

Det er ikke kun musik, film og bøger, der i dag bliver organiseret i store, online databaser. Også forskeres data bliver i stigende grad samlet på nettet, og mere og mere af det bliver efterhånden tilgængeligt for alle interesserede.

Internettet er formentlig en af de største samfundsmæssige revolutioner, vi har set de sidste mange år. Det har sat folk i kontakt med hinanden på kryds og tværs, og i dag er det ingen sag at snakke med en kineser, hvis man sidder i Danmark, eller udveksle viden med én fra Uganda, selvom man sidder bag en skærm i England.

Faktisk udviklede internettet sig oprindeligt blandt forskere, men world wide web, som vi kender det i dag, har også betydet en revolution for forskningsverdenen. Samarbejder på tværs af nationaliteter og fagligheder er nu mere normen end undtagelsen. Hver dag bliver der genereret store mængder videnskabelige data, når forskere over hele verden laver deres forsøg for at blive klogere på netop det, de forsker i. De voksende mængder af information har medført, at flere og flere internetbaserede databaser, der samler alle informationerne, skyder frem. Én af dem hedder GPCRDB (G-Protein Coupled Receptors DataBase), og den bliver styret af postdoc Vignir Isberg og lektor David Gloriam, der sidder på Institut for Lægemedeldesign og Farmakologi på Københavns Universitet.

”GPCRDB er en database, hvor vi samler

eksperimentelle data fra forsøg om en bestemt familie af receptorer, der hedder G-protein koblede receptorer (GPCR). Det er en stor, vigtig familie af proteiner til lægemiddelbrug (læs mere i faktaboksen Hvad er G-protein koblede receptorer?), fortæller Vignir Isberg.

Formålet er at gøre de mange data, der hele tiden bliver genereret verden over tilgængelige og brugbare for alle interesserede og at gøre det nemmere for folk at bruge hinandens data på tværs.

”Normalt ligger de data, den enkelte forsker genererer, bare som proteïnsekvenser et eller andet sted. Vi samler dem, så alle for det første får mulighed for at få del i de mange informationer, og for det andet får mulighed for at omdanne dataene til viden gennem at visualisere dem og analysere dem med nogle værktøjer, vi har udviklet og koblet op på databasen,” forklarer Vignir Isberg.

Arbejdet med GPCRDB og andre forskningsdatabaser kan sammenlignes med The Human Genome Project, et globalt samarbejde, hvor forskere verden over samarbejdede om at sekventere hele det menneskelige genom for allerførste gang.

Populær blandt forskere verden over

Databasen er en fusion af to separate databaser. Den ene blev udviklet af en forsker i Holland for 20 år siden, den anden har Vignir Isberg som ph.d.-projekt udviklet over de sidste tre år. I forbindelse med, at den hollandske forsker er gået på pension, har han overladt sin 20 år gamle database til lektor David Gloriam, og den blev fusioneret med den database, Vignir Isberg har opbygget. Resultatet er et GPCRDB, der både indeholder rå data og værktøjer til at analysere dem.

”Vi havde en intern database, der blandt andet indeholdt nogle proteïnsekvenser fra GPCR-receptorer, og vi fik brug for at kunne lave nogle bestemte analyser. Vi byggede nogle IT-baserede værktøjer for at gøre det nemmere for os selv, og fik så den idé, at vi lige så godt kunne gøre det tilgængeligt på internettet. Efterfølgende fik vi overdraget den hollandske database med en masse data, og så blev det slået sammen,” siger Vignir Isberg.

Der findes mange databaser inden for det farmakologiske område, men GPCRDB adskiller sig, fordi den fokuserer specifikt på familien af G-protein koblede receptorer. Inden for den ene familie >

OM FORSKEREN:



Vignir Isberg er netop blevet færdig som ph.d. ved Institut for Lægemedeldesign og Farmakologi, hvor han nu laver en post.doc. Han blev oprindeligt student i Island fra Islands Handelsgymnasium, og den gang troede han, at han skulle være en slags IT-iværksætter. Men kemi og biologi var alligevel så spændende, at han valgte at læse farmaci, og har opbygget

sin viden om IT og programmering ved siden af. Han planlagde dog at kombinere farmaceutuddannelsen med IT, og det må man sige, han er lykkedes med. For Vignir Isberg er det først og fremmest en stor fascination af, hvad man kan opnå ved at kombinere naturvidenskab og IT, der driver værket.

samler Vignir Isberg og David Gloriam så alt, hvad de kan grave frem af information. Databasen indeholder blandt andet:

- proteinsekvenser
- mutationer på receptorerne
- receptorernes krystalstrukturer
- ligand-bindinger

På nuværende tidspunkt har databasen omkring 1.000 besøgende om måneden, og tilbagemeldingerne er positive – forskerne er glade for de muligheder, de får med GPCRDB, og bruger i vidt omfang databasen i deres videnskabelige arbejde.

Når en forsker går ind og søger efter en bestemt G-protein koblet receptor, får de først et hurtigt overblik. Her kan de se alt, hvad databasen indeholder af informationer. På den måde kan de for eksempel se, om nogen allerede har skabt den mutation, de overvejede at lave. De kan også bruge databasens værktøjer til at analysere nogle af de data, de får præsenteret, og det tilbud benytter mange sig af.

Bygger 3D-strukturer og diagrammer

Et særligt populært værktøj er databasens 'Template selection'. Dette værktøj hjælper forskerne med at skabe en 3D-model for den receptor, de ønsker at arbejde med. I dag ved forskerne nemlig, at den fysiske (3D) struktur af en receptor er ekstremt afgørende, bl.a. for udviklingen af lægemidler. Det kan sammenlignes med en nøgle i en dør – et aktivt lægemiddelstof skal fysisk passe til receptoren, før 'døren' ind til kroppen kan låses op.

Desværre er der meget svært at finde ud af, hvordan en receptor egentlig ser ud i 3D. Det skyldes, at der ikke er nogle metoder, der lige kan tage et billede af en receptor. Kun ved at få proteinet til at krystallisere, kan man bruge røntgenstråler til at skabe en 3D-model af det – metoden hedder røntgenkrystallografi. Det er indtil videre kun lykkedes at kortlægge 3D-strukturen af 25 forskellige GPCR receptorer – og der findes altså 3500 receptorer i GPCR-familien.

"Her kan 'Template selection' hjælpe forskerne. Værktøjet scanner databasen for at finde ud af, hvilken af de 25 kendte 3D-strukturer, forskerens receptor ligner mest. Det foregår ved, at værktøjet sæt-

ter de to proteinsekvenser ved siden af hinanden, så matchende aminosyrer sidder ud for hinanden. Efterfølgende kan programmet bygge en 3D-model, der også tager højde for forskellene. Det er et vigtigt værktøj, for 3D-strukturen for en receptor er afgørende for lægemiddeludvikling," forklarer Vignir Isberg.

Databasens værktøjer kan også generere diagrammer, som mange forskere bruger til at illustrere deres arbejde i videnskabelige artikler.

Et arbejde, der aldrig slutter

Databasen indeholder på nuværende tidspunkt omkring 80.000 proteinsekvenser fra GPCR-familien, både fra mennesker og dyr. Dertil kommer mange tusinde mutationer, ligand-bindinger og meget mere. Og der dukker hele tiden nye data op, der skal inkluderes. Nogle data kommer fra store, velkendte databaser, mens andre gemmer sig hos de enkelte forskere eller i mindre databaser.

Når Vignir Isberg har gravet de relevante informationer frem hos forskerkolleger eller i databaser, skal han formatere dem, så de kan passe ind i databasen. De kan komme i et format, databasen ikke kan arbejde med, og så kan han enten køre informationerne igennem et konverteringsprogram, der omformaterer dem, eller også må han selv designe et lille program til opgaven.

"Der er tre primære arbejdsopgaver forbundet med at drive en database som GPCRDB. Der er programmeringen, der sørger for at databasen fungerer og er brugervenlig. Så er der arbejdet med at finde data, og at 'omformulere det' til et format, vores database kan læse, så de nye informationer kan blive en del af den. Og så er der løbende planlægning og udvikling, hvor vi ser på, hvordan vi kan forbedre databasen og skabe nye analyseværktøjer. Her har vi også otte internationale samarbejdspartnere, der byder ind. Dem mødes vi med to gange årligt, hvor vi viser demoer og diskuterer fremtidige muligheder," fortæller Vignir Isberg.

Fremover vil en stor del af arbejdet med databasen bestå i at skrive programmer, der kan automatisere noget af det arbejde, Vignir Isberg laver med at trække data ind og formatere det. Det er ikke som sådan svært programmeringsar-

bejde, siger Vignir Isberg, men at arbejde med databaser kræver en klart formuleret idé og en sans for struktur.

"En database er en organiseret struktur for data. Så for at lave en database skal man kunne designe en datastruktur: Hvad skal ind og hvordan skal det organiseres? Der er også en smule matematik i nogle af de værktøjer, vi bygger, så alt i alt er arbejdet med en farmakologisk database et godt miks af naturvidenskab og IT," forklarer Vignir Isberg.

Han arbejder som minimum med databasen de næste to år, mens han laver sin postdoc. Bagefter ved han ikke, hvad han selv skal, men arbejdet med en database slutter aldrig.

"Med de voksende mængder af videnskabelig data, der hele tiden bliver skabt over hele verden, bliver det kun vigtigere at have de rigtige værktøjer og systemer til at opbevare dataene og sørge for, at de er tilgængelige for alle interesserede. Der vil også være løbende behov for nye metoder til at analysere, for det er ikke nok at have dataene liggende; de skal selvfølgelig også kunne omdannes til brugbar viden, der f.eks. kan bruges til at lave nye lægemidler," slutter Vignir Isberg. ■

Hvad er "G-protein koblede receptorer"?

GPCR er den største og mest forskelligartede familie af membranreceptorer, der findes. Receptorerne sidder på cellers overflade og agerer som postkasse for beskeder i form af lysenergi, peptider, lipider, sukkerkæder og proteiner. Beskederne informerer cellerne om tilstedeværelse eller fravær af lys og næringsstoffer i miljøet, eller videreformidler information fra andre celler. GPCR-receptorerne spiller en rolle i en lang række funktioner i kroppen, og forskere mener, at mellem en tredjedel og halvdelen af lægemidler på markedet virker ved at binde til GPCR-receptorer.

Kilde: <http://www.nature.com/scitable/topicpage/gpcr-14047471>

Programmeringssprog til databaser

Når Vignir Isberg arbejder med GPCRDB, bruger han mange forskellige programmeringssprog:

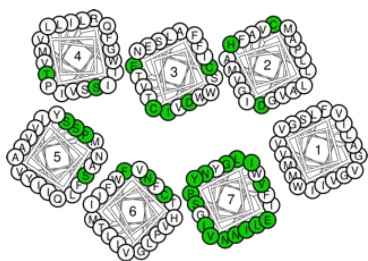
PHP: Bruges til at opbygge interfacen, altså det, man kan se på hjemmesiden.

Python: Bruges bl.a. til at bygge værktøjer til at analysere de data, der kommer ind.

MySQL: Bruges som en slags grundstruktur for at lave en database. Ovenpå det er der bygget et sprog, som man bruger til at hente data ind i databasen.

Java: Bruges bl.a. til at hente og formatere de data, der bliver trukket ind i databasen.

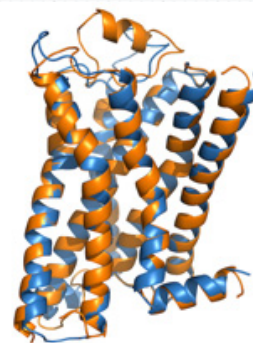
ET SMUGKIG IND I DATABASEN



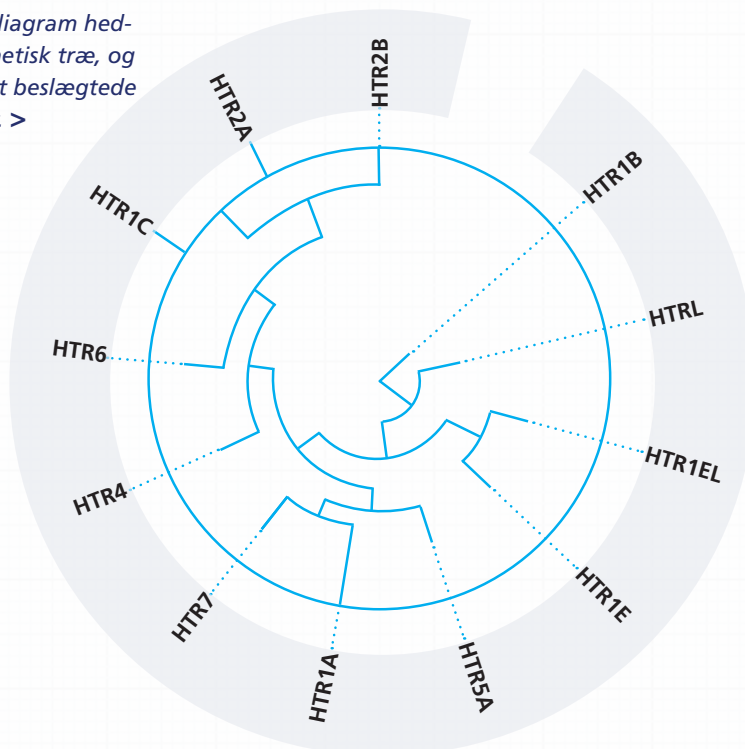
Billedet viser funktionen 'Template selection', hvor en receptor model (blå) er lagt oveni en "skabelon" krystalstruktur (orange). >

Et af de diagrammer, man kan trække ud ad databasen. Denne type kaldes en Helix box og viser receptoren som set igennem celledmembranen i 2D. Cirklene repræsenterer aminosyrer. Der findes mutationsdata for de farvede aminosyrer.

<



Denne type diagram hedder phylogenetisk træ, og viser hvor tæt beslægtede receptorer er. >



SRP-inspiration

- Har man mod på programmering, kan man lave en egen mini-database ved hjælp af de forskellige programmeringssprog og trække informationer fra GPCRDB ind i den.
- Eller brug databasens funktion 'web services' på <http://www.gpcr.org/7tml> til at skrive et lille program, der kan interagere direkte med databasen.
- Relevante fag kunne være biologi, bioteknologi eller samfundsfag, hvis man ser på den samfundsmæssige betydning af, at store mængder data bliver tilgængelige for alle.
- Hvis du vil lave SRP-øvelser hos os, så kan du læse om mulighederne for det på pharmaschool.ku.dk/SRP.

RELATERET LÆSNING

<http://www.farma.ku.dk/index.php/Naar-filmens-fiktion-bliver-vi/1521/0/>
<http://www.farma.ku.dk/index.php/Krystallografi-i-farmaceutisk/1499/0/>