Antigen-Specific Antibody Design with Diffusion-Based Generative Models

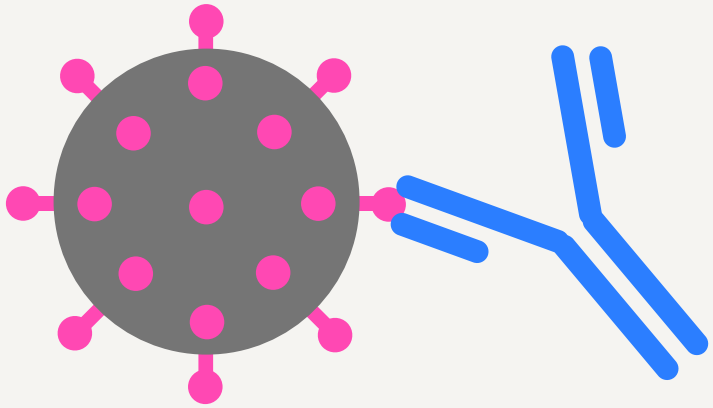
SEMINAR ADVANCED MACHINE LEARNING IN BIG DATA ANALYTICS

Marcel Nieveler

Based on: Shitong Luo, et. al., "*Antigen-Specific Antibody Design and Optimization with Diffusion-Based Generative Models for Protein Structures*"

Intoduction

- Antibodies are a vital part of immune system
- Immune response
- Antobodies detect pathogenes



Contents

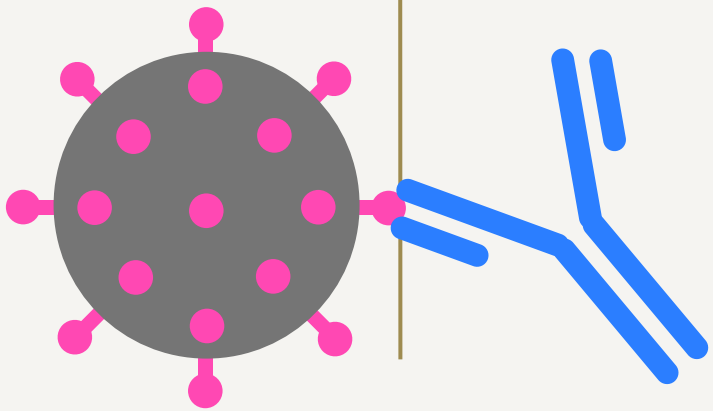
- What are antigens and antibodies?
- What are diffusion models?
- How does the diffusion model for antibody design work?
- What are applications of it?

Antigens and

- Large molecules
- For this presentation: only proteins
- On the surface of cells, viruses, fungi, or bacteria

Antibodies

- Protein structures
- Detect and bind to antigens
- Trigger immune response



3D Structure of an Antibody

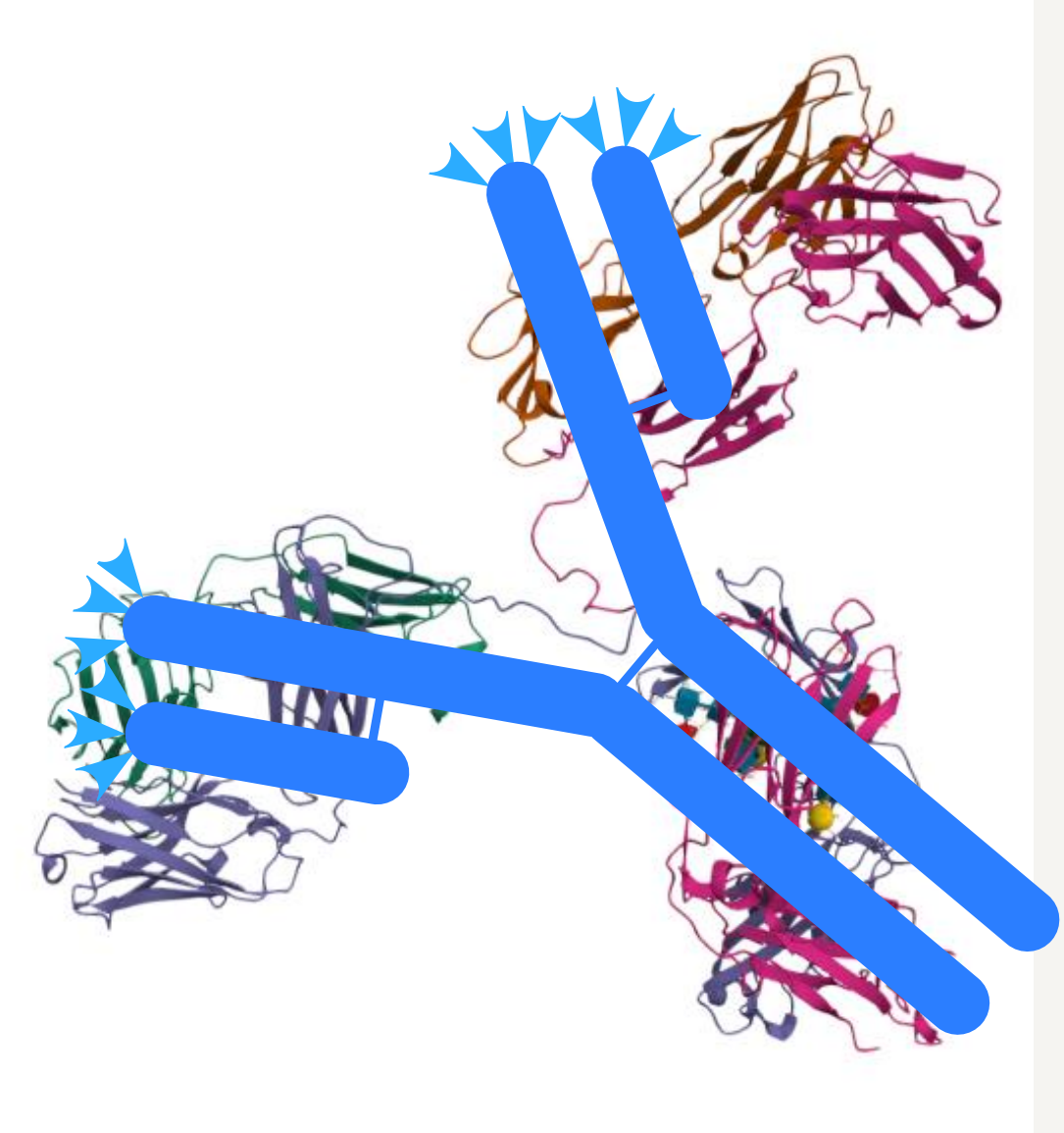
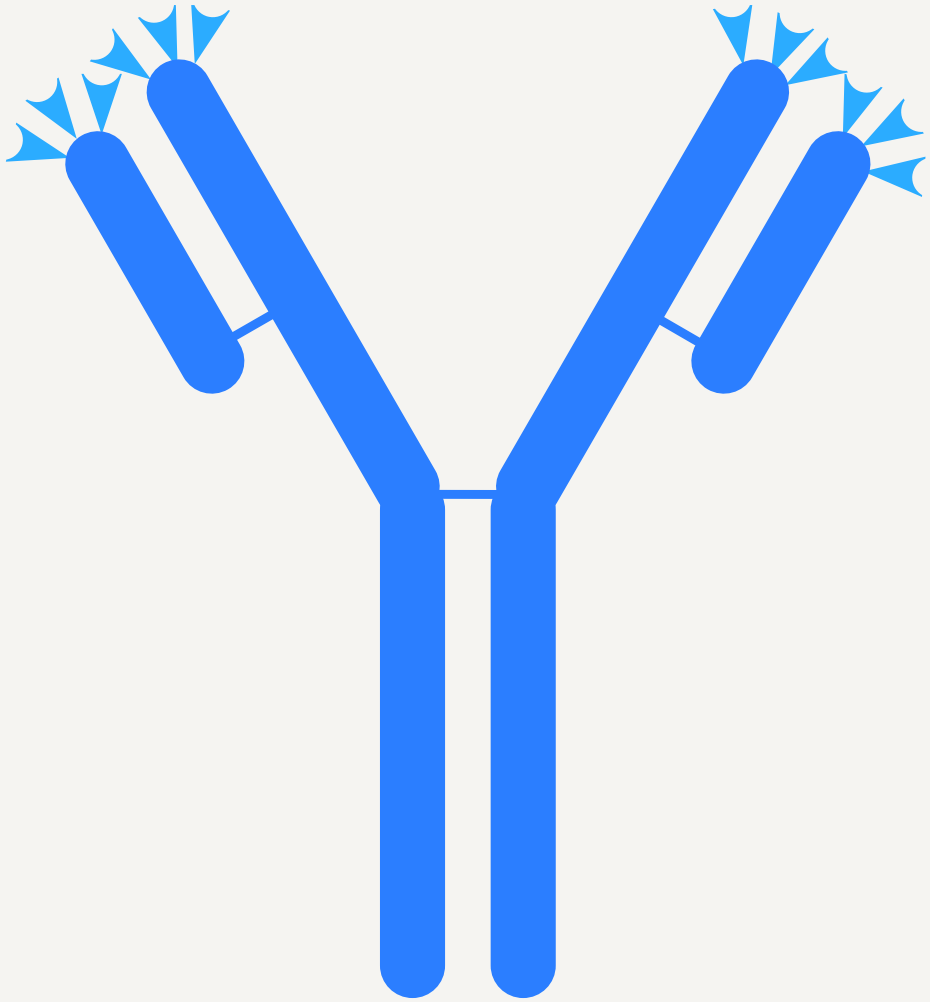


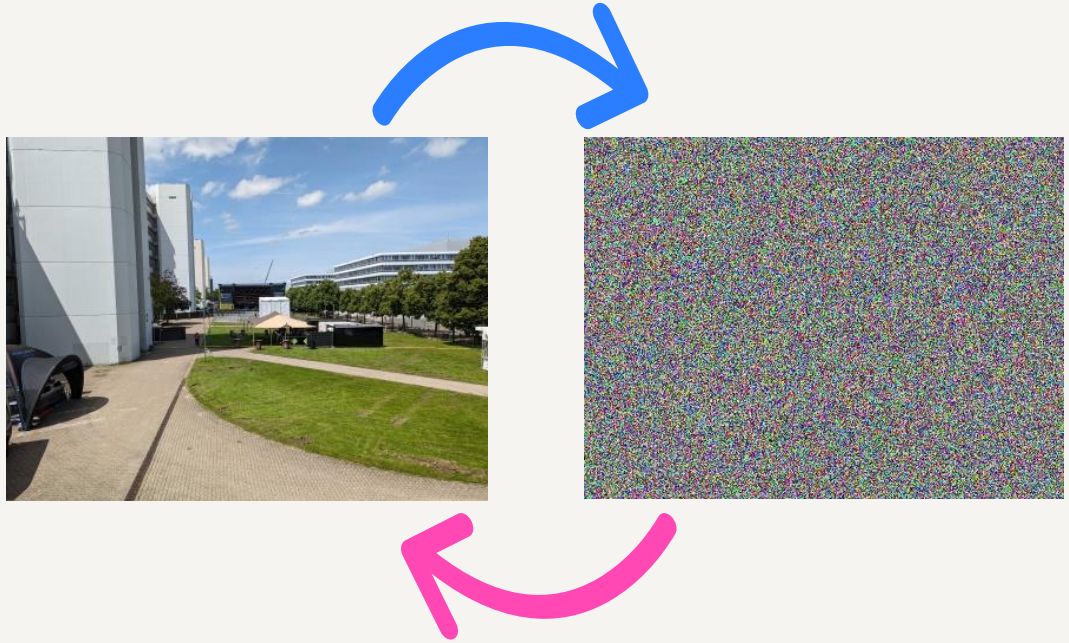
Image Source: RCSB PDB entry 1IGT

Structure of Anibodies

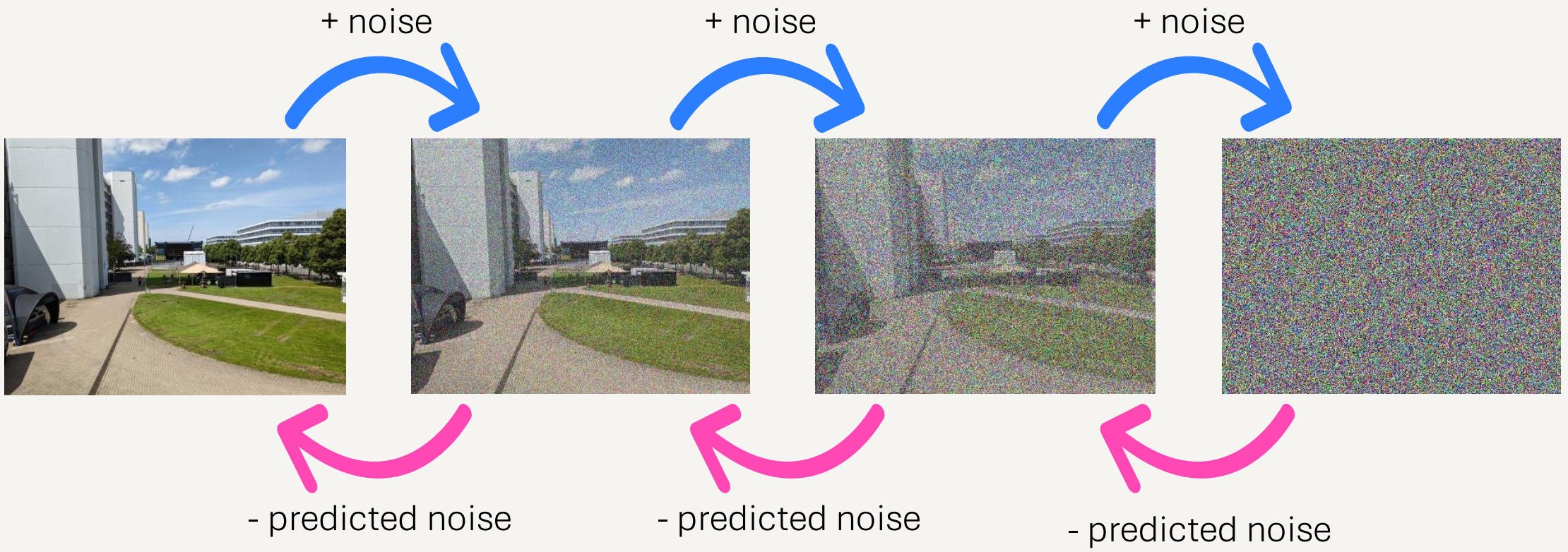
- 2 light chains
- 2 heavy chains
- 6 different complementary determining regions (CDRs)



Diffusion (Models) for Images



Diffusion Models for Images



Model for Antibody Design

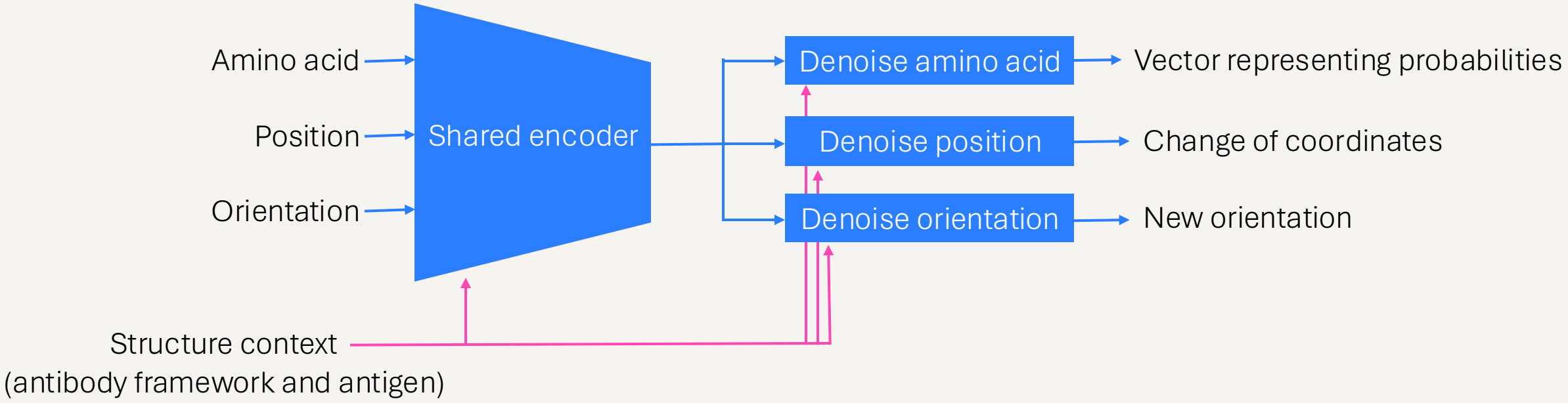
- Conditioning on 3D structure of the antigen to fit it
- State space of model
  - Amino Acid
  - Position
  - Orientation

Training

Variable	Domain	Diffusion Noise	Objective
Amino acid	{ACDEFGHIKLMNPQRSTVWY}	Multinomial	Kullback-Leibler divergence
(Normalized) Position	$\mathbb{R}^3$	Gaussian	Mean squared error
Orientation	SO(3)	Iterative pertubation scheme	Discrpancy of real and predicted orientation matrix

Architecture

- 4 multi layer perceptrons



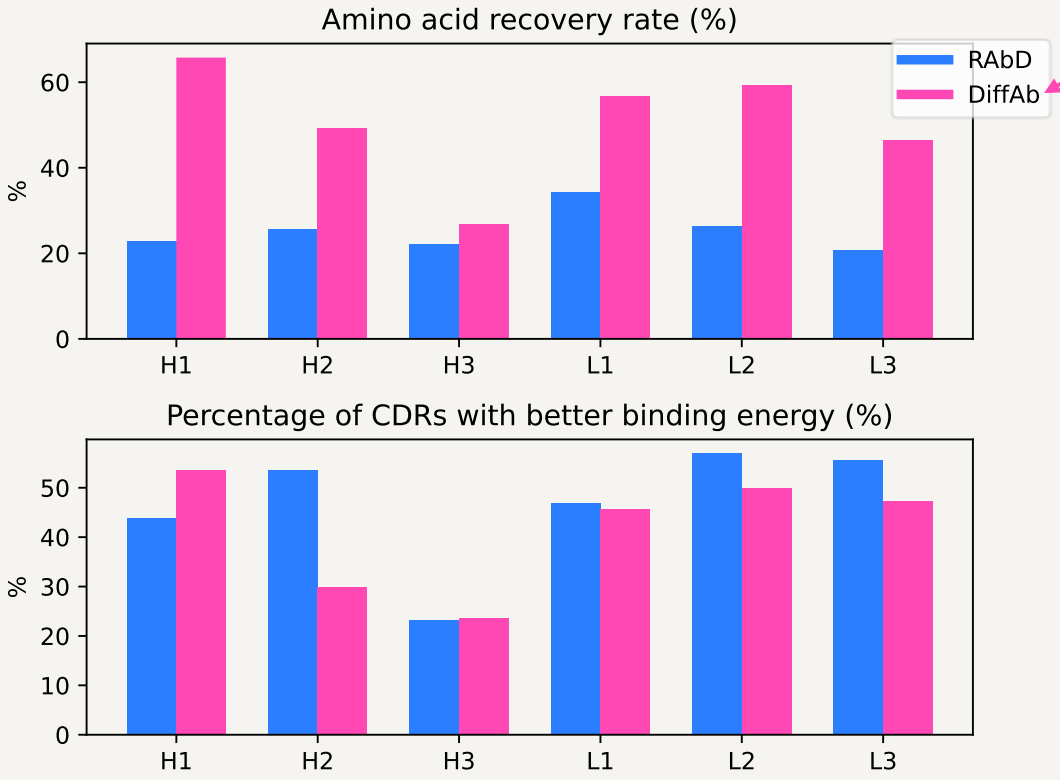
Sampling

1. Initialize with arbitrary
    - Sequence
    - Positions
    - Orientations
2. Iteratively update values
3. Refinement with OpenMM and Rosetta

Sequence-Structure Co-design

1. Remove CDR
2. Sample new sequence and structure
    - Set length to original length

Sequence-Structure Co-design: Results



Method from paper

- Sequences are recovered more accurately

- Lower increase in binding energy
  - But RAbD optimizes binding energy directly

Antibody Optimization

1. Pertub the CDR sequence and structure for *t* steps
2. Denoise for *t* steps

Altering Images

+ noise (60%)

- predicted noise

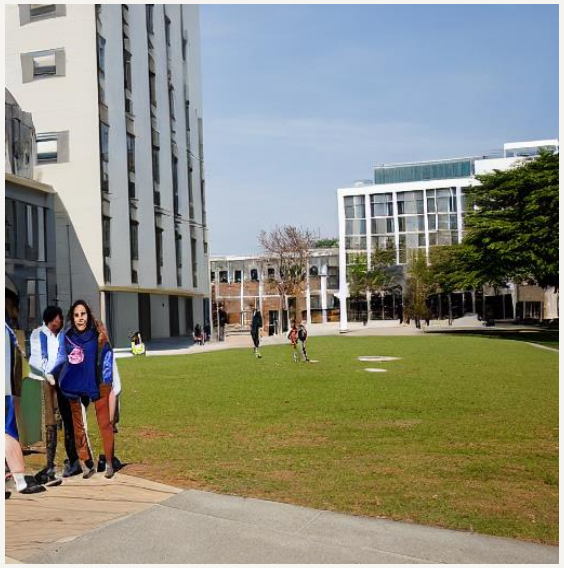
Conditioned on: "University campus with students"

Altering Images

Original

30%

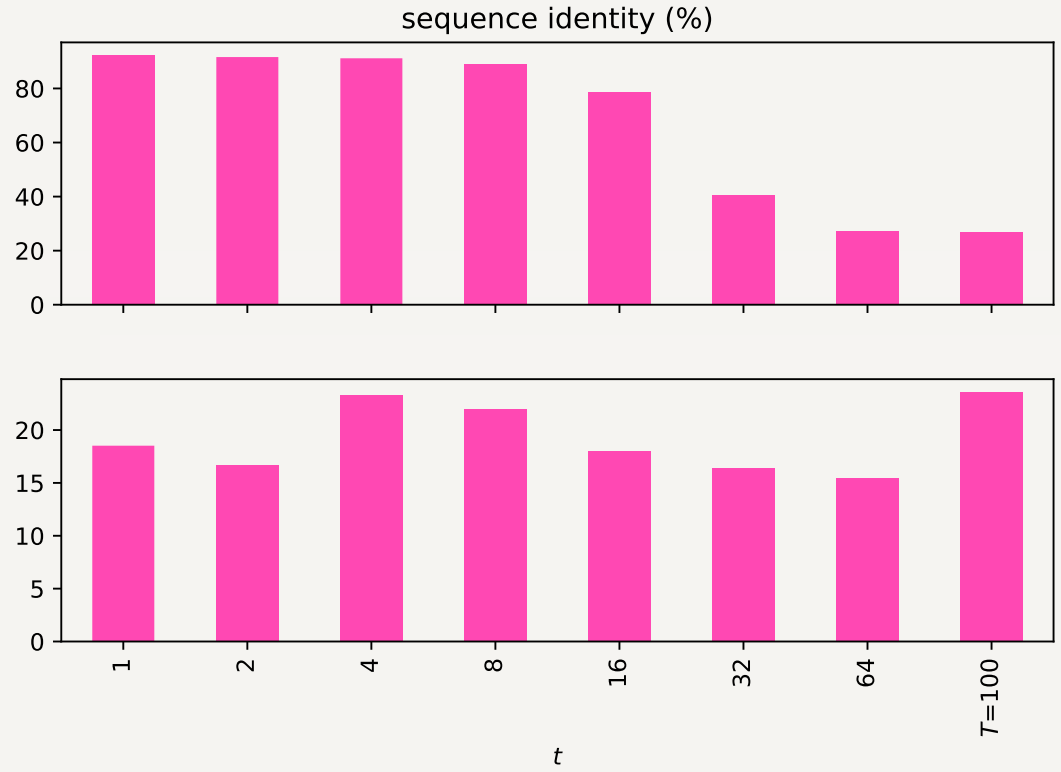
60%



80%

Conditioned on: "University campus with students"

Antibody Optimization: Results



- Sequence similar to original
- Improved binding energy

Summary

- Diffusion models reverse random pertubation
- Design of antibody CDR with competative performance
- Multiple use-cases