

Elev vejledning:

Byg Din Egen Slangemodgift med Kunstig intelligens

Introduktion

I forløbet Gift og Modgift har du lært om, hvad gift er, og måske har du også læst om, hvordan man bruger laborieteknikker som fagdisplay til at lave modgift. Med fagdisplay kigger man i store antistofbiblioteker efter antistoffer, der kan binde og neutralisere slangegift. Det er en god metode, men den er tidskrævende i laboriet. For nylig er forskere derfor begyndt at udforske, om denne laborietid kan reduceres ved hjælp af computerteknologi.

Det er faktisk lykkedes forskere at udvikle et program ved navn RFdiffusion, der benytter kunstig intelligens til at simulere proteiner. At simulere proteiner betyder i dette tilfælde at opfinde helt nye proteiner helt fra bunden!

Selvom du ikke nødvendigvis behøver at forstå, hvordan programmet fungerer, er det godt at have en generel idé om konceptet for hvordan det er muligt at skabe proteinstrukturer fra bunden. Forestil dig et billede af en kat. Billedet består af tusindvis af pixels, som tilsammen danner et firkantet billede. Hver pixel har et koordinat. Hver pixel har 3 værdier der svarer til rød, blå og grøn farve og de tre tal bestemmer "blandingen" og dermed pixelens farve. Se hvordan nedenstående farve blandes med rød, grøn og blå:

Red: 153

Green: 51

Blue: 255

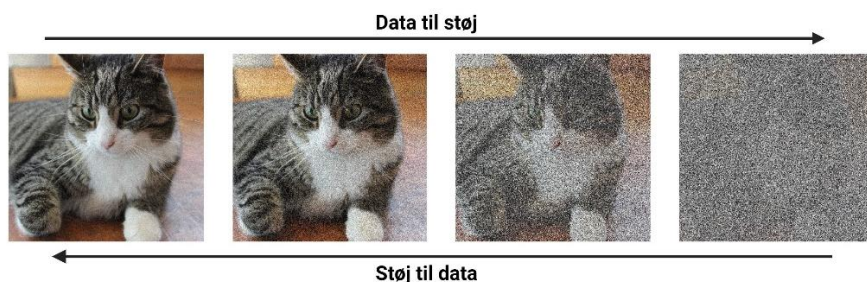


Hvis alle disse pixels farvewærdier ganges med en matematisk faktor, vil de forskyde sig, hvilket gør billedet mere rodet. Hvis du gentager dette mange gange, vil du til sidst få et billede, der er så rodet, at du ikke kan se, hvad det er – denne proces kaldes noising. Et billede der er så rodet at man ikke kan se hvad det forestiller kaldes "ren støj."

Forestil dig igen kattebilledet: Først lader man sin diffusions model lave billedet om til ren støj. Derefter træner man maskinlærings modellen (AI model) til at gå fra det "rene støj" billede og tilbage til det

originale billede af katten. Hvis man giver en AI modellen nok opgaver af denne art vil den kunne lære hvordan man går fra et "ren støj billede" til et rigtigt billede. Dette kaldes en diffusionsmodel og den kan lave rigtige billeder "ud fra tilfældig støj" som er nemt at generere. Man kan forestille sig at modellen har "lært noget om hvad der karakteriserer et rigtigt billede som ikke er støj." Nedenfor kan du se hvordan støj bliver til blomster og hvordan et kattebillede bliver noised og denoised. Du kan se en demonstration af denne type model på Bings Image generator [her](#). Prøv at skrive en sætning som fx "A class room filled with students who are working on their computer, studying AI generation of protein structures" og se, hvad den finder på!

Fra støj til blomster: Hvis du [klikker her](#), kan du se en gif, der viser hvordan en diffusionmodel, der er trænet på billeder af blomster omdanner støj til kunstige billeder af blomster



RFdiffusion opererer ud fra det samme grundprincip, men er specifikt trænet på proteiner. Modellen er dog en smule mere kompleks, da proteiner er tredimensionelle og dermed ikke todimensionelle som billeder. Men på samme måde som med billeder, har modellen lært noget om hvordan et protein ser ud. Den har både evnen til at forestille sig tilfældige proteiner, men den kan også skabe forestillinger om proteiner, der har evnen til at binde sig til andre proteiner, hvilket gør den særligt brugbar i slangemodgift udviklingen.

I denne øvelse vil du via din egen computer bruge RFdiffusion til at udvikle proteiner de kan binde slangegift.

Du skal bruge

- En computer
- Internetforbindelse
- Google konto
- Ikke obligatorisk: Programmet Pymol som kan downloades [her](#). Du skal bruge dette til at visualisere proteiner.

Sådan gør du

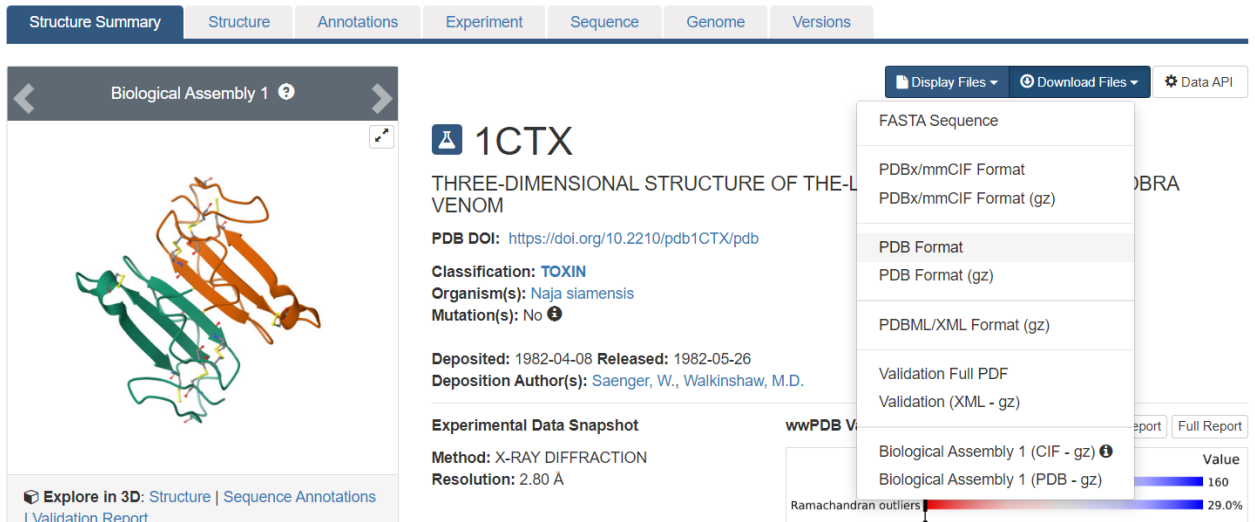
Del 1: Lær proteinstrukturen at kende

For at kunne lave en modgift til en slangegift er det vigtigt at gøre sig bekendt med slangegiftens struktur. I denne øvelse vil I arbejde med toksinet α -cobratoksin, som er en del af Kobra slangegift. Dette gift-protein

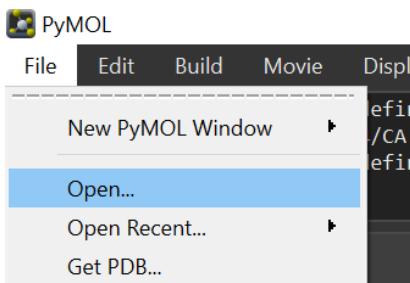
kender du måske, hvis du har om "nervegift" i forløbet. Dette toksin er faktisk ret svært at lave gode bindere til, men vi bruger det her, fordi vi allerede har arbejdet med det i forløbet.

Mange proteiner kan findes online på proteindatabanken kaldet RCSB. Hvert protein i databanken har en unik kode som man kan bruge til at finde strukturer. I dette tilfælde skal vi arbejde med α -cobratoksin som har koden 1CTX.

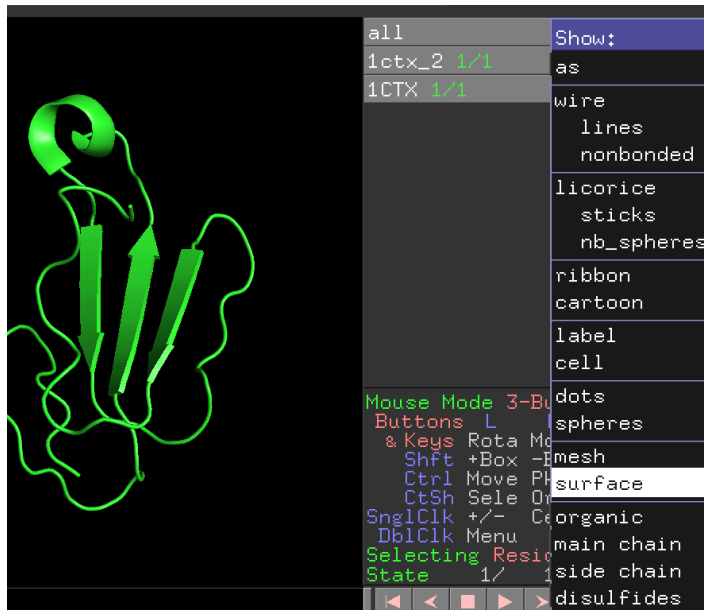
1. Find 3D strukturen for α -cobratoksin ved at bruge proteindatabanken <https://www.rcsb.org/>. Brug søgefeltet og søg efter 1CTX.
2. Download proteinstrukturen. Tryk "download files" → PDB format



3. Åben strukturen i Pymol. Dette gøres ved at åbne Pymol → File → Open → find din fil og åben den. Alternativt kan man i prompt feltet skrive fetch 1CTX



4. Brug musen til at dreje proteinet.
5. Du kan både se proteinets sekundære struktur eller overfladen af proteinet. Dette kan du gøre ved at klikke S → Surface.
6. Du kan igen se den sekundære struktur ved at klikke H → Surface



7. Brug nu informationen fra proteindatabanken til at besvare følgende spørgsmål.
 - Find et billede af den slange som giften stammer fra og indsæt her:
 - Er giften en nervegift eller en hæmotoksisk gift?:
 - Hvor mange aminosyrer består proteinet af?:
 - Hvilke sekundære strukturer genkender du?
 - Indsæt et screenshot at α -cobratoksin med og uden "surface" (se skridt 5 & 6):
 - Ekstra spørgsmål: Hvordan virker giften? Hvad er "mekanismen" der gør at giften udøver effekten?

Del 2: Design dit eget protein med RFdiffusion

I dette afsnit skal du bruge RFdiffusion til at designe et tilfældigt protein. På den måde lærer du programmet at kende. Programmet er skrevet i Python kode og opsat i en "Notebook." Det gode er, at du ikke skal kunne skrive eller forstå pythonkode for at bruge programmet!

1. Gå til [LINK](#).
2. Log ind med en Google konto
3. Læs intro til hvordan man bruger Notebooken og Google Colab
4. Klik nu "play-ikonet" knappen ved "Start RFdiffusion". Dette starter programmet hvilket kan tage lidt tid (maks 10 min). Når play knappen bliver grøn er programmet klar. Hvis du er mere interesseret, kan du klikke "show code" eller "vis kode" – men koden er ret kompleks så ikke brug

tid på at forstå den. Vent herefter nogen sekunder før cellen nedenunder "Definer RFdiffusion" køres. En nærmere beskrivelse er givet i Notebooken.



Play ikonet ser sådan ud!

> Start RFdiffusion (Tidsforbrug: ~ 2 - 10 minutter)



[Show code](#)

5. I den næste sektion "Brug RFDiffusion til at lave et protein" skal vi nu udfylde feltet således:

> Brug RFdiffusion til at lave et protein



name:

contigs:

pdb:

hotspot:

num_designs:

[Show code](#)

Indstillinger

Name: Ved indstillingen "**name**" kan du indsætte et hvilket som helt navn. Fx "**test**" som vist her.

Contigs: Indstillingen "**contigs**" fortæller programmet at vores protein skal være **100 aminosyrer** langt.

Num_designs: Indstillingen "num_designs" fortæller hvor mange proteiner vi vil designe. I dette tilfælde skal vi blot designe et protein.

De andre indstillinger beholder vi på standardindstillingerne indtil videre ;-)
Når du er klar, kan du trykke play. Du vil nu kunne se proteinet blive designet:



6. Når play knappen bliver grøn er proteinet færdig. Tillykke: Du har nu designet dit første protein!

Indsæt et screen shot af dit protein her:

7. Vi har nu genereret en proteinstruktur, men hvad er rent faktisk aminosyresekvensen til dette helt nye protein. Find ud af dette ved at køre den næste del af koden der bruger de to programmer kaldet Protein mpnn og Alpha Fold til at finde ud af netop dette.
Brug koden til at generer én sekvens til din protein struktur ved at sætte num_seqs=1

> Brug **ProteinMPNN** og **AlphaFold** til at genere en aminosyre sekvens



num_seqs: 1

[Show code](#)

8. Du har nu aminosyresekvensen for dit nye protein!

Indsæt sekvensen her:

Spørgsmål: Med hvilken "kode" står aminosyrerne? (hint: hvor mange bokstaver bruges der til at vise en aminosyre)

Find enbokstavskoden for disse aminosyrer

Arginin:

Asparaginsyre:

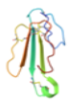
Tyrosin:

Serin:

Del 3: Design Slangemodgift

Nu har vi lært hvordan man generer proteiner med computeren. Nu skal vi prøve at lave et protein der potentielt kan binde og neutralisere slangen toksinet α -cobratoksin.

- Gå tilbage til afsnittet "Brug RFdiffusion til at lave et protein". Her skal vi ændre lidt i indstillingerne.
 - Du kan give din opgave et navn under "name". Brug evt: " α -cobratoksin modgift"
 - Under sektionen "contigs" i programmet skal vi specificere størrelsen på det protein, vi ønsker at generere, samt hvad dette protein skal binde sig til.
 - I vores specifikke tilfælde ønsker vi at konstruere en aminosyrekæde med en længde på 100-150 aminosyrer, og vi vil have, at dette protein skal kunne binde sig til vores α -cobratoksin-kæde. På PDB-hjemmesiden kan vi se, at denne kæde er navngivet "A." Derfor skriver vi "A:100-150" i feltet.

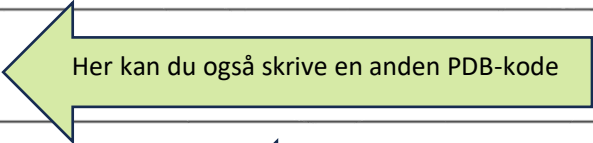
Entity ID: 1					
Molecule	Chains ⓘ	Sequence Length	Organism	Details	Image
ALPHA-COBRATOXIN	A 	71	Naja siamensis	Mutation(s): 0 ⓘ	

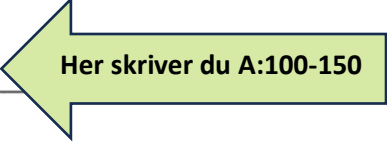
- Under "pdb" punktet indtaster vi pdb-indikatoren, som i vores eksempel er "1CTX."
 - OBS:** Hvis du vil lave en modgift til et andet protein fra PDB, så skriver du PDB-koden for det pågældende protein her (eks. 4LFT er et toxin fra *Black Mamba*)
- Til sidst angiver vi num_designs som 1, da vi kun ønsker at designe ét enkelt protein.
- Når du er klar, skal du trykke på afspil-knappen

> Brug RFdiffusion til at lave et protein

 **name:**

contigs:

pdb:  Her kan du også skrive en anden PDB-kode

hotspot:  Her skriver du A:100-150

num_designs:

[Show code](#)

2. Du har nu designet et slange modgift protein!

Indsæt et billede af din modgift herunder. Hvis på billedet hvilken del der er din modgift og hvilken del der er a-cobratoxin:

3. Nu har du dannet dit helt eget protein der potentielt kan binde til slangetoksinet a-cobratoksin. Men hvad er sekvensen? Tryk nu igen play næste kode for at finde sekvensen.

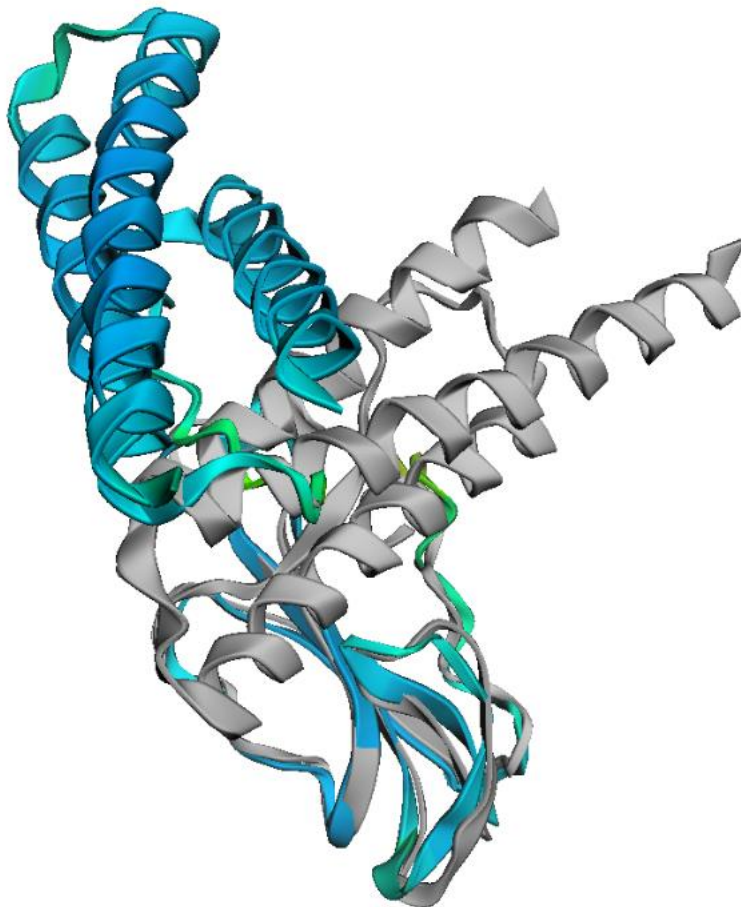
Indsæt din sekvens her:

4. I skal nu i klassen undersøge hvem der har designet det bedste protein (hvilket jo i høj grad afhænger af hvor gode i har været, lidt ligesom en konkurrence om at slå det højeste tal...)! Det kan i gøre ved at kigge på den værdi der hedder `i_pae`. Dette er en faktor der beskriver hvor god interaktionen er mellem dit protein og din slangegift. Jo lavere et tal jo bedre en interaktion. Husk at dette er en simuleret interaktion, og det er vanskeligt at forudsige protein-protein interaktioner, så det skal verificeres i et laboratorie! (Hvis tallet er under 5 er det højst sandsynligt at dit protein i den virkelige verden vil kunne binde slange toksinet. Dette er dog sjældent så hver ikke overrasket over at kun få eller ingen har en `i_pae` under 5)


```
{'pdb':'outputs/test_1h5s0_0.pdb','loc':'outputs/test_1h5s0','contigs':'A1-71:127-
protocol=binder
running proteinMPNN...
running AlphaFold...
design:0 n:0 mpnn:1.122 plddt:0.883 i_ptm:0.103 i_pae:26.283 rmsd:35.548 IRCFITPD
CPU times: user 282 ms, sys: 29.2 ms, total: 311 ms
Wall time: 47.9 s
```

Skriv din i_pae score her og evt. også på tavlen så klassen kan sammenligne:

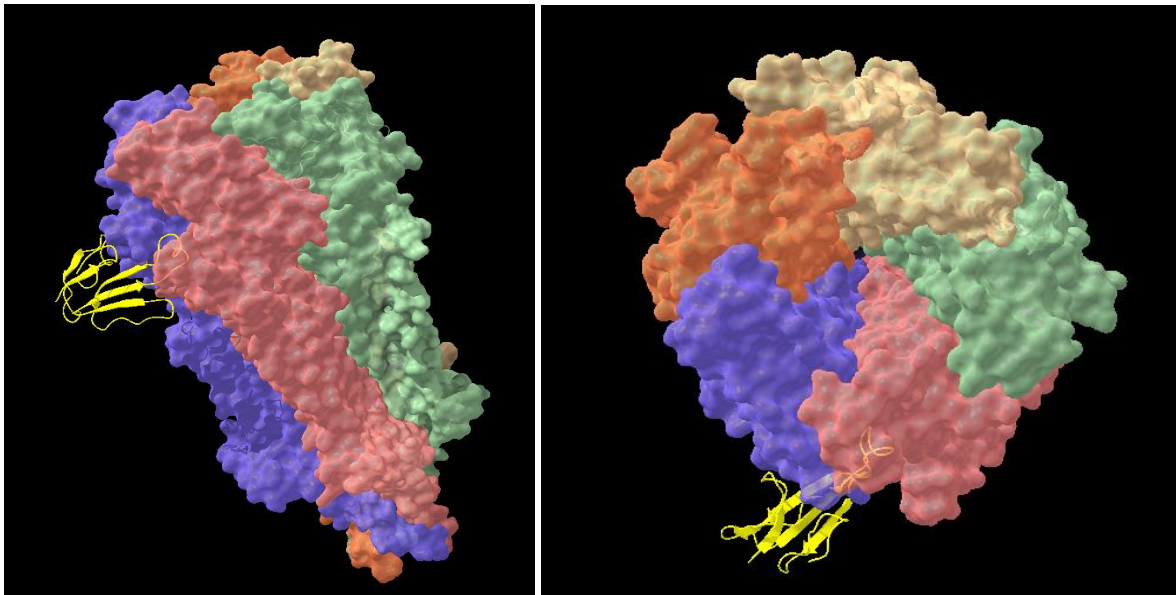
På figuren ses to forskellige strukturer. Den grå struktur er den, som RFdiffusion har designet, mens den farvede struktur er den som AlphaFold og proteinMPNN har forudset efter at have lavet en aminosyresekvens. Disse to strukturer er normalt meget ens, men i dette eksempel ligner det, at den farvede struktur er roteret i forhold til den grå.



Del 4: Undersøg om din modgift er neutraliserende:

Nu skal du undersøge hvordan din modgift binder sig til slangegiften, hvilket man kalder orienteringen. Denne information giver dig mulighed for at vurdere, om din slangemodgift vil have neutraliserende egenskaber eller ej. Et protein kan godt binde til slangegiften uden at neutralisere den! For at genopfriske, hvad en neutraliserende binding indebærer, kan du tage et nærmere kig på figur 10 under afsnittet om [Antistoffer og passiv immunisering](#).

1. Når man bliver forgiftet med α -cobratoksin vil det binde sig til den nikotinerge acetylcholin receptor (nAChR) hvilket fører til lammelse i kroppen. På nedenstående figur kan du i gul α -cobratoksin (i gult) bundet til den store receptor nAChR (det store med mange farver). Begge billede viser den samme binding men fra to vinkler.



2. Ud fra din viden om måden hvorpå α -cobratoksin normal binder nAChR vurder da om dit designet protein vil være neutraliserende eller ej.

Extra opgave

- Vi har endnu ikke brugt funktionen hotspot endnu. Denne funktion fortæller programmet hvor præcist at den skal binde til proteinet. Prøv at køre en design runde hvor du vælger hotspot A33 (Dette gør du ved at skrive A33 i feltet "Hotspot" inden du genererer din modgift.)
 - Hvad sker der?
 - Hvorfor er dette smart?

Refleksion

- Forstil dig hvad man eller kunne bruge RFdiffusion til. Kom med tre forslag.
- Hvorfor kan et protein binde til slangemodgiften uden at neutralisere den?
- Hvilke ting afgør om to proteiner kan binde til hinanden? Hvilke kræfter er på spil?



- Kender du nogle eksempler på proteiner der binder til andre proteiner (hint: peptidhormoner)
- Hvordan ville man kunne verificere i et laboratorie om ens proteiner faktisk binder til slangegiften?