank staat, waaronder: de eiwitnaam (*Protein name*), herkomst (*From*), functie (onder *Comments*, *function of catalytic activity*).

Probeer te achterhalen om welk eiwit het hier gaat:

```
>Protein1
QYSSNTQQGR TSIVHLFEWR WVDIALECER YLAPKGFGGV QVSPNENVA IHNPFPRPWE
RYQPVSYKLC TRSGNEDEFR NMVTRCENVG VRIYVDAVIN HMCNAVSAG 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 IRNMVNFNRV VDGQPFNTWY DNGSNQVAFG
RGNRGFIVFN NDDWTFSLTL QTGLPAGTYC DVISGDKING NCTGIKIYVS DDGKAHFSIS
NSAEDPFIAT HAESKL
```


→ Zijn er organismen met vergelijkbare eiwitten?

1. Ga naar <http://mrs.cmbi.ru.nl>
2. Zoek het eiwit dat je wil gebruiken
- 3a. Klik op 'Find similar'
- 3b. Klik op 'Blast'

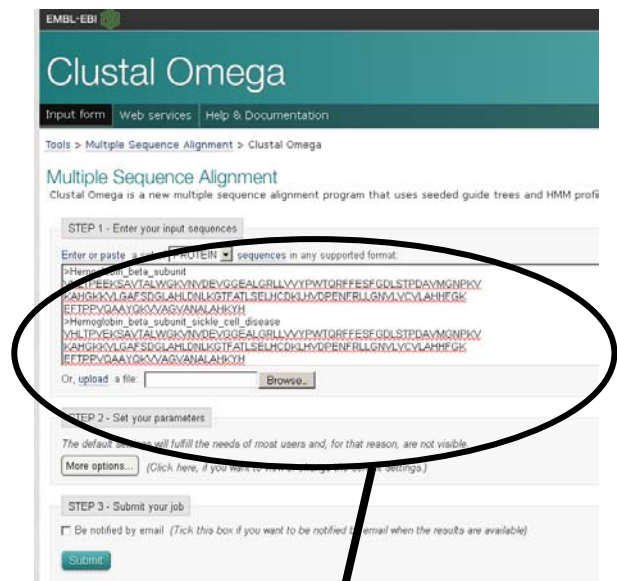
1. Via de website <http://mrs.cmbi.ru.nl> kun je in allerlei databanken zoeken naar eiwitten en genen. Als je zoekt naar een eiwit kun je het beste de databank Swiss-Prot selecteren. Typ in het zoekveld de Engelse naam van het eiwit in.
2. Je krijgt nu een lijst met hits. Deze eiwitten zijn allemaal gerelateerd aan jouw zoekopdracht. Dat kan betekenen dat het het juiste eiwit is, maar ook dat het een eiwit is dat een interactie heeft met het gezochte eiwit. Bekijk dus goed of de korte omschrijving klopt met jouw wensen. De ID kan hierbij helpen. De eerste letters staan voor de afkorting van het eiwit. De letters na het liggend streepje geven het organisme waarin het eiwit voorkomt aan. Als je alleen maar resultaten wilt zien van één soort organisme, voeg dan *os:human* (origin species: human) toe aan de zoekopdracht. Nu vind je alleen maar eiwitten van de mens. Hetzelfde trucje geldt natuurlijk ook voor andere organismen.
Het programma kent ook een soort relevantie toe aan de hits. Hoe verder het balkje gevuld is, hoe hoger de relevantie.
3. Klik op de ID-code van jouw gekozen eiwit. Je ziet nu een overzicht van gegevens die over dit eiwit bekend zijn. Controleer bij de *protein name* of je inderdaad het juiste eiwit hebt gekozen. Als dit het geval is kun je kiezen welke volgende stap je neemt. Beide stappen leveren ongeveer dezelfde resultaten op, maar soms is de ene methode beter dan de andere. Bovenaan de pagina kun je kiezen voor:
 - a. Klik op 'Find similar'. Je ziet nu een overzicht van soortgelijke eiwitten. Je kunt de eiwitten bekijken door op de ID-code te klikken. De eiwitten in deze lijst zijn geselecteerd op basis van hun omschrijving en trefwoorden (overeenkomsten in tekst)
 - b. Klik op 'Blast' en daarna op 'Run Blast'. Klik vervolgens op het resultaat. Je ziet nu een overzicht van soortgelijke eiwitten. Je kunt eiwitten bekijken door op de ID-code te klikken. De eiwitten in deze lijst zijn geselecteerd op basis van hun aminozuurvolgorde (overeenkomsten in sequentie).
4. Bij het bekijken van de gedetailleerde eiwit-informatie kun je vooral kijken naar 'Protein name' en 'From'. Soms kun je ook nuttige termen vinden bij 'Keywords'. Om te kijken of het verwante eiwit dezelfde functie als jouw zoekopdracht heeft, kun je verder naar beneden scrollen en kijken bij 'Function'.

→ **Waarin verschillen deze eiwitten ten opzichte van elkaar?**

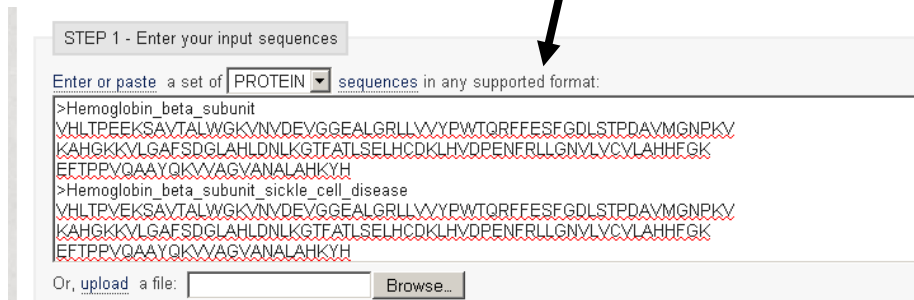
→ **Waarin verschillen deze genen ten opzichte van elkaar?**

1. Ga naar: <http://www.ebi.ac.uk/Tools/msa/clustalo/>
2. Knip en plak de sequenties in het grote invulvlak
3. Klik op 'submit' (blauwe knop links onderaan)
4. Na even wachten wordt de alignment weergegeven

1. Met behulp van een alignment-tool kun je aminozuur- of nucleotidevolgordes met elkaar vergelijken.
Ga naar de website van Clustal Omega: <http://www.ebi.ac.uk/Tools/msa/clustalo/>
2. Dit programma maakt een alignment. Dat wil zeggen, het vergelijkt de verschillende ingevoerde sequenties met elkaar. Voor een alignment heb je minimaal twee sequenties nodig. In het tekstveld typ je eerst *>sequentie1*. Je kunt hierbij zelf de naam bepalen; het is handig om deze overeen te laten komen met het organisme (bijvoorbeeld *>pig*). Let hierbij wel op dat de naam bestaat uit één woord.



3. Je kunt verschillende parameters aanpassen, maar dat is voor een simpele alignment niet nodig. Wel moet je bovenaan het invulveld controleren of het blok op 'Protein' of 'DNA' staat. Klik na het invoeren van de sequenties op 'Submit'. Het programma gaat nu rekenen, dit kan een paar minuten duren.
4. De resultaten zijn opgebouwd uit verschillende onderdelen, aan te klikken met behulp van de tabbladen.
5. Het eerste tabblad 'Alignments' toont de daadwerkelijke alignment. Alle sequenties zijn hier onder elkaar gezet. Als de aminozuren van de verschillende eiwitten overeen komen, of de basen van de verschillende genen overeen komen, dan zie je op de onderste regel een *. Als



ze niet overeen komen, dan staat er niets. Bij sommige verschillen zie je een : of . staan. Dit betekent dat de aminozuren niet overeen komen, maar wel erg veel op elkaar lijken; de aminozuren hebben soortgelijke eigenschappen, ze zijn bijvoorbeeld allebei hydrofoob. De alignment kun je inzichtelijker maken door kleur toe te voegen met behulp van de knop Show Colors.

Het tweede tabblad 'Result Summary' toont een Jalview. Als je op Jalview klikt, kun je de alignment in een andere weergave (in een nieuw venster) bekijken en zien welke gebieden van het eiwit of gen in deze alignment geconserveerd zijn.

Het derde tabblad 'Guide Tree' wordt verder uitgelegd bij *Hoe ziet een fylogenetische boom op basis van dit eiwit eruit?*, p.12.

Probeer eens de beta-subunit van gezond hemoglobine te vergelijken met die van een sikkelcelanemie- patiënt.

```
>Hemoglobin_beta_subunit
VHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPKV
KAHGKKVLGAFSDGLAHL DNLKGTFFATLSELHCDKLHVDPENFRLLGNVLCVLAHFFGK
EFTPPVQAAYQKVVAGVANALAHKYH
>Hemoglobin_beta_subunit_sickle_cell_disease
VHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPKV
KAHGKKVLGAFSDGLAHL DNLKGTFFATLSELHCDKLHVDPENFRLLGNVLCVLAHFFGK
EFTPPVQAAYQKVVAGVANALAHKYH
```

→ Hoe ziet een fylogenetische boom op basis van dit eiwit eruit?

→ Hoe ziet een fylogenetische boom op basis van dit gen eruit?

Zie 'Waarin verschillen deze eiwitten/genen ten opzichte van elkaar?', p.10'. Na de alignment (stap 4) ga je verder met deze instructie:

Klik op het derde tabblad: 'Phylogenetic Tree'.

Je kunt onder 'Phylogram' kiezen uit twee verschillende opties: "Cladogram" of "Real". Standaard laat hij het Cladogram zien. Er is een klein verschil tussen beiden.

Voor een cladogram berekent Clustal Omega het minste aantal 'stappen' (veranderingen in de lettervolgorde) om van een bepaalde sequentie naar een andere te komen. Op basis hiervan maakt het programma een stamboom.

Bij "Real" tekent Clustal Omega een phylogram, dat is bijna hetzelfde als een cladogram, maar hiervoor verwerkt het programma de 'evolutionaire afstanden' tussen eiwitten of genen. Dat wil zeggen: als bijvoorbeeld twee eiwitten erg op elkaar lijken, zullen de stokken korter zijn, dan wanneer er veel verschil is tussen de eiwitten. De berekende afstand tussen de eiwitten kun je ook weergeven met 'Show distances'.

Beide stambomen moeten altijd worden bekeken in het licht van de data die eraan ten grondslag liggen. Als je een eiwit of gen uit de vergelijking haalt, kan dit een compleet andere stamboom geven.

Probeer eens een stamboom te maken aan de hand van het eiwit myoglobine.

```
>mens
MGLSDGGEWQL VLNVWGKVEA DIPGHGQEV L IRLFKGHPET LEKFDKFKHL KSEDEMKASE
DLKKHGATVL TALGGILKKK GHHEAEIKPL AQSHATKHKI PVKYLEFISE CIIQVLQSKH
PGDFGADAQG AMNKALELFR KDMASNYKEL GFQG

>konijn
MGLSDAEWQL VLNVWGKVEA DLAGHGQEV L IRLFHTHPET LEKFDKFKHL KSEDEMKASE
DLKKHGNTVL TALGAILKKK GHHEAEIKPL AQSHATKHKI PVKYLEFISE AIIHVLHSKH
PGDFGADAQA AMSKALELFR NDIAAQYKEL GFQG

>haai
MABWDKVNVS WSAVEQNITA IGQNILLRLF EQYPESEDYF PKLKNKSLGE LKDTADIKAQ
ADTVLRALGN IVKKKGDSHQ PVKALAATHI TTHKIPPHYF TKITTIAVGV LSEMYPSEM N
AQAQAAFSGA FKNICSDIEK EYKAANFQG

>tonijn
MADFDAVLKC WGPVEADYTT MGGLVLTRLF KEHPETQKLF PKFAGIAQAD IAGNAAISAH
GATVLLKLGE LLKAKGSHAA ILKPLANSHA TKHKIPINNF KLISEVLVKV MHEKAGLDAG
GQTALRNVMG IIIADLEANY KELGFSG

>gibbon
MGLSDGGEWQL VLNVWGKVEA DIPSHGQEV L IRLFKGHPET LEKFDKFKHL KSEDEMKASE
DLKKHGATVL TALGGILKKK GHHEAEIKPL AQSHATKHKI PVKYLEFISE CIIQVLQSKH
PGDFGADAQG AMNKALELFR KDMASNYKEL GFQG

>baviaan
MGLSDGGEWQL VLNVWGKVEA DIPSHGQEV L IRLFKGHPET LEKFDKFKHL KSEDEMKASE
DLKKHGATVL TALGGILKKK GHHEAEIKPL AQSHATKHKI PVKYLELISE SIIQVLQSKH
PGDFGADAQG AMNKALELFR NDMAAKYKEL GFQG

>karper
MHDAELVLKC WGGVEADFEG TGGEVLTRLF KQHPETQKLF PKFVGIASNE LAGNAAVKAH
GATVLLKLGE LLKARGDHAA ILKPLATTHA NTHKIALNMF RLITEVLVKV MAEKAGLDAG
GQSALRRVMD VVIGDIDTY Y KEIGFAG

>zebra
MGLSDGGEWQQ VLNVWGKVEA DIAGHGQEV L IRLFTGHPET LEKFDKFKHL KTEAEMKASE
DLKKHGTVVL TALGGILKKK GHHEAELKPL AQSHATKHKI PIKYLEFISD AIIHVLHSKH
PGDFGADAQG AMTKALELFR NDIAAKYKEL GFQG
```

→ Hoe ziet de 3D-structuur van het eiwit eruit?

1. Download de pdb-file van het eiwit via bijvoorbeeld <http://www.pdb.org>
 2. Open het bestand in Yasara
-
1. Een pdb-bestand is het meest gebruikte type bestand om eiwitten in 3D te bekijken. Dit type bestanden kun je downloaden vanuit de databank PDB (Protein Data Bank). Behalve de PDB-databank zijn er andere mogelijkheden om pdb-bestanden te downloaden. Bijvoorbeeld via Google (www.google.nl): Typ de naam van het gewenste eiwit in en voeg hierbij de zoekterm 'pdb'.
 2. Typ bij 'Search' het gewenste eiwit in (Engels). Als jouw zoekopdracht geen goede resultaten oplevert, kun je meer zoektermen invoeren. Bijvoorbeeld '*lipase human*'. Je kunt ook kiezen uit de '*Query refinements*' en kiezen voor een organisme of publicatiedatum. Van lang niet alle eiwitten is de structuur bekend, dus soms is het eiwit ook helemaal niet te vinden in de databank.

Iedere Query in de databank heeft een eigen ID-code. Bijvoorbeeld 1H76 voor Transferine, een carrier van ijzer-ionen in het bloed. Door deze code in te typen in het zoekveld van de databank, krijg je direct de juiste hit. Dat is handig als je vanuit een andere databank of wikipagina deze ID-code voorhanden hebt.
 3. Klik op het zoekresultaat waarnaar je zocht, bijvoorbeeld op de eerste hit. Rechts zie je een kleine afbeelding van het eiwit. Zo kun je snel controleren of je het juiste bestand te pakken hebt. Bij '*Molecular description*' staat de inhoud van het bestand (*Molecule*). Klik rechtsboven op 'Download files' en vervolgens op 'PDB file (text)' en sla het bestand op.
 4. Open het programma Yasara en open het gedownloade bestand. Instructies om Yasara te downloaden en te configureren kun je vinden op pagina 14.

Tip: Bekijk ook eens de sectie 101 van de PDB-database

(http://www.pdb.org/pdb/101/structural_view_of_biology.do). Hierop zijn veel interessante eiwitten te vinden met een uitgebreide en begrijpelijke uitleg. Bijvoorbeeld over DNA ligase:

<http://www.pdb.org/pdb/101/motm.do?momID=55>

→ Ik weet de structuur van het eiwit; en wil die bekijken

Algemene Yasara handleiding

Met Yasara kun je op de computer laten zien hoe een eiwit er in 3D uitziet. Als de school het programma Yasara niet heeft, kun je het gratis downloaden en installeren:

- 1 Ga naar www.yasara.org en klik op 'Products' in het menu.
- 2 Kies vervolgens 'freely download now' achter 'Yasara View'.
- 3 Vul het formulier in. Bij 'department' kun je de naam van jouw school invullen. Jouw e-mailadres wordt alleen gebruikt voor het verzenden van de download-link.
- 4 Via een email krijg je een download-link, hiermee kun je Yasara in de gewenste map op jouw computer plaatsen.
- 5 Installeer Yasara volgens de instructies.
- 6 Eventueel kun je aanvullende informatie (in het Engels) over Yasara vinden op: <http://www.cmbi.ru.nl/~hvensela/yasara/>

Met Yasara kun je veel bekijken en bedienen. Hieronder tref je de meest gebruikte functies.

Bediening/configuratie

Om het molecuul te draaien gebruik je de muis en de pijltjestoetsen op het toetsenbord, houdt de linkermuisknop ingedrukt en beweeg de muis.

Om in- en uit te zoomen houd je de rechtermuisknop ingedrukt en beweeg je de muis naar voren en achter.

Met behulp van de pijltjestoetsen kan het molecuul bewegen over het scherm

Bestanden inladen

Yasara herkent verschillende typen bestanden. Deze kun je herkennen aan de uitgang (de extensie, bijvoorbeeld bestanden die Word kan lezen altijd eindigen op .doc): pdb, sce, yob. PDB-bestanden worden het meeste gebruikt om eiwitten in te laden in visualisatie-programma's.

Om een bestand te openen in Yasara klik je op 'File' en vervolgens op 'Load'. Dan kun je kiezen welk type bestand je opent, bijvoorbeeld 'PDB file' voor een pdb-bestand of 'YASARA Scene' voor een sce-bestand. In het menu dat opent kun je het bestand opzoeken en openen door op 'OK' te klikken.

Als je klaar bent met het bekijken van het eiwit en je wilt een nieuw eiwit openen, klik dan eerst op 'File' en dan 'New', voordat je het nieuwe bestand inlaadt. Hiermee sluit je alle openstaande bestanden af en krijg je geen verkeerde weergaves.

Verschillende weergaves

Yasara kan verschillende weergaves van het eiwit laten zien. Gebruik hiervoor de toetsen F1 t/m F8

- F1 Ball-weergave
- F2 Ball-and-stick-weergave
- F3 Stick-weergave
- F4 Trace-weergave
- F5 Tube-weergave
- F6 Ribbon-weergave
- F7 Cartoon-weergave

Met F8 kun je in iedere bovengenoemde weergave de restgroepen van de aminozuren zichtbaar of onzichtbaar maken.

Sommige bestanden zijn zo gemaakt, zodat bepaalde delen van het molecuul goed zichtbaar zijn. Deze instellingen wis je door gebruik te maken van de verschillende weergaves. Wanneer je de vooraf ingestelde weergave weer terug wil krijgen, kun je het bestand opnieuw openen.

Extra opties

Aankleuren van negatief geladen aminozuren	<ul style="list-style-type: none"> - Zet de weergave van waterstof aan (<i>Edit > Add > Hydrogenst o All</i>). - Selecteer <i>view > color > residue</i> - Selecteer in de derde kolom (<i>belongs to or has Charge < 0</i>) en klik op <i>Ok</i>. - Kies een kleur en klik <i>Apply Unique color</i>.
Aankleuren van positief geladen aminozuren	<ul style="list-style-type: none"> - Zet de weergave van waterstof aan (<i>Edit > Add > Hydrogenst o All</i>). - Selecteer <i>view > color > residue</i> - Selecteer in de derde kolom (<i>belongs to or has Charge > 0</i>) en klik op <i>Ok</i>. - Kies een kleur en klik <i>Apply Unique color</i>.
Aankleuren van aminozuren met hydrofiele restgroepen	<ul style="list-style-type: none"> - Selecteer <i>view > color > residue</i> - Selecteer in de tweede kolom Arg, Asp, Asn, Glu, Gln, His, Lys, Ser en Thr (houdt Ctrl ingedrukt) en klik op <i>Ok</i>. - Kies een kleur en klik <i>Apply Unique color</i>.
Aankleuren van aminozuren met hydrofobe restgroepen	<ul style="list-style-type: none"> - Selecteer <i>view > color > residue</i> - Selecteer in de tweede kolom Ile, Leu, Met, Phe en Val (houdt Ctrl ingedrukt) en klik op <i>Ok</i>. - Kies een kleur en klik <i>Apply Unique color</i>.
Weergeven waterstofmoleculen	Selecteer <i>Edit > Add > Hydrogens to All</i>
Weergeven waterstofbruggen	<ul style="list-style-type: none"> - Selecteer <i>Edit > Add > Hydrogens to All</i> - Selecteer <i>View > show interactions > Show hydrogen bonds > All</i>
Weergeven secundaire structuren	F6
Weergeven restgroepen	F8
Verwijderen (ruis van) water	Selecteer <i>Edit > Delete > Waters</i>

Vanaf hier kun je doorlezen voor het beantwoorden van de vragen:

- ➔ (enzym) *Waar bindt het substraat?* – p.16
- ➔ (receptor/kanaal) *Waarmee hecht het eiwit zich aan het membraan?* – p.17

→(enzym) Waar bindt het substraat?

1. Zoek of bepaal de structuur van het substraat
2. Open het eiwit in Yasara
3. Bepaal waar het substraat kan binden, bijvoorbeeld in een holte
4. Zoek mogelijkheden van interactie tussen het enzym en het substraat

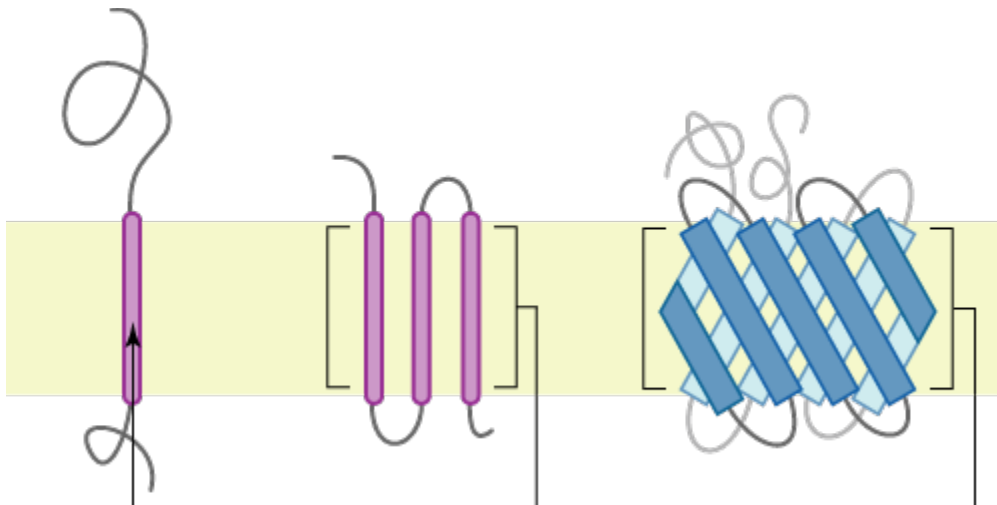
1. Ieder enzym breekt een substraat (of ligand) af. Via Google, Wikipedia of het biologieboek kun je erachter komen welk substraat jouw gekozen enzym afbreekt. Vaak wordt in Wikipedia de vorm van het substraat gegeven. Of je kunt met Google zoeken op afbeeldingen. Het gaat erom dat je achterhaalt hoe groot het substraat ongeveer is en welke vorm het heeft.
2. Open het enzym in Yasara.
3. Het substraat bindt vaak in een holte aan het enzym, zo heeft het substraat meer bindingsoppervlak met het enzym. Zoek een holte in het enzym.
Kun je geen holte vinden? Zoek dan naar andere opvallende structuren in het eiwit. Het vinden van het actieve centrum aan de hand van enkel de structuur is niet altijd eenvoudig. Vaak heb je andere bronnen nodig om de juiste plek in het enzym te bepalen.
4. Controleer of de gevonden holte ook het daadwerkelijke actieve centrum is. Dat kun je doen door te zoeken naar mogelijke bindingsplekken. Als er bijvoorbeeld een ion in het actieve centrum zit en een ion met tegenovergestelde lading in het substraat, dan heb je hoogstwaarschijnlijk de goede locatie te pakken. Andere interacties die je kunt zoeken tussen atomen in het actieve centrum van het enzym en het substraat zijn waterstofbruggen, zwavelbruggen en hydrofobe interacties.
5. Van sommige, heel erg bekende, eiwitten kun je het pdb-bestand downloaden waaraan het substraat reeds gebonden is. Zoek hiernaar in de PDB-databank, of via Google.

Probeer het actieve centrum te vinden van het eiwit lactase. Dit eiwit breekt lactose (melksuiker) af en kun je vinden in de PDB-databank met de ID-code: 3E1F

→(receptor/kanaal) Waarmee hecht het eiwit zich in het membraan?

1. Open het eiwit in Yasara
2. Druk op F6
3. Bekijk welke typische secundaire structuurelementen door het membraan kunnen gaan

1. Open het receptor- of kanaal-eiwit in Yasara
2. Druk op F6, dit is de sneltoets voor de zogenaamde cartoon-weergave. In deze weergave zie je duidelijk welke secundaire structuren er zijn: α -helices en β -sheets. Een helix is blauw, een sheet rood.
3. Er zijn drie structuren bekend waarmee eiwitten door een membraan steken. Een enkele helix, een helix-bundel of een β -barrel (zie de figuur hieronder).
Een helix-bundel is een bundel α -helices bij elkaar, die gezamenlijk door het membraan gaan. Bij de vouwing van het eiwit kunnen de aparte helices al door het membraan steken. Doordat het energetisch gunstiger is om samen door het membraan te gaan, klitten ze samen tot een bundel. De grootte van helix-bundels kan variëren van 3 helices tot enkele tientallen.
Een β -barrel bestaat uit β -strands die om elkaar heen draaien en zo een kanaal vormen.



Een schematische weergave van een enkele helix, een helix-bundel en een β -barrel

Bron afbeelding: http://en.wikipedia.org/wiki/Transmembrane_protein

Zoek de helix-bundel in de acetylcholinereceptor. PDB ID-code: 2BG9

→ Wat is de naam van het gen?

1. Ga naar de website <http://www.ensembl.org/Multi/blastview>
 2. Kopieer de basenvolgorde in het scherm en klik op RUN
 3. Bekijk de contigview van de bovenste 'hit' om het gen te achterhalen
-
1. Op internet staan verschillende bioinformatica-tools. Om een basenvolgorde te identificeren, gebruiken we de tool Blast. Blast is een zoekmachine die zoekt in een databank vol basenvolgordes waarvan de functie of herkomst bekend is. Door jouw gezochte basenvolgorde te vergelijken met al die volgordes in de databank, krijg je een aantal 'hits'. De beste hit komt nagenoeg overeen met de door jou ingevoerde basenvolgorde. De conclusie is dat dit gevonden gen hetzelfde is als het gen dat je hebt ingevoerd. Hier zitten een aantal stappen tussen die vrij ingewikkeld worden. Een gen is niet altijd eenduidig een gen, omdat het allerlei exonen, intronen, transcripten etc. bevat. Deze handleiding geeft een globaal overzicht van de stappen die je moet nemen.
 2. Blast-tools bestaan in verschillende vormen en kun je op verschillende locaties op het internet vinden. Zo is de Blast-tool van Ensembl handig om te gebruiken voor basenvolgordes, maar kun je beter de Blast-tool van <http://mrs.cmbi.ru.nl> gebruiken voor het identificeren van aminozuurvolgordes.
Ga naar de website van Ensembl om jouw basenvolgorde te identificeren:
<http://www.ensembl.org/Multi/blastview>
 3. Kopieer de basenvolgorde in het tekstveld:
Begin de volgorde met een tekstregel die er uit ziet als *>basenvolgorde 1*. Druk dan op enter en voer de basenvolgorde in. Het gebruik van deze eerste tekstregel heet het FastA-format. Omdat Blast een computerprogramma is, moeten commando's heel precies worden uitgevoerd, anders kan het achterliggende programma niets uitzoeken.
Deze Blast-tool heeft ook een functie waarbij je een referentiecode kunt invoeren van een andere databank. Als je bijvoorbeeld een gen hebt gevonden in een databank, maar deze verder wilt analyseren met behulp van Blast, dan kun je de ID-code van het gen invoeren bij *Enter a sequence ID or accession (EMBL, UniProt, RefSeq)*. Klik vervolgens op *Retrieve*.
 4. Vervolgens kun je verschillende opties uitvoeren:
 - *Select the databases to search against* – Omdat deze Blast-tool ook een alignment kan uitvoeren (zie *Welke organismen hebben verwante genen?*) kun je hier kiezen in welke databanken je de ingevoerde basenvolgorde wil vergelijken. Voor het identificeren van een gen hoef je hier niets te selecteren.
 - *Select te Search Tool* – Omdat er verschillende versies van Blast bestaan (de ene Blast zoekt op een net iets andere manier dan de andere Blast), kun je net iets verschillende resultaten krijgen. Automatisch staat de Search Tool op BLAT (Blast Like Alignment Tool). Deze Blast is het snelste en kun je aanhouden voor jouw zoekopdracht. Andere Search Tools zijn BLASTN (Blast Nucleotides) en TBLASTX (Translate Blast X). Beiden zijn relatief langzaam, omdat zij extra optimalisatie-berekeningen doen.
 - *Search sensitivity* – Je kunt de zoekresultaten aanpassen met dit selectie-menu. Voor het identificeren van een gen, kun je hem laten staan op *Near-exact matches* of als je bij deze optie te veel resultaten krijgt op *Exact matches*. Bij te weinig resultaten kun je eventueel

kiezen voor *No optimisation*, maar dan loop je het risico dat je heel veel zoekresultaten krijgt die niets te maken hebben met de ingevoerde basenvolgorde.

- Als alle opties op het scherm staan zoals je wenst, kun je op RUN klikken. De zoekopdracht kan een paar minuten duren, omdat het programma bezig is met enorm veel basenvolgorde te doorzoeken. Als hij klaar is, krijg je een nieuw scherm.

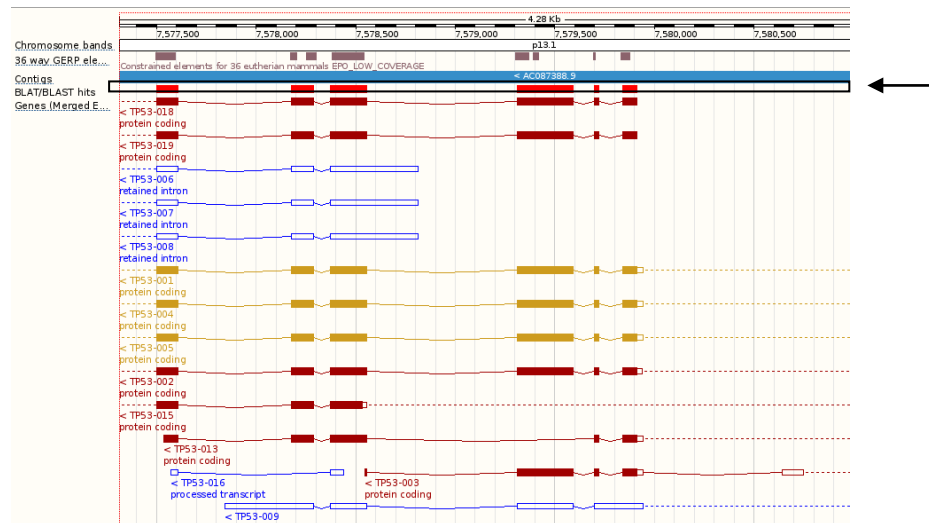
Dit scherm begint met *Alignment Locations vs. Karyotype*. De rode pijlen geven plekken aan waar op de chromosomen een gen zit dat overeenkomt met jouw ingevoerde basenvolgorde. Het pijltje dat een rood vakje om zich heeft staan, lijkt het meeste op jouw ingevoerde volgorde.

Daarna volgt *Alignment Locations vs. Query*. Deze geeft de locatie van de HSPs (High-scoring Sequence Pairs) aan. Dit zijn stukjes van de basenvolgorde die Blast gebruikt voor het zoeken in de databank. Dit overzicht heb je niet nodig en kun je desgewenst wegklikken met het pijltje.

Tot slot volgt de *Alignment Summary*. Bovenaan staat de beste hit. Dit kun je controleren door te kijken naar de *Stats*, helemaal rechts in beeld. Hoe hoger de *Score*, hoe beter de match. *%ID* geeft aan hoeveel procent van jouw ingevoerde volgorde overeenkomt met de volgorde in de databank. De *Length* geeft de lengte van het overeenkomstige deel aan. Helemaal aan de linkerkant van een hit staan de letters A, S, G en C.

Klik op C (Contigview) bij de beste hit.

- Een nieuw scherm opent met een overzicht van de regio op het chromosoom.



Als je iets naar beneden scrolt krijg je een schematisch overzicht van genen die lijken op jouw ingevoerde gen. De rode balkjes zijn de hit die je hebt geselecteerd (in dit diagram aangegeven met de zwarte pijl). De stukken DNA eronder zijn genen of stukken van genen. Donkerrood staat voor eiwit-coderende genen. Aan de linkerkant van het schema staat de naam van een betreffend gen. Klik op de naam van het gewenste gen en klik vervolgens op de code naast *Gene*.

- Bovenaan de nieuwe pagina staat welk gen we geselecteerd hebben. Hieronder staat de locatie op het chromosoom en daarna hoeveel transcripten wetenschappers van dit gen kennen.

Probeer dit eens voor deze basenvolgorde:

>Basenvolgorde1

```

ATGGAGGAGCCGCAGTCAGATCCTAGCGTCGAGCCCCCTCTGAGTCAGGAAACATTTTCAGACCTATGGA
AACTACTTCTGAAAACAACGTTCTGTCCCCCTTGCCGTCCAAGCAATGGATGATTTGATGCTGTCCCC
GGACGATATTGAACAATGGTTCACTGAAGACCCAGGTCCAGATGAAGCTCCCAGAATGCCAGAGGCTGCT
CCCCCGTGGCCCCCTGCACCAGCAGCTCCTACACCGGCGGCCCTGCACCAGCCCCCTCCTGGCCCCCTGT
CATCTTCTGTCCCTTCCCAGAAAACCTACCAGGGCAGCTACGGTTTCCGTCTGGGCTTCTTGCATTCTGG
GACAGCCAAGTCTGTGACTTGCACGTA CTCCCCCTGCCCTCAACAAGATGTTTGGCCAACTGGCCAAGACC
TGCCCTGTGCAGCTGTGGGTTGATTCCACACCCCCGCCCGCACCCGCTCCGCGCCATGGCCATCTACA
AGCAGTCACAGCACATGACGGAGGTTGTGAGGCGCTGCCCCACCATGAGCGCTGCTCAGATAGCGATGG
TCTGGCCCCCTCCTCAGCATCTTATCCGAGTGGAAGGAAAATTTGCGTGTGGAGTATTTGGATGACAGAAAC
ACTTTTCGACATAGTGTGGTGGTGCCCTATGAGCCGCTGAGGTTGGCTCTGACTGTACCACCATCCACT
ACA ACTACATGTGTAACAGTTTCTGCATGGGCGGCATGAACCGGAGGCCCATCCTCACCATCATCACACT
GGAAGACTCCAGTGGTAATCTACTGGGACGGAACAGCTTTGAGGTGCGTGTTTGTGCCTGTCTGGGAGA
GACCGGCGCACAGAGGAAGAGAATCTCCGCAAGAAAGGGGAGCCTCACCACGAGCTGCCCCAGGGAGCA
CTAAGCGAGCACTGCCAACAACACCAGCTCCTCTCCCAGCCAAAGAAGAAACCACTGGATGGAGAATA
TTTACCCTTCAGATCCGTGGGCGTGAGCGCTTCGAGATGTTCCGAGAGCTGAATGAGGCCTTGGA ACTC
AAGGATGCCCAGGCTGGGAAGGAGCCAGGGGGGAGCAGGGCTCACTCCAGCCACCTGAAGTCCAAAAAGG
GTCAGTCTACCTCCCGCCATAAAAAACTCATGTTCAAGACAGAAGGGCCTGACTCAGACTGA
    
```

Let op: dit is alleen het coderende gedeelte van het gen (alleen exons)

- Wat is de basenvolgorde van het gen?
- Hoeveel transcripten zijn er van dit gen?
- Hoeveel coderende gebieden heeft dit gen (introns/exons)?
- Wat is de chromosoomlocatie van dit gen?

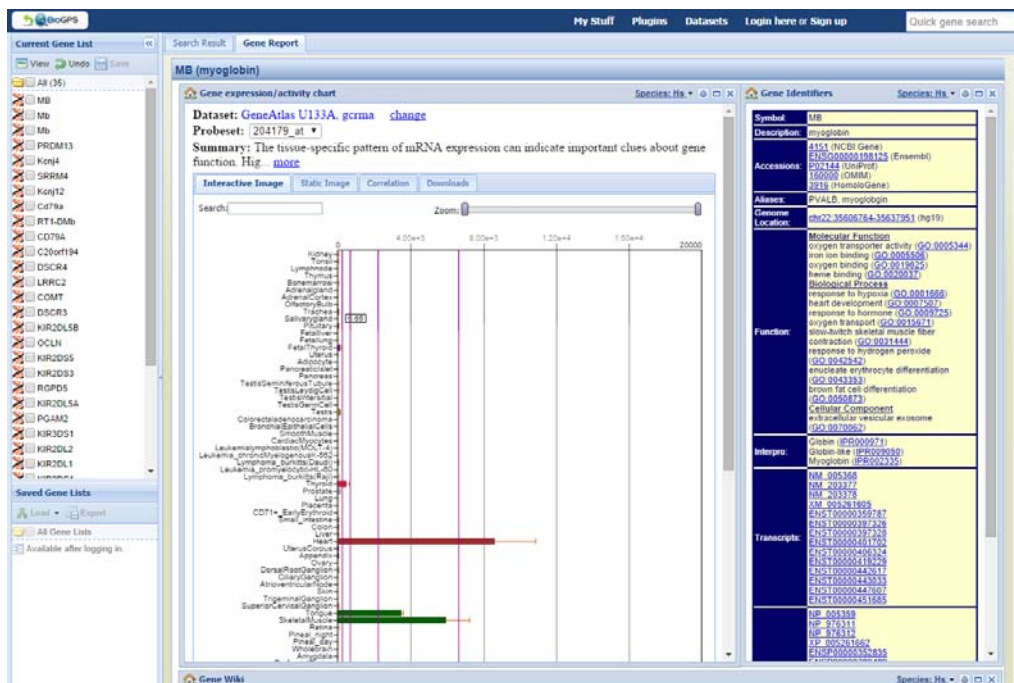
1. Ga naar: <http://www.ensembl.org>
2. Type de naam van het gen in het tekstveld
3. Bekijk de beste hit onder 'gene'
4. Voor de basenvolgorde: klik op 'Sequence' in het menu

1. Ensembl is een zogenoemde genome browser waarin je allerlei informatie kunt opzoeken over genen. Let op: in Ensembl kun je alleen zoeken in vertebraten en andere eukaryoten. Planten zijn bijvoorbeeld niet opgenomen in de databanken.
Ga naar <http://www.ensembl.org>
2. Gebruik de zoekfunctie op de hoofdpagina van Ensembl om het gen van jouw interesse te vinden. Je kunt hiervoor verschillende zoektermen gebruiken, zoals de naam van het gen, het gensymbool of de coördinaten van de locatie van het gen. Klik op 'Go' om de zoekopdracht te starten.
3. Je krijgt nu een overzicht van de gevonden items. Zoek naar de naam van jouw gen, met daarachter (*Human Gene*). Als je een gen zoekt van een ander organisme dan de mens, moet je uiteraard uitkijken naar (*soortnaam Gene*). Klik op dit item voor meer informatie.
4. Het informatieblad wat je nu ziet geeft allerlei informatie.
De bovenste sectie geeft een omschrijving van het gen, synoniemen en de locatie.
Vervolgens staan in een tabel alle bekende transcripten van het gen. Hier staat vermeld hoeveel baseparen (bp) het transcript lang is en of het transcript codeert voor een eiwit (*protein coding*). Als het transcript codeert voor een eiwit, staat vermeld uit hoe veel aminozuren dat eiwit bestaat. Er kan ook *No protein* staan, en dan staat vermeld wat er dan gebeurt met het transcript. Met de muis kun je boven de tekst zweven voor meer uitleg.
Onder de tabel staat een *Summary* dat begint met het gensymbool. Vervolgens zijn er verschillende links die verwijzen naar andere databases of aangeven welke methode is gebruikt om te voorspellen of het gen ergens voor codeert.
Onderaan de pagina staat een afbeelding waarin schematisch is weergegeven waar de transcripten zich bevinden. Niet alleen jouw gen staat hierin afgebeeld, maar ook genen die in de buurt liggen. Het gen is weergegeven als een lijn met blokken. De blokken zijn exonen, de tussenliggende lijn zijn intronen.
5. Om de basenvolgorde van een transcript te achterhalen, moet je eerst op de transcript ID van het gewenste transcript klikken. Vervolgens kun je in het menu aan de linkerkant klikken op *cDNA*, dit staat onder *Sequence*. Onderaan de pagina verschijnt nu de basenvolgorde, met de daarbij horende aminozuurvolgorde.

Probeer het gen *TP53* in de mens te vinden, en zoek de basenvolgorde, het aantal transcripten, het aantal exonen en de chromosoomlocatie.

→ In welk weefsel is het gen actief?

1. Ga naar <http://biogps.org>
 2. Zoek naar het gen met behulp van het gensymbool
 3. Bekijk het RNA expressiepatroon in de grafiek.
1. Als we kijken naar de activiteit van een gen, dan kijk je naar het voorkomen van RNA op een bepaalde plek in het lichaam. Wordt er veel RNA gevonden in bepaald weefsel, dan wordt het gen op die plek vaak afgelezen. Dit noemt men expressie. Een expressieprofiel van ene gen kun je op verschillende plekken vinden. Op een gen-pagina in de Engelse Wikipedia kun je bijvoorbeeld in de kolom aan de rechterzijde van de pagina het expressieprofiel zien (*RNA expression pattern*). Je kunt ook direct in BioGPS kijken. Ga naar <http://biogps.org>.
2. Voer het gensymbool in van jouw gewenste gen. Als je het gensymbool niet weet, kun je via www.google.nl zoeken op de naam van jouw gen en 'gensymbool'. Meestal komt het symbool dan ergens in de eerste hits tevoorschijn. Het gaat om een afkorting van letters (en cijfers). Druk op *Search*. Kies vervolgens de juiste hit in de tabel. In de laatste kolom staat de soortnaam.
 3. Er verschijnt een grafiek met balken. Op de Y-as staan verschillende weefsels waarin de expressie is gemeten. Op de X-as staat de mate van expressie.



In welke weefsels komt myoglobine (gensymbool: MB) vooral tot expressie?

→ Zijn er organismen met vergelijkbare genen?

1. Ga naar <http://www.ensembl.org>
 2. Type de naam van het gen in het tekstveld
 3. Bekijk de beste hit onder 'gene'
 4. Klik op 'Orthologues' in het menu aan de linkerkant
 5. Selecteer van welke soortgroepen je gedetailleerde informatie wilt zien
-
1. Ensembl is een zogenoemde genome browser waarin je allerlei informatie kunt opzoeken over genen. Let op: in Ensembl kun je alleen zoeken in vertebraten en andere eukaryoten. Planten zijn bijvoorbeeld niet opgenomen in de databanken.
Ga naar <http://www.ensembl.org>
 2. Gebruik de zoekfunctie op de hoofdpagina van Ensembl om het gen van jouw interesse te vinden. Je kunt hiervoor verschillende zoektermen gebruiken, zoals de naam van het gen, het gensymbool of de coördinaten van de locatie van het gen. Klik op 'Go' om de zoekopdracht te starten.
 3. Je krijgt nu een overzicht van de gevonden items. Zoek naar de naam van jouw gen, met daarachter (*Human Gene*). Als je een gen zoekt van een ander organisme dan de mens, moet je uiteraard uitkijken naar (*soortnaam Gene*). Klik op dit item voor meer informatie.
 4. Het informatieblad wat je nu ziet geeft allerlei informatie. Om te vinden welke organismen vergelijkbare genen hebben, moet je op *Orthologues* klikken in het menu aan de linkerkant, deze kun je vinden onder *Comparative Genomics*.
 5. Onderaan de pagina verschijnt nu een tabel met groepen. Je kunt per groep selecteren (*Show details*) of je meer informatie wilt zien. Als je iets aanvinkt, verschijnt nog een tabel met een lijst van soortnamen van soorten die een soortgelijk gen hebben als het gen dat jij hebt ingevoerd.

Probeer te vinden welke organismen allemaal het gen voor rhodopsine (betrokken bij het detecteren van licht; zien) hebben.