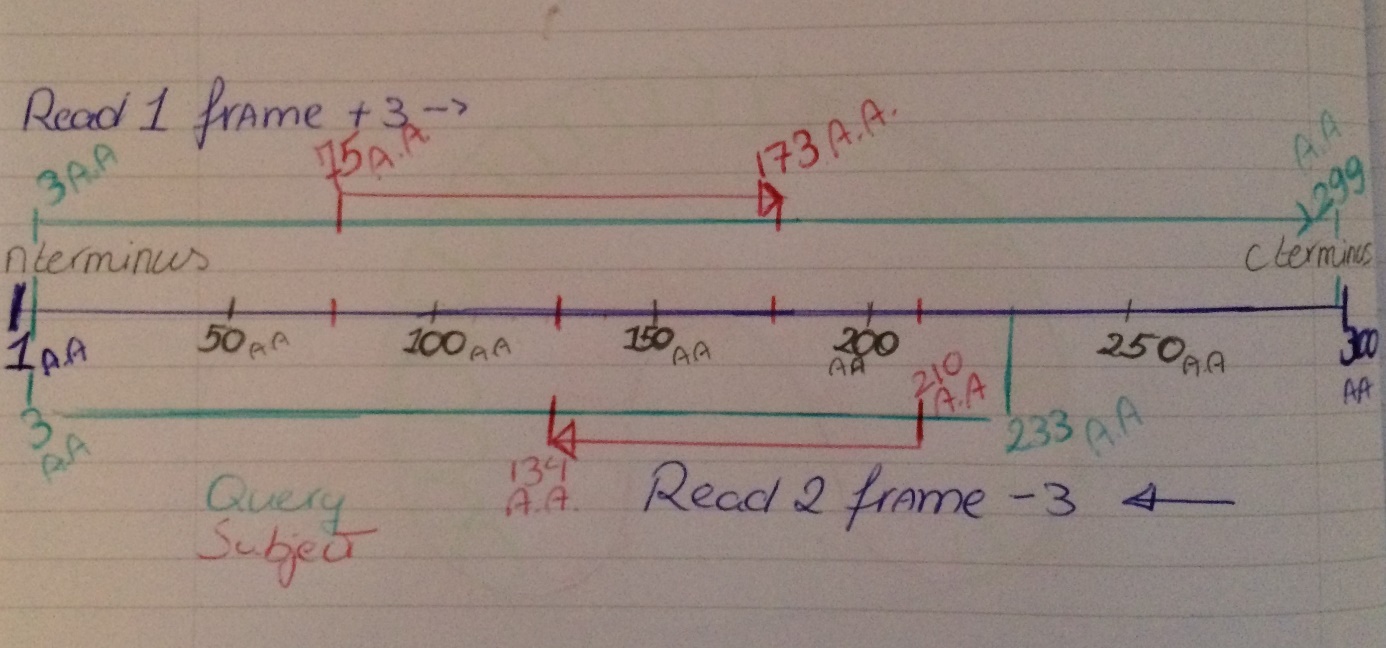
Resultaten onderzoeksverslag

Het doel was om 200 sequenties van annotatie te voorzien door middel van een applicatie die de 200 sequenties geautomatiseerd kan blasten en ze vervolgens in een database kan laden. Deze applicatie kan geraadpleegd worden via een site waarmee de database doorzocht kan worden. Eventueel kunnen blast resultaten die via de site zijn verkregen ook worden ingeladen in de database. Het uiteindelijke doel van dit onderzoek was om de champignonteelt te stabiliseren door meer inzicht te krijgen in welke micro-organismen zich in de verschillende stadia van de compost bevinden. Een van deze micro-organismen met geproduceerd eiwit wordt verder toegelicht, zoals dit in het verdere stadium van dit onderzoek gedaan zou kunnen worden met alle geannoteerde sequenties.

Uit de dataset zijn 222 organismen en 222 eiwitten geïdentificeerd. De 3 meest voorkomende organsimen zijn: *Paenibacillus amylolyticus* (komt 5 keer voor), *Methanosarcina mazei* (komt 4 keer voor) en *Treponema bryantii* (komt 3 keer voor). De 3 meest voorkomende eiwitten zijn: molybdenum cofactor *biosynthesis protein B* (komt 42 keer voor), *sugar ABC transporter permease* (komt 42 keer voor) en *purine-nucleoside phosphorylase* (komt 42 keer voor).

Het organisme waar verder op ingegaan zal worden is een van de 200 sequenties dat door Illumina MiSeq technologie als paired-end reads is gesequenst, vervolgens geblast met de module Blastx en zo van annotatie voorzien. In de database is te zien dat het organisme *Actinotalea fermentans ATCC 43279 = JCM 9966 = DSM 3133 (accession: KGM15509)* regelmatig voorkomt. In de annotatie is te zien dat de blast resultaten een hit hebben gegeven op het eiwit dat dit organisme produceert: 2-isopropylmalate synthase, geproduceerd door het leuA gen. In het onderstaande figuur (figuur 1) is te zien hoe de forward en reverse read op de sequentie van de gevonden hit liggen. Met in het groen de query en in het rood het subject.



Figuur 1 Overzicht van de forward en reverse read, hoe deze ten opzichte van de sequentie van de hit liggen.

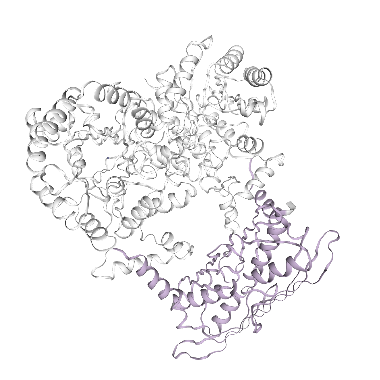
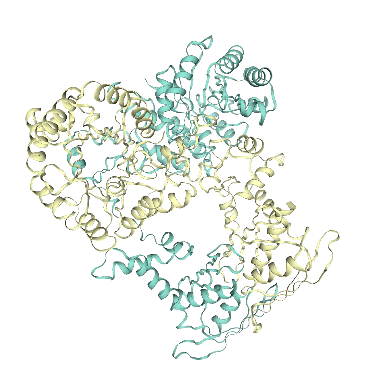
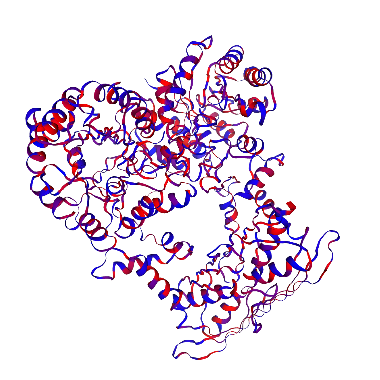
Het organisme *Actinotalea fermentans* behoort tot het rijk van de bacteriën en is onderverdeeld in de stam van *terrabacteriën*, die er om bekend staan dat ze goed bestand zijn tegen milieugevaren zoals uitdroging, ultraviolette straling en hoge zoutgehaltes. Verder behoort de bacterie tot de klasse van de *actinobacteria*, de order van de *micrococcineae* en de familie van de *cellulomonadaceae*, die behoren tot de groep grampositieve bodemorganismen (Nightingale A, 2017). Geslacht en soort zijn van de naam af te leiden. De cellular component van deze bacterie beslaat alle structuren van een cel uitgesloten van het plasma membraan en de nucleus. De bacterie is zo goed als onschadelijk voor een gezonde volwassene en kan in een BioHazard level 1 laboratorium gehouden worden. *Actinotalea fermentans* is een aeroob organisme wat inhoud dat het zuurstof nodig heeft voor zijn energie productie. Zuurstof wordt gebruikt in cellulaire processen zoals de Kreb’s cycle en de elektronen transport om ATP te produceren. Wanneer de zuurstofatomen niet meer nodig zijn worden er steeds twee zuurstofatomen aan een carbonatoom gebonden, zodat het makkelijk als CO2 vervoerd kan worden. Het Organisme heeft een maximale activiteit bij 45 graden Celcius en bij een Ph van 8.

Het eiwit *2-isopropylmalate synthase* heeft 2 GO termen: het ‘Biologisch proces’ en de ‘Moleculaire functie’. Het biologische proces van het eiwit resulteert in de chemische reactie die leidt tot de vorming van het aminozuur leucine: 2-amino-4-methylpentanoic acid(leuA). *2-isopropylmalate synthase* is niet de enige katalysator, er zijn meerdere proteïnes betrokken bij het omzetten van 3-methyl-2-oxobutanoate + acetyl-CoA naar leucine. De volgorde is parallel met de L-leucine biosynthesis pathway: [*3-isopropylmalate dehydratase small subunit (****leuD****)*](http://www.uniprot.org/uniprot/A0A0A0C6Q5)*,* [*3-isopropylmalate dehydratase large subunit (****leuC****)*](http://www.uniprot.org/uniprot/A0A0A0C8C0)*,* [*3-isopropylmalate dehydrogenase (****leuB****)*](http://www.uniprot.org/uniprot/A0A0A0CB52) *en* [*Branched-chain-amino-acid aminotransferase**(Accession:* ***N867\_09825****)*](http://www.uniprot.org/uniprot/A0A0A0CAF9)*. 2-isopropylmalate synthase* wordt door de ribosomen van de cel geproduceerd.

Het eiwit *2-isopropylmalate synthase* is de 1e katalysator in de L-leucine biosynthesis pathway. De moleculaire functie van het eiwit is het katalyseren van de eerste stap in de L-leucine biosynthesis pathway zoals in de reactievergelijking (een aldol condensatie) is weergegeven:

*3-methyl-2-oxobutanoate + acetyl-CoA + H2O 🡪(2S)-2-isopropylmalate + CoA + H+*

In de bijlagen is een figuur weergegeven met daarin de complete reactie van *2-isopropylmalate synthase* naar leucine (Bijlage2) en de GO termen die horen bij deze reactie(Bijlage3). In figuur 3 is te zien dat *2-isopropylmalate synthase* uit 2 unieke groepen bestaat (geel en groen), wat opvalt is dat de gele chain een homo2-mer is en de groene chain ook uit 2 keer dezelfde polypeptiden bestaat. Dit is logisch want *2-isopropylmalate synthase* heeft een subunit structure genaamd homotetramer (Chen F., 2017). In figuur 4 is de hydrofobiciteit van het eiwit te zien.



Figuur 2. 2-isopropylmalate synthase LeuA,

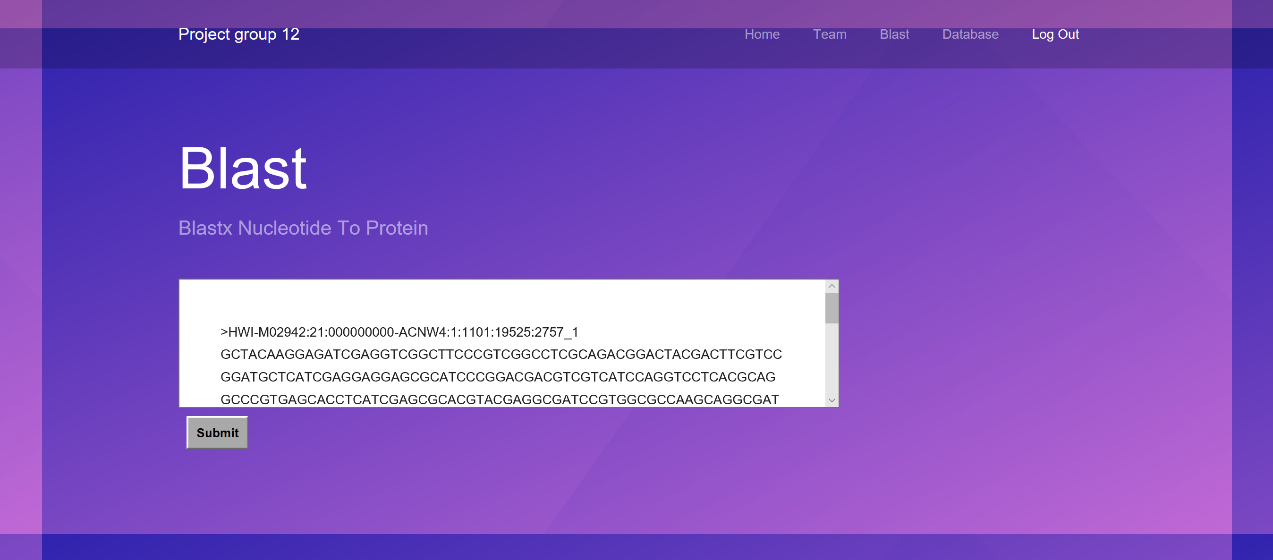
Figuur 4. 2-isopropylmalate synthase hydrofobic

Figuur 3. 2-isopropylmalate synthase

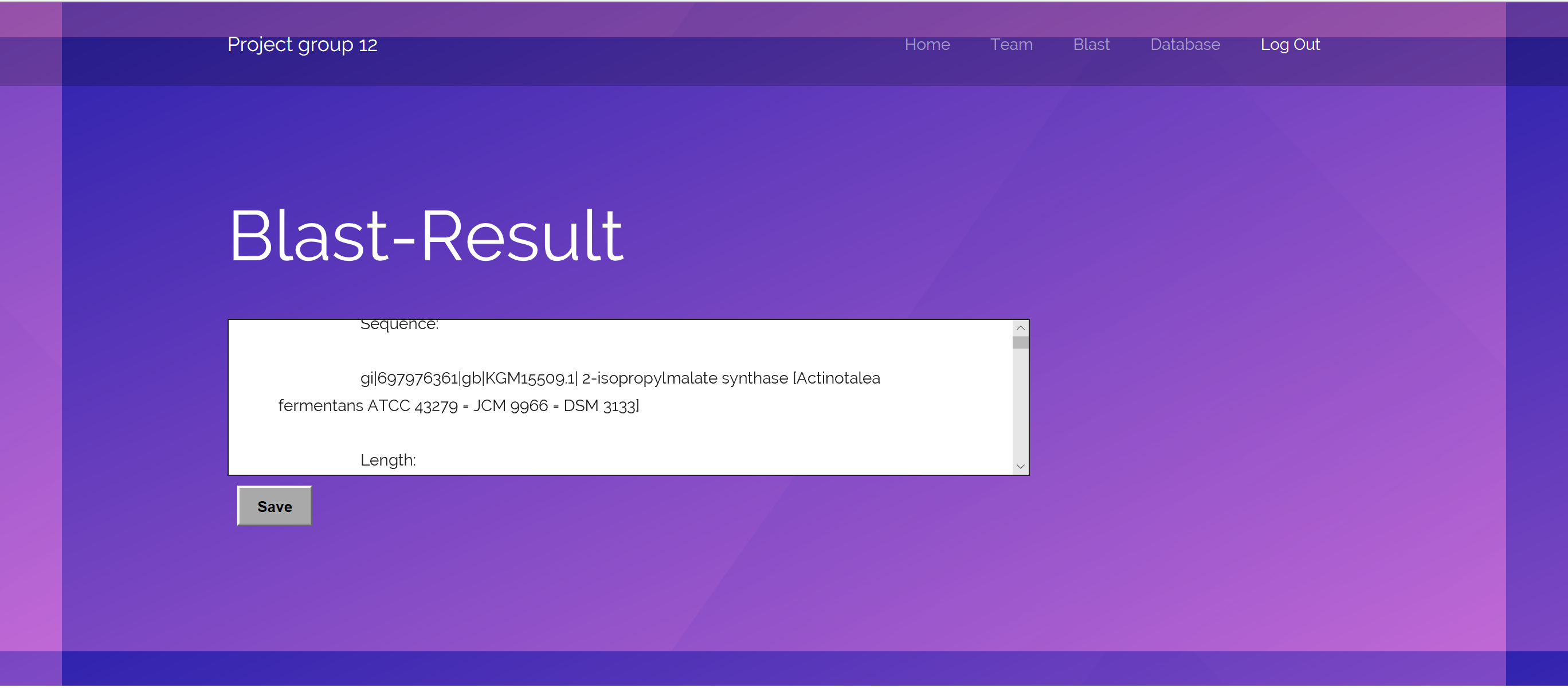
alloste ric (dimerisation) domain

Het python script is onderverdeeld in 6 functies. De eerste functie is zo geschreven dat het een txt file met ruwe data ophaalt en dit omzet naar een overzichtelijk fasta file. Dit fasta file wordt vervolgens weggeschreven en kan door de volgende functie weer aangeroepen worden. Parameters voor het blasten worden aan de module NCBIWWW.qblast meegeven en de resultaten van deze blast worden vervolgens opgeslagen in een variabele, die wordt weggeschreven als xml file. Wat door volgende functies makkelijk kan worden aangeroepen. De benodigde informatie (hit\_id, organisme, accessioncode, title, score, E-value, identities, positives, gaps, frame, match en subject) wordt uit het xml file opgevraagd en via een SQL query ingeladen naar de SQL database. Voor de taxonomie wordt per hit de accessiecode opgevraagd uit het eerder gegenereerde xml file. Door de accessiecode mee te geven aan de module Entrez kan de taxonomie van de hit opgehaald worden. De taxonomie wordt opgesplitst in de variabelen: Familie, Geslacht en Soort, deze worden dan ook via een SQL query ingeladen in de SQL database. Dit python script is geschreven om de 200 sequenties makkelijk van een annotatie te kunnen voorzien en ze gemakkelijk in de SQL database te laden. Een zelfde principe is gebruikt in de webapplicatie FLASK maar hier is het de bedoeling dat de gebruiker van de site een(of meerdere) sequenties blast en ze daarna eventueel update naar de database.

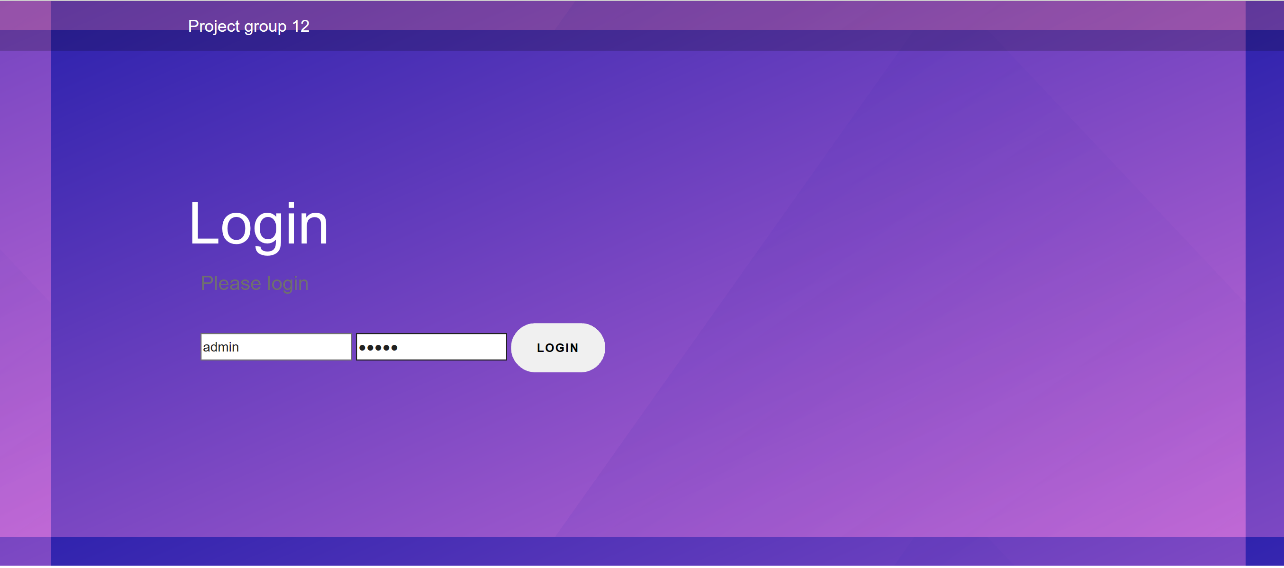
In de webapplicatie FLASK wordt er gebruik gemaakt van een zelfde soort script, alleen is er dan al een fasta-input aanwezig (figuur 5). De fasta-input moet voorzien zijn van een header en de sequentie moet een nucleotide sequentie zijn. Met vastgestelde parameters zal de input geblast worden tegen de NR-database. De output wordt op een nieuwe pagina overzichtelijk weergegeven (figuur 6). Nadat de Blastoutput is weergegeven kan de gebruiker er voor kiezen om de gegevens in de SQL database te laden(Bijlage ERD). Om er voor te zorgen dat hier geen misbruik van gemaakt wordt is er voor gekozen om een inlogscherm(Figuur 7) te gebruiken. De gebruiker zal de loginnaam en het wachtwoord eerst correct moeten invullen voordat er gebruik kan worden gemaakt van de website.



Figuur 5 Blast pagina, input is een nucleotide sequentie die met Blastx geblast wordt tegen de NR-database.



Figuur 6 Blast-Result, op deze pagina worden de blast resultaten overzichtelijk weergegeven. Mochten de resultaten interessant zijn voor het onderzoek, dan kunnen de resultaten worden ingeladen in de SQL database door op 'save' te klikken.



Figuur 7 Inlogscherm, wanneer iemand gebruik wil maken de site zal hij/zij eerst een correct wachtwoord moeten invullen. Dit om misbruik te voorkomen.

Bijlage 2

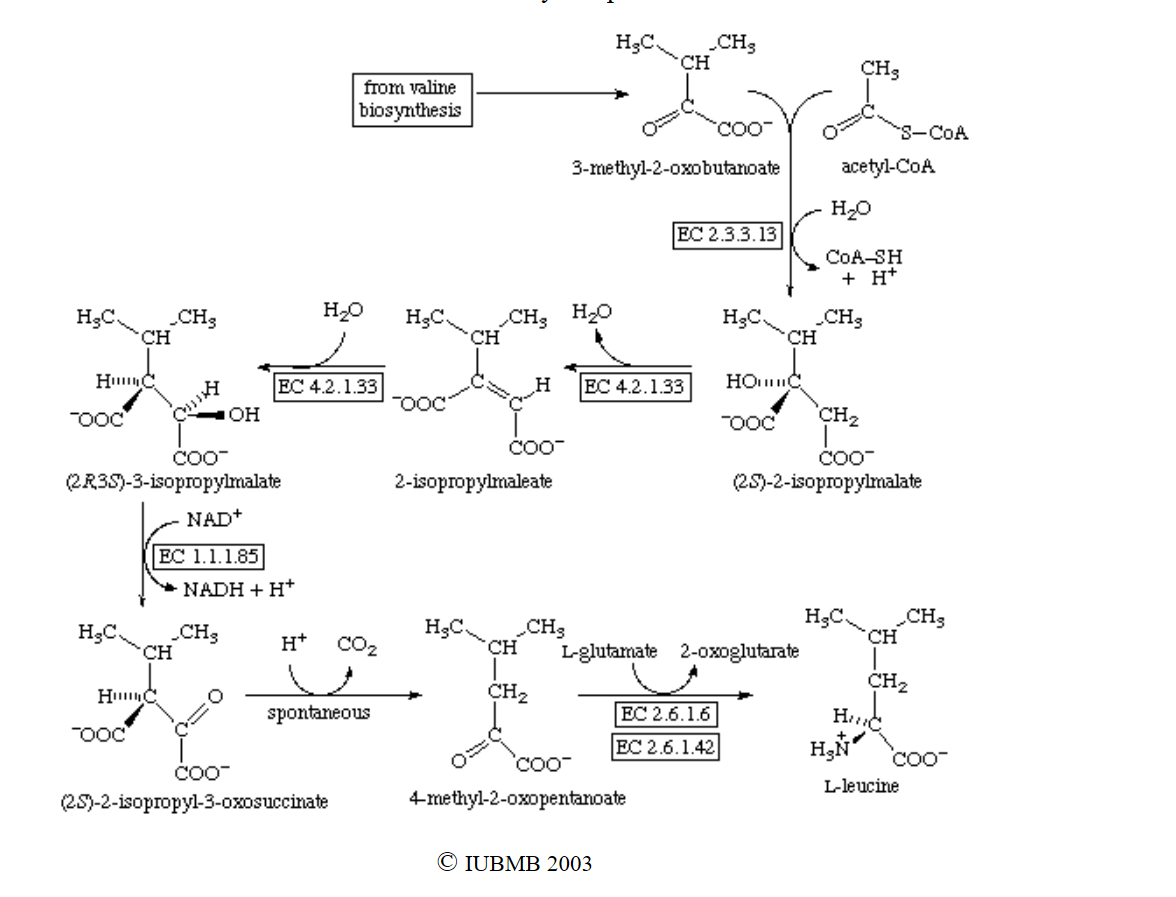


Figure De chemische reactie tussen 2-isopropylmalate synthase naar leucine. <http://www.sbcs.qmul.ac.uk/iubmb/enzyme/reaction/AminoAcid/Leu.html>

Bijlagen 3

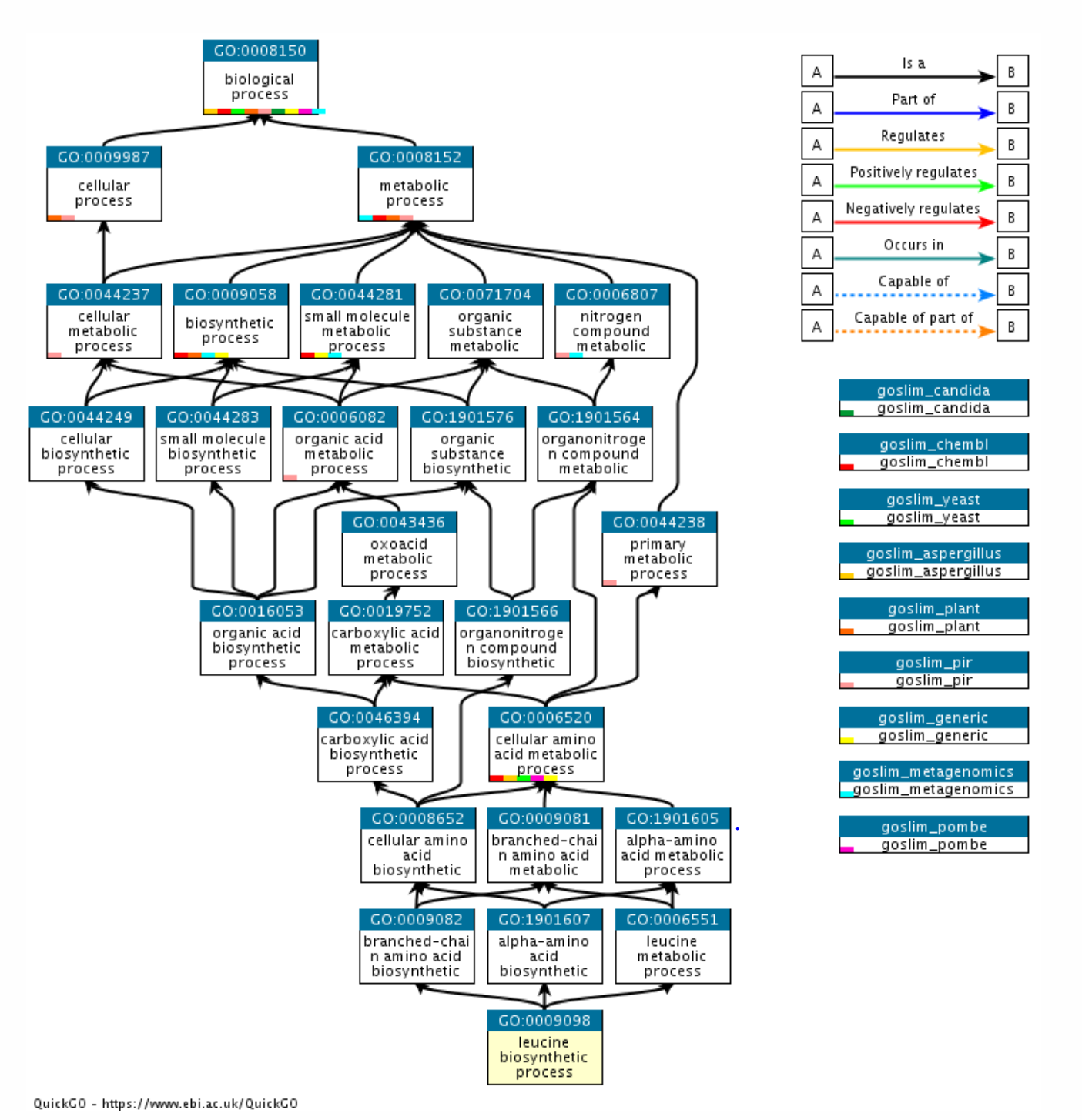


Figure GO termen bij de vorming van Leucine d.m.v. een chemische reactie met 2-isopropylmalate synthase. <https://www.ebi.ac.uk/QuickGO/term/GO:0009098>