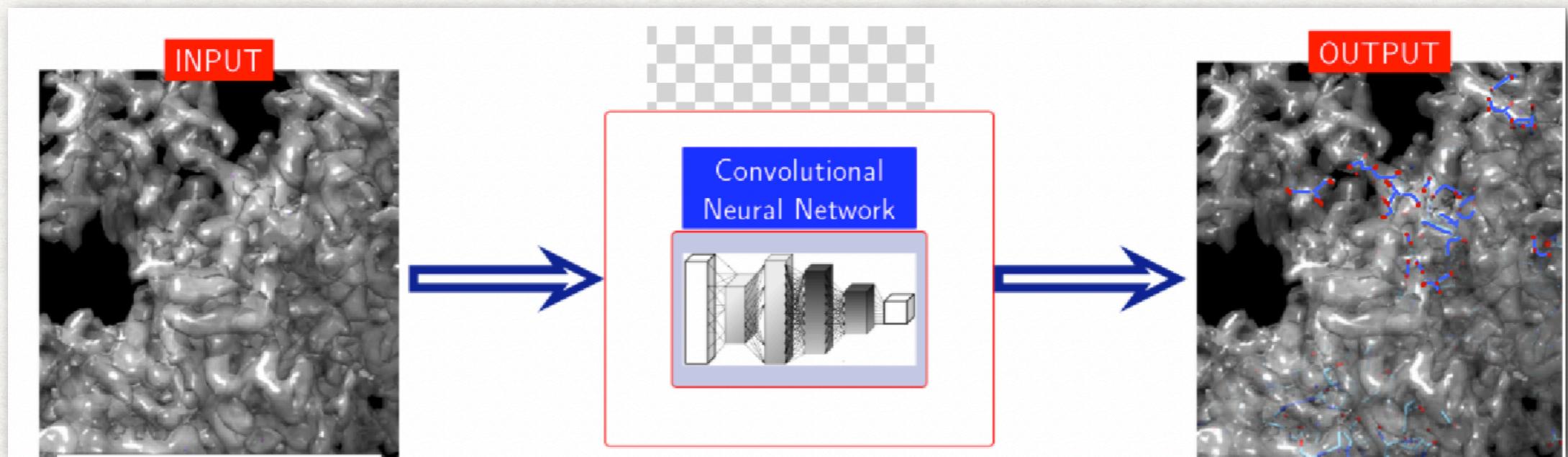
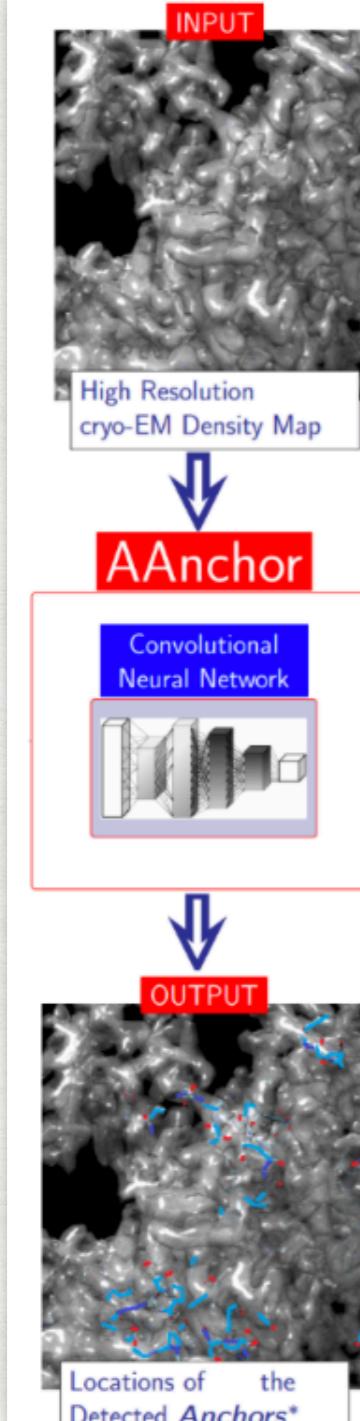


AANCHOR 2

IMPROVE THE DETECTION USING REALISTIC SIMULATION





AAnchor •

The Task

- **Input:** cryo EM map at $\sim 3\text{\AA}$ resolution
- **Output:** Locations and types of amino acids in the map with less than 20% of errors.

Motivation

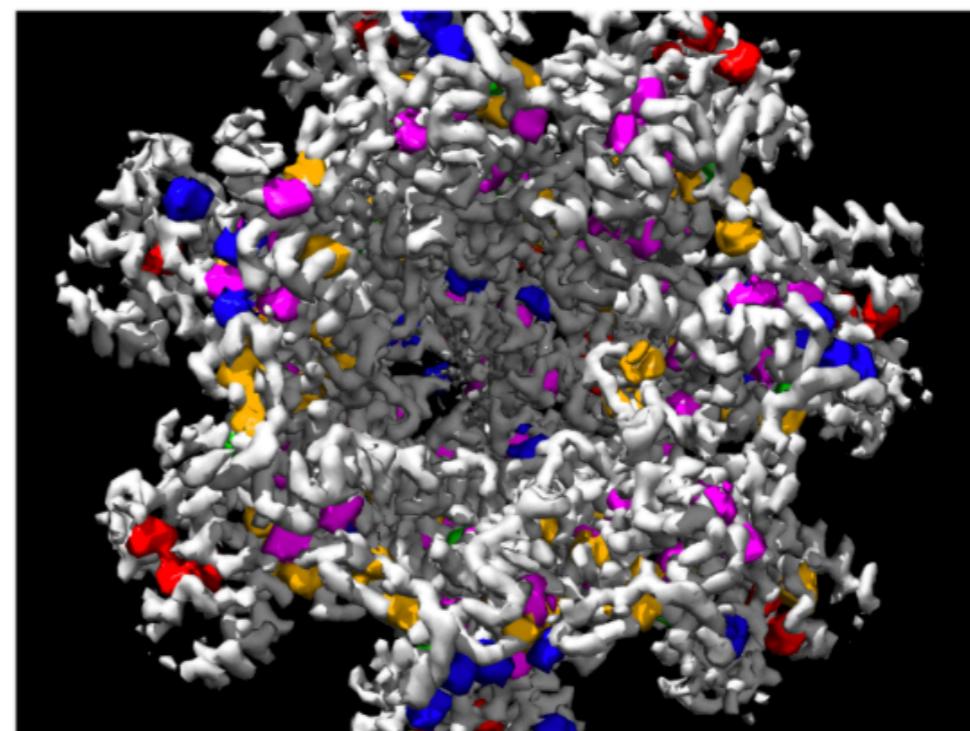
- Assistance in manual structure modelling
- Feed to a de-novo modelling method
- Input to a refinement method

Results: resolution 2.9 Å

- Test Map : EMD-6224, 2.9Å resolution cryo-EM structure of anthrax toxin protective antigen pore
- Only Anchors Shown: Less then 20% errors in reported detections

Reported Detections:

<i>ASN</i>	5	<i>out of</i>	288
<i>ARG</i>	20	<i>out of</i>	119
<i>LEU</i>	40	<i>out of</i>	241
<i>LYS</i>	30	<i>out of</i>	189
<i>PRO</i>	80	<i>out of</i>	133
<i>TYR</i>	18	<i>out of</i>	91
<i>VAL</i>	35	<i>out of</i>	161

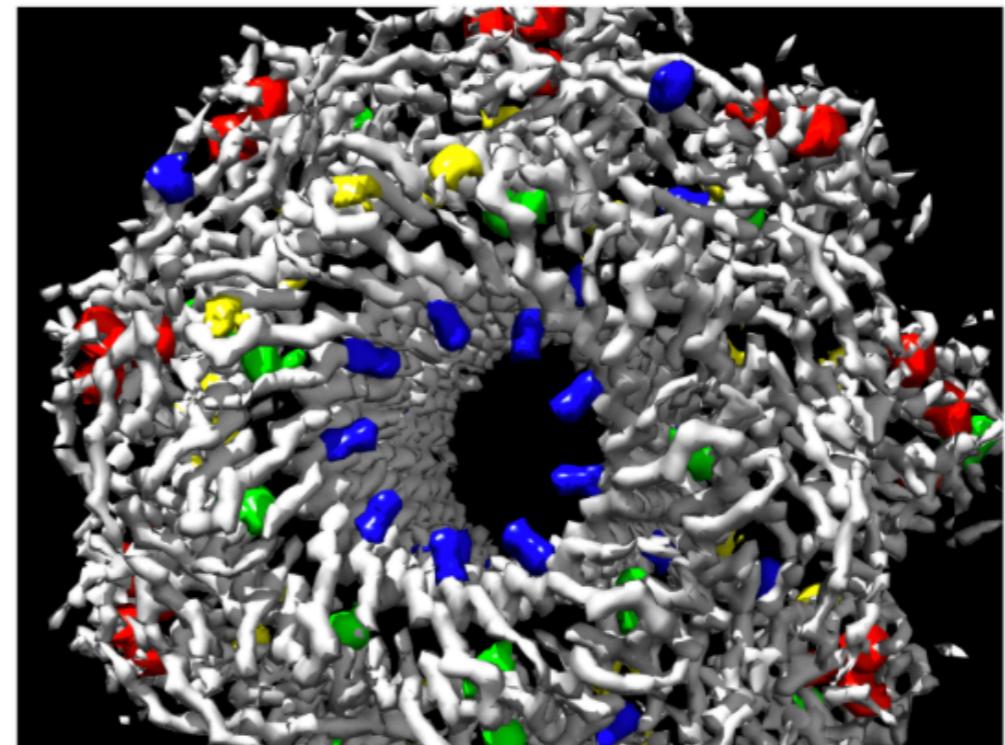


Results: resolution 3.1 Å

- Test Map : EMD-8015, 3.1 Åresolution cryo-EM structure of the Lysenin Pore
- Only Anchors Shown: Less then 20% errors in reported detections

Reported Detections:

<i>ASN</i>	5	<i>out of</i>	288
<i>ARG</i>	20	<i>out of</i>	119
<i>LEU</i>	40	<i>out of</i>	241
<i>LYS</i>	30	<i>out of</i>	189
<i>PRO</i>	80	<i>out of</i>	133
<i>TYR</i>	18	<i>out of</i>	91
<i>VAL</i>	35	<i>out of</i>	161



HIGH RESOLUTION: LOCATING AMINO ACIDS CLASSIFICATION CNN

The Task

- **Input:** cryo EM map at $\sim 3\text{\AA}$ resolution
 - **Output:** Locations and types of amino acids in the map with less than 20% of errors.
- AANCHOR PERFORMANCE IMPROVEMENT
1. Use X-ray crystallography Data
 2. Realistic cryo-EM simulation

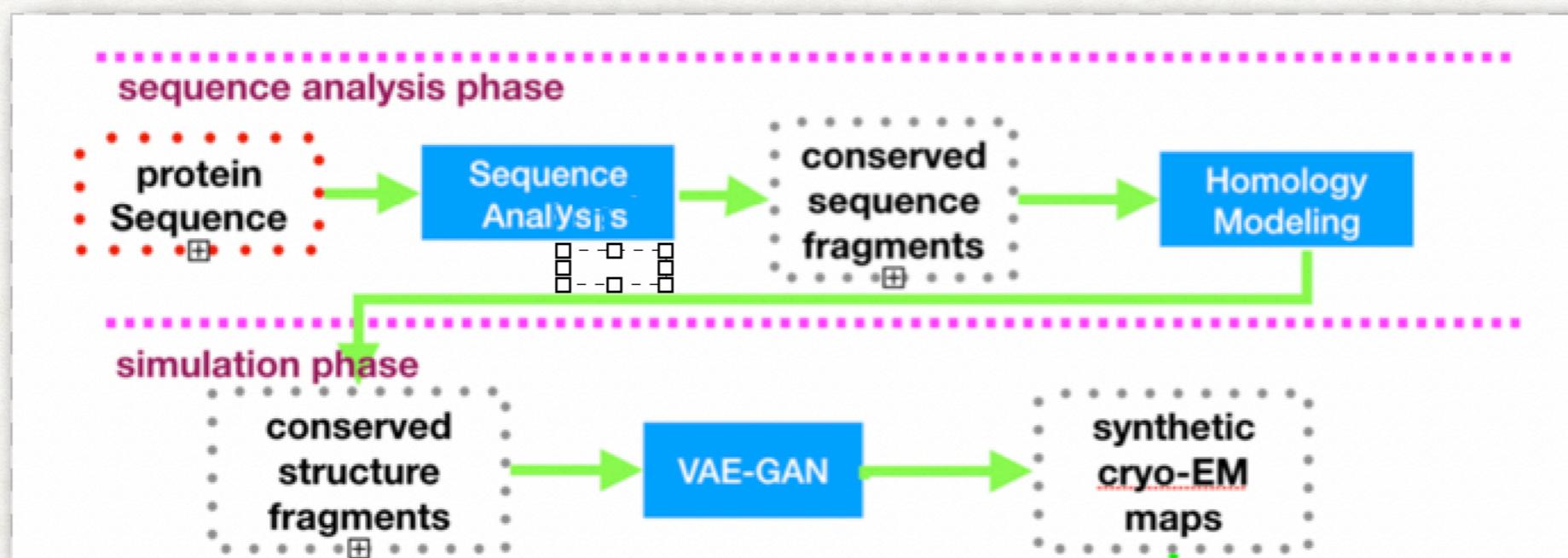
HIGH RESOLUTION: LOCATING AMINO ACIDS ADVANTAGES

- Sequence information is utilized in the anchor location task.
- Using homolog structures ensures that the training dataset and the query map have similar feature distributions.
- Training datasets of arbitrary size can be created.

RESEARCH GOALS

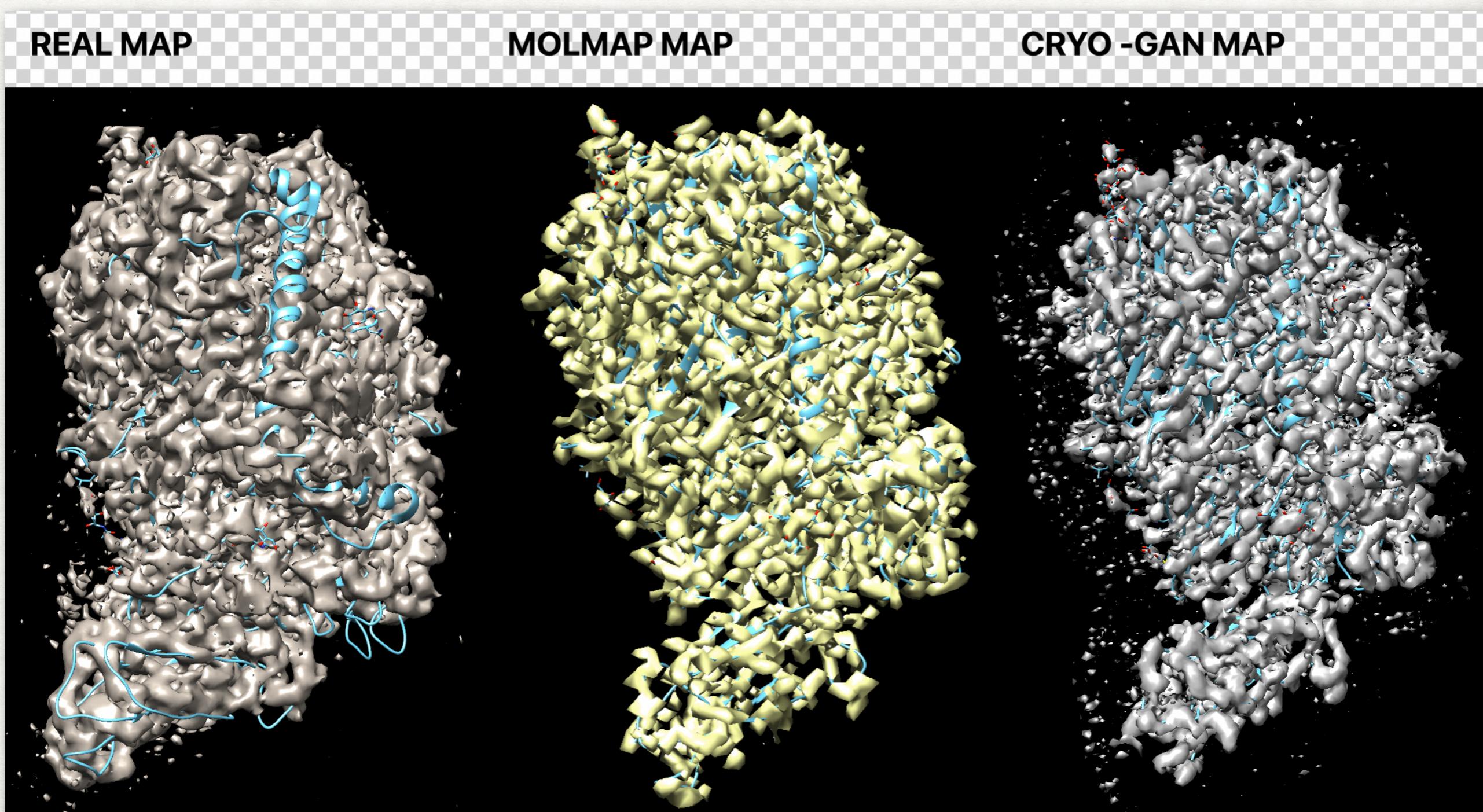
CREATE TRAINING DATASET WITH SIMILAR FEATURES
DISTRIBUTION
USE SEQUENCE DATA

- Find Sequence homologs
- Extract \Create Structures
- **Simulate Cryo-EM map**

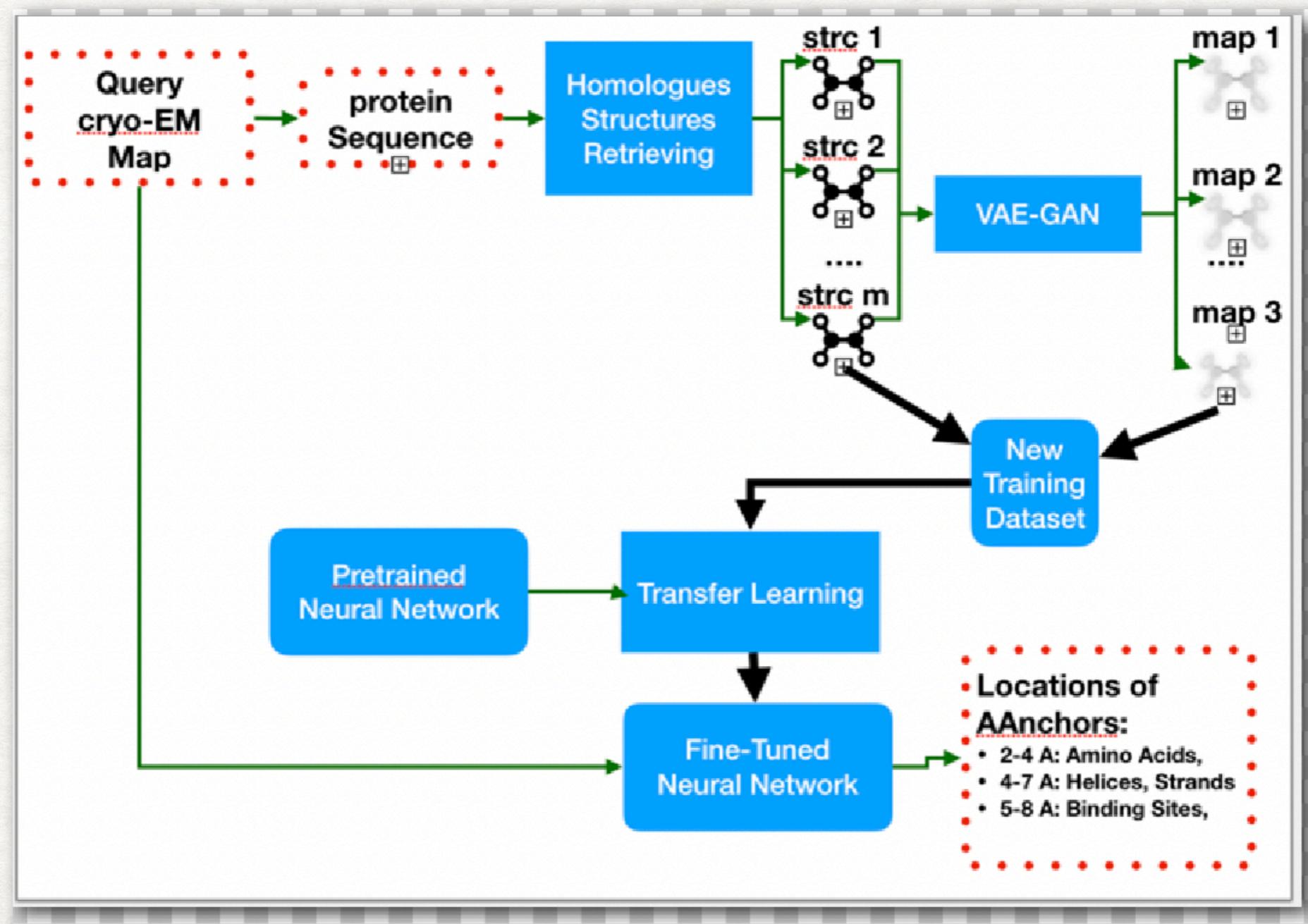


CURRENT STATUS

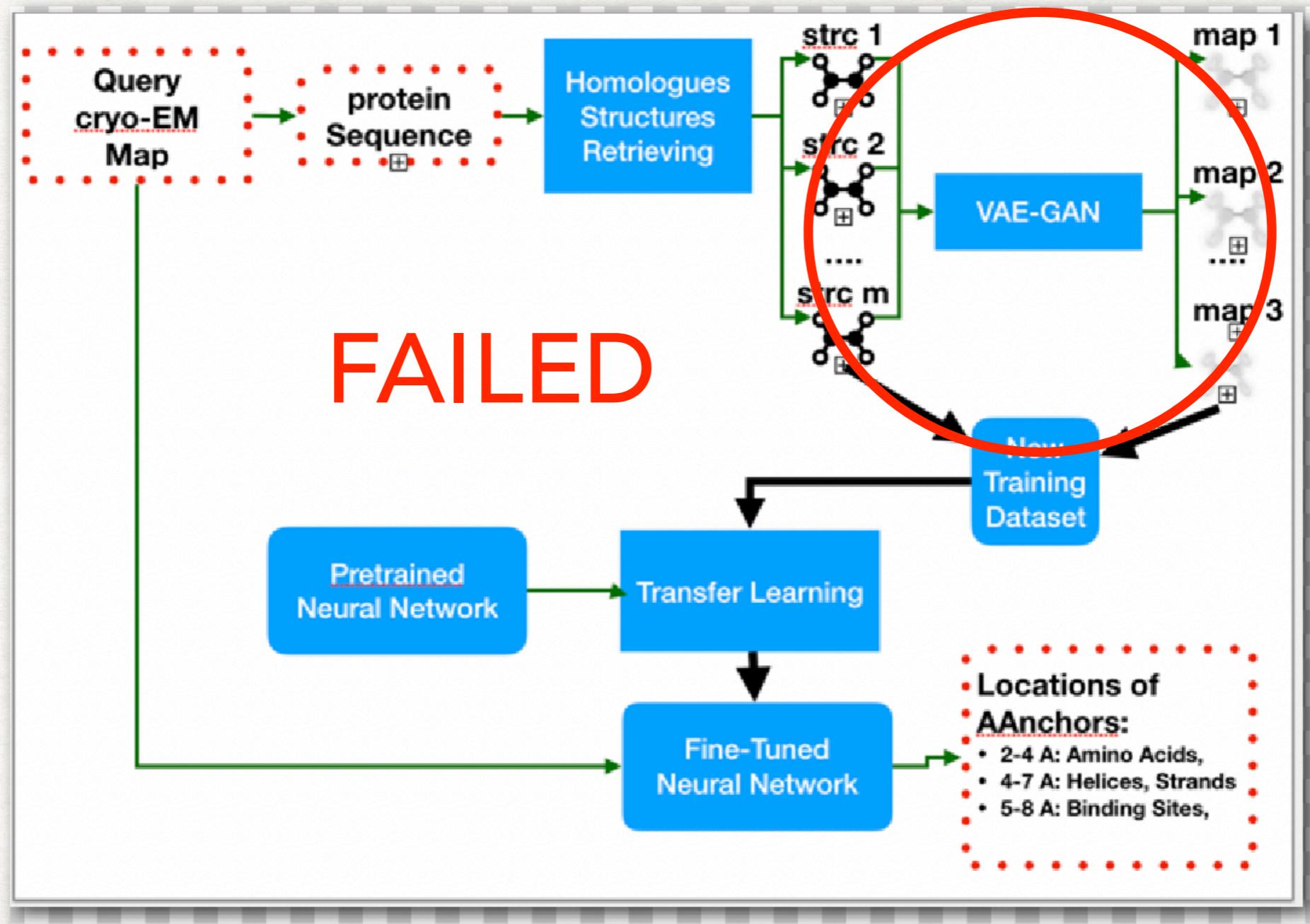
3 ANGSTREM SIMULATION



SEMI SUPERVISED LEARNING -



SEMI SUPERVISED LEARNING -



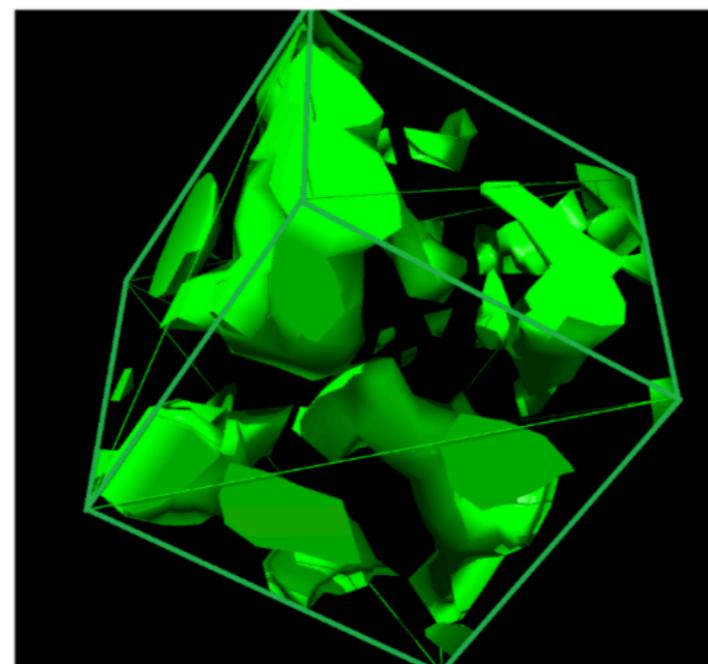
PROBLEMS (?):

1: NEGATIVE SAMPLES

2: SAMPLING

Detection: The Cube

- 3D Matrix $10\text{\AA} \times 10\text{\AA} \times 10\text{\AA}$
- A cube is labeled as an amino acid of type X if:
 - A center of mass of an Amino Acid of type X is within 1.5\AA from the cube's center
- otherwise the cube is labeled as NONE



PROPOSED SOLUTION

- Change Anchor to 20 labels (without NONE)
- Create “boxes” directly from simulation

PROBLEM (?) - NEGATIVE SAMPLES

- GAN simulates box from structure, but can't simulate "empty" space

CURRENT PROCESSING LINE:



NEW PROCESSING LINE:

