

Student Guide:

Build your own Snake Antivenom using Artificial Intelligence

Introduction

In the course Venom and Antivenom, you have learned about what venom is, and perhaps you have also read about how to use laboratory techniques such as a phage display to make antivenom. With phage displays, large antibody libraries look for antibodies that can bind and neutralize snake venom. It is a good method, but it is time-consuming in the laboratory. Therefore, researchers have recently begun to explore whether this laboratory time can be reduced using computer technology.

In fact, researchers have succeeded in developing a program called RFdiffusion, which uses artificial intelligence to simulate proteins. Simulating proteins in this case means inventing completely new proteins from scratch!

Although you don't necessarily need to understand how the program works, it's good to have a general idea of the concept of how it's possible to create protein structures from scratch. Imagine a picture of a cat. The image consists of thousands of pixels, which together form a square image. Each pixel has a coordinate. Each pixel has 3 values corresponding to red, blue and green color and the three numbers determine the "blend" and thus the color of the pixel. See how the color below mixes with red, green and blue:

Red:

153

Green:

51

Blue:

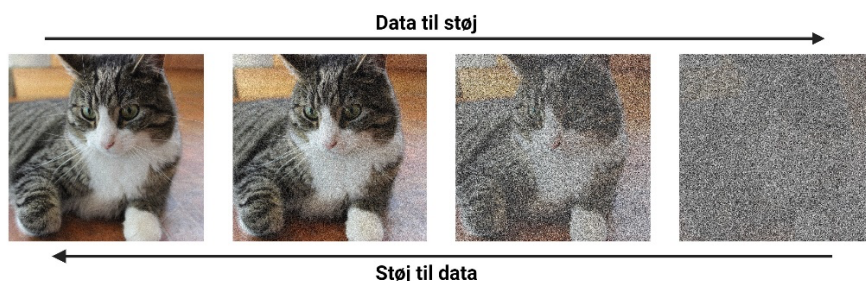
255



If all these pixels are multiplied by a mathematical factor, they will shift, making the image messier. If you repeat this many times, you'll eventually get an image that's so messy you can't see what it is – this process is called noising. An image so messy that you can't see what it depicts is called "pure noise."

Imagine the cat image again: First, you let your diffusion model turn the image into pure noise. Then you train the machine learning model (AI model) to go from the "pure noise" image and back to the original image of the cat. If you give an AI model enough tasks of this nature, it will be able to learn how to go from a "pure noise image" to a real image. This is called a diffusion model, and it can create real images "from random noise" that are easy to generate. One can imagine that the model has "learned something about what characterizes a real image, which is not noise." Below you can see how noise turns into flowers and how a cat image becomes noised and denoised. You can see a demonstration of this type of model on Bing's Image generator [here](#). Try writing a sentence like "A class room filled with students who are working on their computer, studying AI generation of protein structures" and see what it comes up with!

From noise to flowers: If you [click here](#), you'll see a gif showing how a diffusion model trained on images of flowers transforms the noise into artificial images of flowers



RFdiffusion operates on the same basic principle, but is specifically trained on proteins. However, the model is a bit more complex, as proteins are three-dimensional and thus not two-dimensional like images. But just like with pictures, the model has learned something about what a protein looks like. It has the ability to imagine random proteins, but it can also create images of proteins that bind to other proteins, which makes it particularly useful in snake antivenom development.

In this exercise, you will use RFdiffusion via your own computer to develop proteins that can bind snake venom.

You will need

- A computer
- Internet connection
- Google Account
- Not mandatory: The program Pymol which can be downloaded [here](#). You will need this to visualize proteins.

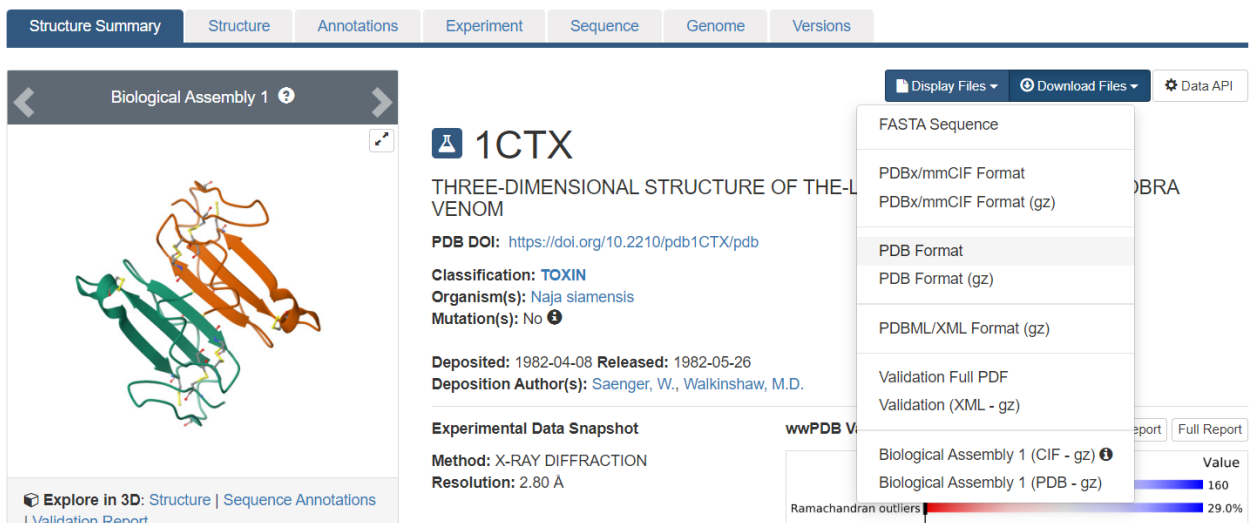
Here's how to do it

Part 1: Get to know the protein structure

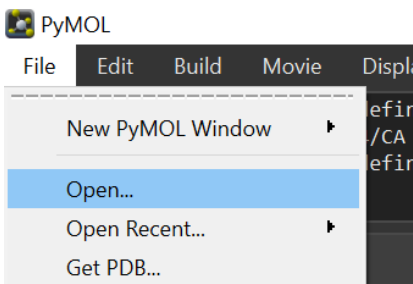
In order to make an antivenom for a snake venom, it is important to familiarize yourself with the structure of snake venom. In this exercise, you will work with the toxin α -cobratoxin, which is part of Cobra snake venom. You may know this venom protein if you have read about "neurotoxins". This toxin is actually quite difficult to make good binders for, but we use it here because we have already worked with it in the process.

Many proteins can be found online at the protein data bank called RCSB. Each protein in the databank has a unique code that can be used to find structures. In this case, we are going to work with α -cobratoxin, which has the code 1CTX.

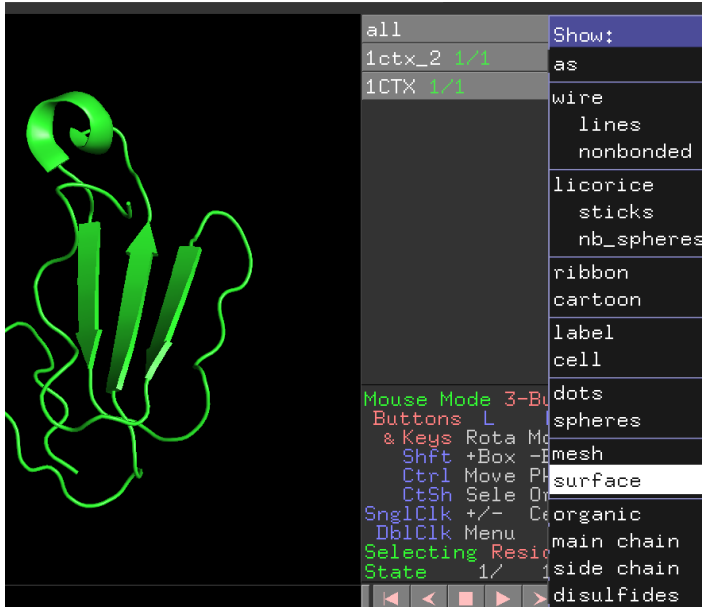
1. Find the 3D structure for α -cobra toxin using the protein data bank <https://www.rcsb.org/>. Use the search bar and search for 1CTX.
2. Download the protein structure. Press "download files" → PDB format



3. Open the structure of Pymol. This is done by opening Pymol → File → Open → find your file and open it. Alternatively, in the prompt field you can type `fetch 1CTX`



4. Use the mouse to rotate the protein.
5. You can see both the secondary structure of the protein or the surface of the protein. To do this, click `S` → Surface.
6. You can again see the secondary structure by clicking `H` → Surface



7. Now use the information from the protein data bank to answer the following questions.
- Find a picture of the snake from which the venom originates and insert here:
 - Is the venom a neurotoxin or a hemotoxic venom?
 - How many amino acids does the protein consist of?
 - What secondary structures do you recognize?
 - Insert a screenshot of α -cobratoxin with and without "surface" (see steps 5 &6):
 - Extra question: How does the venom work? What is the "mechanism" that makes the venom exert the effect?

Part 2: Design your own protein with RFdiffusion

In this section, use RFdiffusion to design a random protein. That way you get to know the program. The program is written in Python code and set up in a "Notebook." The good thing is that you don't have to be able to write or understand python code to use the program!

1. Go to [LINK](#).
2. Sign in with a Google Account
3. Read the intro to how to use the Notebook
4. Now click the "play icon" button at "Start RFdiffusion". This will start the program, which may take some time (max 10 min). When the play button turns green, the program is ready. If you're more interested, you can click "show code" or "show code" – but the code is quite complex so don't spend time understanding it.



The play icon looks like this!

> **Start RFDiffusion** (Tidsforbrug: ~ 2 - 10 minutter)



[Show code](#)

5. In the next section "Use RFDiffusion to make a protein" we must now fill in the field like this:

> **Brug RFDiffusion** til at lave et protein



name: " Insert text here _____

contigs: " 100 _____

pdb: " Insert text here _____

hotspot: " Insert text here _____

num_designs: 1 _____

[Show code](#)

Settings

Name: At the "name" option, you can insert any whole name. **E.g. "test" as shown here.**

Contigs: The "contigs" option tells the program that our protein must be **100 amino acids** long.

Num_designs: The "num_designs" option tells how many proteins we want to design. In this case, we just need to design a protein.

We will keep the other settings at the default settings for now ;-)

When you're ready, you can press play. You will now be able to see the protein being designed:



- When the play button turns green, the protein is done. Congratulations: You have now designed your first protein!

Insert a screen shot of your protein here:

- We have now generated a protein structure, but what is actually the amino acid sequence for this brand new protein. Find this out by running the next part of the code that uses the two programs called Protein mpnn and Alpha Fold to figure this out. Use the code to generate one sequence for your protein structure by setting num_seqs=1

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



num_seqs: 1

[Show code](#)

- You now have the amino acid sequence for your new protein!

Paste the sequence here:

Question: By what "code" do the amino acids stand? (hint: how many letters are used to display an amino acid)

Find the one-box silence code for these amino acids

Arginine:

Aspartic acid:



Tyrosine:

Serine:

Part 3: Design snake antivenom

Now we have learned how to generate proteins with the computer. Now we are going to try to make a protein that can potentially bind and neutralize the snake toxin α -cobratoxin.

- Go back to the "Use RFDiffusion to make a protein" section. Here we need to change a little in the settings.
 - You can give your task a name under "name". You may want to use: " α -cobritoxin antivenom"
 - Under the "contigs" section of the program, we must specify the size of the protein we want to generate, as well as what this protein should bind to.
 - In our specific case, we want to construct an amino acid chain with a length of 100-150 amino acids, and we want this protein to be able to bind to our α -cobratoxin chain. On the PDB website, we can see that this chain is named "A." Therefore, we write "A:100-150" in the field.

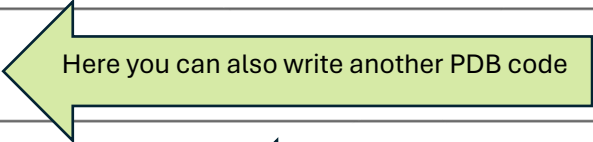
Entity ID: 1					
Molecule	Chains ⓘ	Sequence Length	Organism	Details	Image
ALPHA-COBRATOXIN	A 	71	Naja siamensis	Mutation(s): 0 ⓘ	

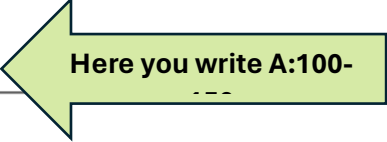
- Under the "pdb" point, we enter the pdb indicator, which in our example is "1CTX."
 - NOTE:** If you want to make an antivenom to another protein from PDB, then write the PDB code for that protein here (e.g. 4LFT is a toxin from *Black Mamba*)
- Finally, we list num_designs as 1, as we only want to design a single protein.
- When you're ready, hit the play button

> Brug RFdiffusion til at lave et protein

▶ **name:** " Slange_Modgift

contigs: " 100

pdb: " 1CTX 

hotspot: " Insert text here 

num_designs: 1

[Show code](#)

- You have now designed a snake antivenom protein!

Insert a picture of your antivenom below. If in the photo which part is your antivenom and which part is α -cobratoxin:

- Now you have formed your very own protein that can potentially bind to the snake toxin α -cobritoxin. But what is the sequence? Now press again play next code to find the sequence.

Paste your sequence here:

- With your fellow students, you will now investigate who designed the best protein (which depends largely on how good you have been, a bit like a competition to beat the highest number...)! You can do this by looking at the value called i_{pae} . This is a factor that describes how good the interaction is between your protein and snake venom. The lower a number, the better the interaction. Remember that this is a simulated interaction, and it is difficult to predict protein-protein interactions, so it must be verified in a laboratory! (If the number is below 5, it is most likely that in the real world your protein will be able to bind the snake toxin. However, this is rare, so each not surprised that few or none have a i_{pae} below 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
```

Write your `i_pae` score here and possibly also on the board so that the class can compare:

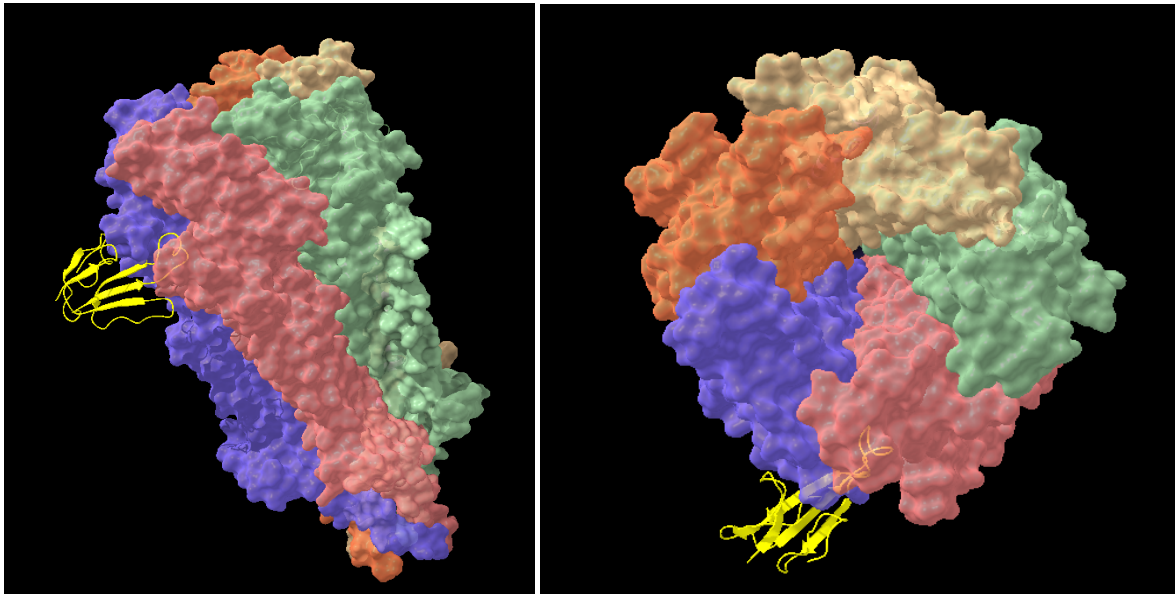
The figure shows two different structures. The gray structure is the one that RFdiffusion has designed, while the colored structure is the one that AlphaFold and proteinMPNN have predicted after making an amino acid sequence. These two structures are usually very similar, but in this example it looks like the colored structure is rotated relative to the gray one.



Part 4: Check if your antivenom is neutralizing:

Now you need to examine how your antivenom binds to snake venom, which is called orientation. This information allows you to assess whether your snake antivenom will have neutralizing properties or not. A protein can bind to snake venom without neutralizing it! To refresh what a neutralizing bond entails, take a closer look at Figure 10 under the section on [Antibodies and passive immunization](#).

1. When envenomed with α -cobrotoxin, it binds to the nicotinic acetylcholine receptor (nAChR), leading to paralysis in the body. In the figure below, in yellow you can α -cobrotoxin (in yellow) bound to the large receptor nAChR (the large one with many colors). Both images show the same bond but from two angles.



2. Based on your knowledge of the way α -cobrotoxin normally binds, assess whether your designed protein will be neutralizing or not.

Extra task

- We haven't used the hotspot feature yet. This function tells the program how exactly to bind to the protein. Try running a design round where you select hotspot A33 (You do this by typing A33 in the "Hotspot" field before generating your antivenom.)
 - What's going on?
 - Why is this smart?

Reflection

- Imagine what one or could use RFDiffusion for. Make three suggestions.
- Why can a protein bind to the snake antivenom without neutralizing it?
- What things determine if two proteins can bind to each other? What forces are at play?
- Do you know any examples of proteins that bind to other proteins (hint: peptide hormones)

- How would you be able to verify in a laboratory whether your proteins actually bind to snake venom?