ICML

International Conference
On Machine Learning

Antibody Design Using a Score-based Diffusion Model Guided by Evolutionary, Physical and Geometric Constraints

Authors: Tian Zhu[1,2], Milong Ren[1,2], Haicang Zhang[1,2,#]

Presenter: Tian Zhu

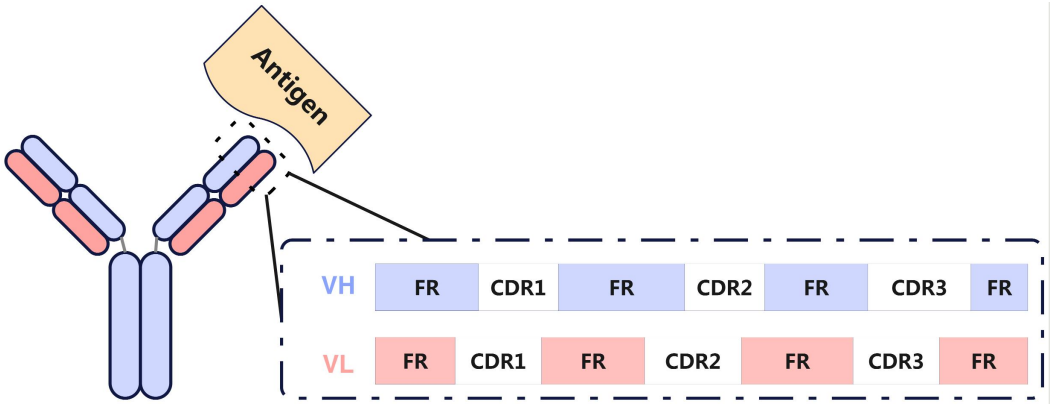
1. Institute of Computing Technology, Chinese Academy of Sciences.
2. University of Chinese Academy of Sciences.

Background

- Antibodies are central proteins in adaptive immune responses, responsible for protecting against viruses and other pathogens.

- The framework regions of antibodies exhibit high conservation, their complementarity-determining regions (CDRs) are variable, mainly determining the binding affinity and specificity to antigens

- Therefore, the primary objective of rational antibody design is to optimize the CDRs for effective binding to the targeted antigen.



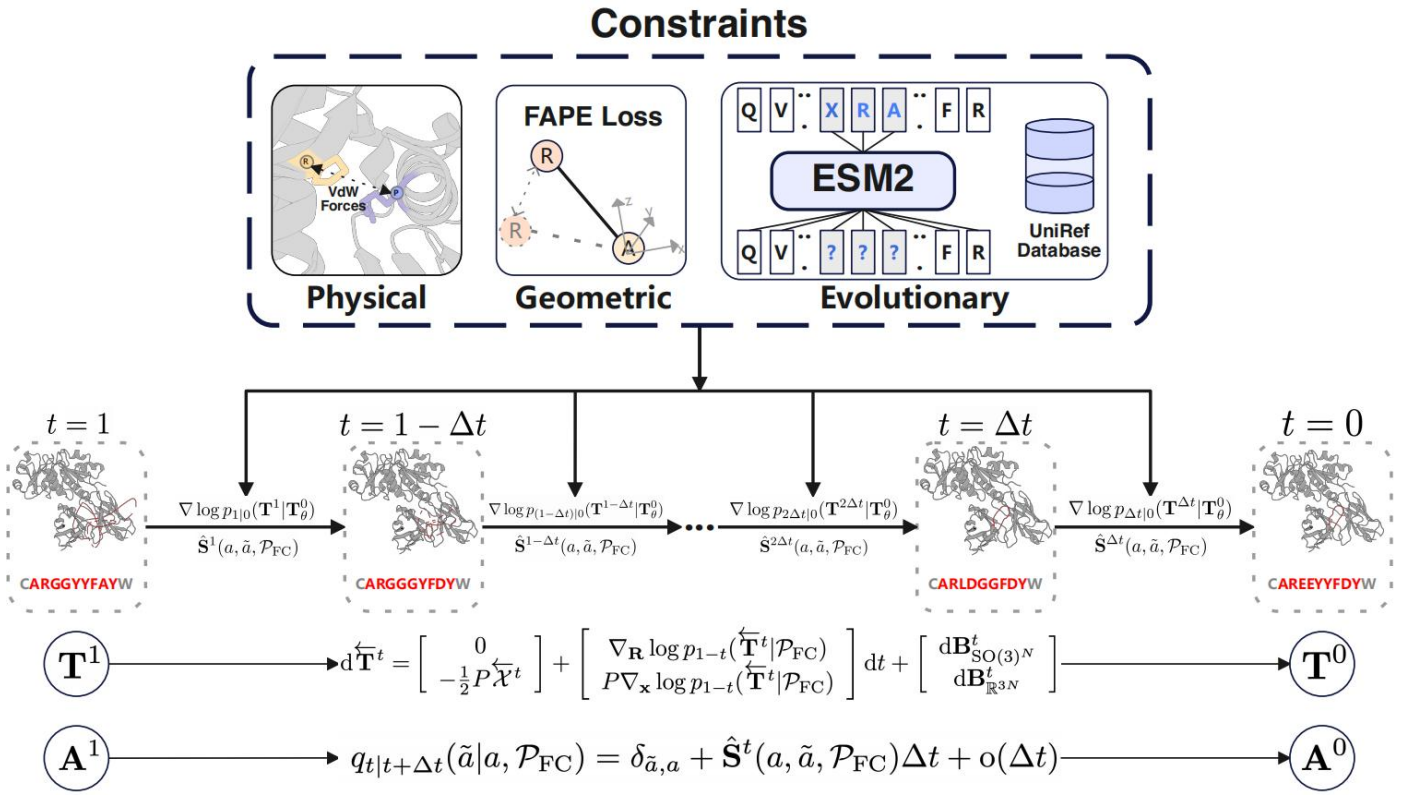
Motivation

- The scarcity of available antibody-antigen complex data poses a significant challenge for these diffusion-based generative models.

- For the case of antibody design, the widely-used SAbDab database of antibody-antigen complexes comprises only thousands of non-redundant samples, leading to a high risk of overfitting.

Method-AbX

- We develop AbX, a new score-based diffusion model that is guided by evolutionary, physical, and geometric constraints.
- These constraints serve as sequence and structure priors, narrowing the exploration to a more plausible space and mitigating the risk of overfitting.



Method-AbX

ICML
International Conference
On Machine Learning

- **Evolutionary Constraints:**

  Recognizing the superiority of general protein language models in enhancing the evolutionary plausibility of antibody designs. We have integrated ESM-2 (3B) as our evolutionary constraint in the score network.

- **Geomtric Constraints:**

  The geometric constraint is specifically formulated to accurately depict the rigidity and flexibility inherent in antibody structures. For the CDRs structures, we incorporate the FAPE loss, distogram loss and lDDT-Ca loss, aiming to generate more rational structure. The geometric constraint is thus defined as:

$$\mathcal{L}_{\text{Geometric}} = \mathcal{L}_{\text{FAPE}} + 0.5\mathcal{L}_{\text{distogram}} + 0.1\mathcal{L}_{\text{lDDT}}.$$

- **Physical Constraints:**

  To guide the generation of antibodies with high binding affinity to target antigens, we included a structural violation loss to prevent violations in covalent peptide bond angles and lengths among neighboring residues, and a van der Waals loss to approximate the van der Waals forces within neighboring non-bonded backbone atoms.

$$\mathcal{L}_{\text{Physical}} = 0.03\mathcal{L}_{\text{VdW}} + 0.03\mathcal{L}_{\text{violation}}.$$

Results

- **Experiment 1: Sequence and Structure Co-design**
  - Dataset: RAbD test dataset
  - Conclusion
    - AbX outperforms other methods in each metric.
    - AbX exhibits a significant improvement in IMP and Plausibility, indicating the efficacy of the introduced constraints in generating more plausible antibodies capable of binding to target antigens.

Table 1. Evaluation of *de novo* designed CDRs in RAbD test dataset.

Metrics	DiffAb	dyMEAN	AbX
IMP(%)↑	12.07	0.00	**18.64**
Plausibility ↑	-1.38	-1.21	**-1.01**
Loop AAR↑	21.25	22.25	**30.80**
Loop RMSD↓	3.45	5.14	**3.24**

Table 2. Evaluation of *de novo* designed CDRs across each CDR in RAbD test dataset.

CDR	Method	AAR(%) ↑	RMSD(Å)↓	CDR	Method	AAR(%) ↑	RMSD(Å) ↓
H1	DiffAb	70.01	0.88	L1	DiffAb	61.07	0.85
	dyMEAN	75.71	1.09		dyMEAN	75.55	1.03
	AbX	**80.72**	**0.85**		AbX	**79.37**	**0.78**
H2	DiffAb	38.52	0.78	L2	DiffAb	58.58	0.55
	dyMEAN	68.48	1.11		dyMEAN	83.09	0.66
	AbX	**70.73**	**0.76**		AbX	**84.53**	**0.45**
H3	DiffAb	28.05	2.86	L3	DiffAb	47.57	1.39
	dyMEAN	37.50	3.88		dyMEAN	52.11	1.44
	AbX	**45.18**	**2.50**		AbX	**65.92**	**1.18**

Results

- **Experiment 2: Antibody Optimization**
  - Dataset: DiffAb test dataset
  - Conclusion
    - The antibodies optimized by AbX consistently exhibit higher binding affinity than those optimized by DiffAb across different noise scales. This highlights the superior efficacy of AbX in antibody optimization.

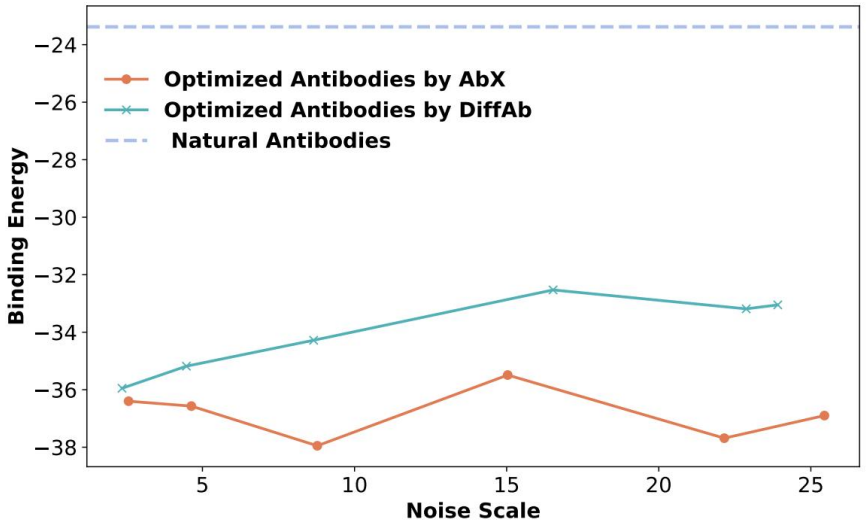


Figure 3. Binding energy of designed antibody-antigen complex in different noise scales. Binding energy is used as an approximation for binding affinity in antibody-antigen interactions.

Results

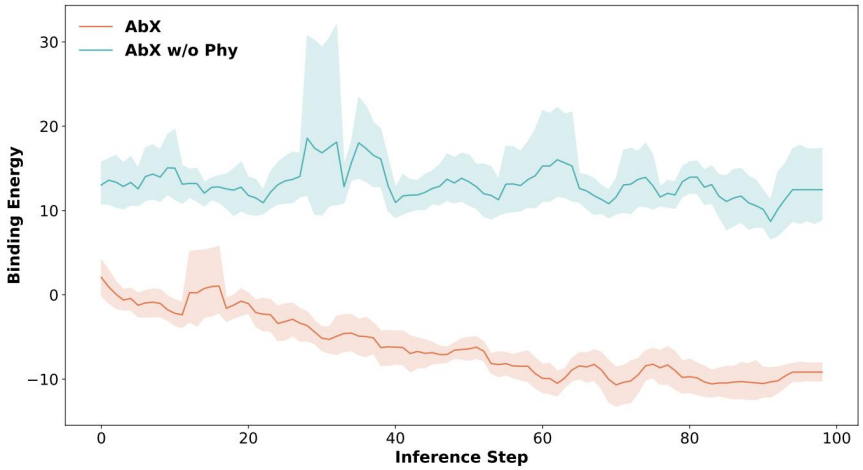
- **Experiment 3: Ablation Studies**
  - Dataset: DiffAb test dataset
  - Conclusion
    - It illustrates that the inclusion of geometric, physical, and evolutionary constraints significantly contributes to the enhanced performance of AbX.

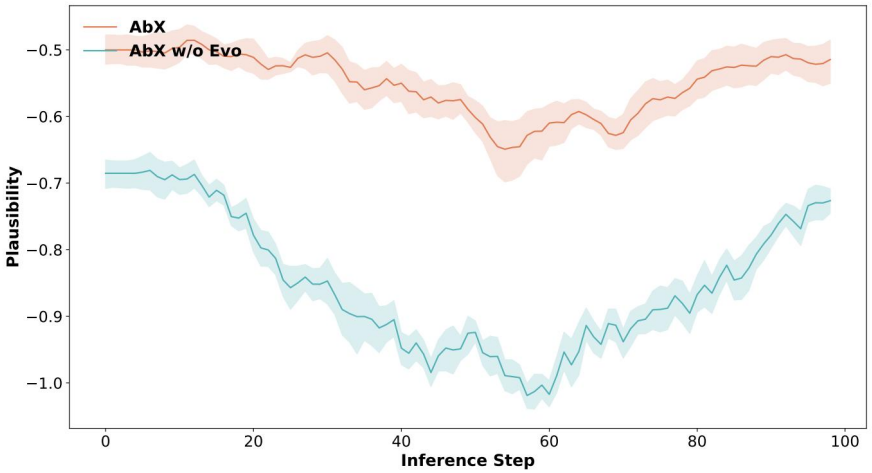
Table 4. Ablation studies for AbX in DiffAb test dataset.

Geometric Constraint	Physical Constraint	Evolutionary Constraint	IMP (%)	Plausibility	H3 AAR(%)	H3 RMSD
✓	✓	✓	**54.82**	**-0.67**	49.17	**2.68**
✓	✓	✗	46.50	-0.77	45.32	3.19
✗	✗	✓	19.36	-0.70	53.21	3.62
✓	✗	✓	52.02	-0.69	**53.84**	2.86

Results

- **Experiment 4: Case Studies on Trajectories of Antibody Design**
  - Case: 5TLJ
  - Conclusion
    - The binding energy decreases progressively during the inference process for all models.
    - AbX consistently yields antibodies with higher evolutionary plausibility.





Results

- **Experiment 4: Case Studies on Trajectories of Antibody Design**

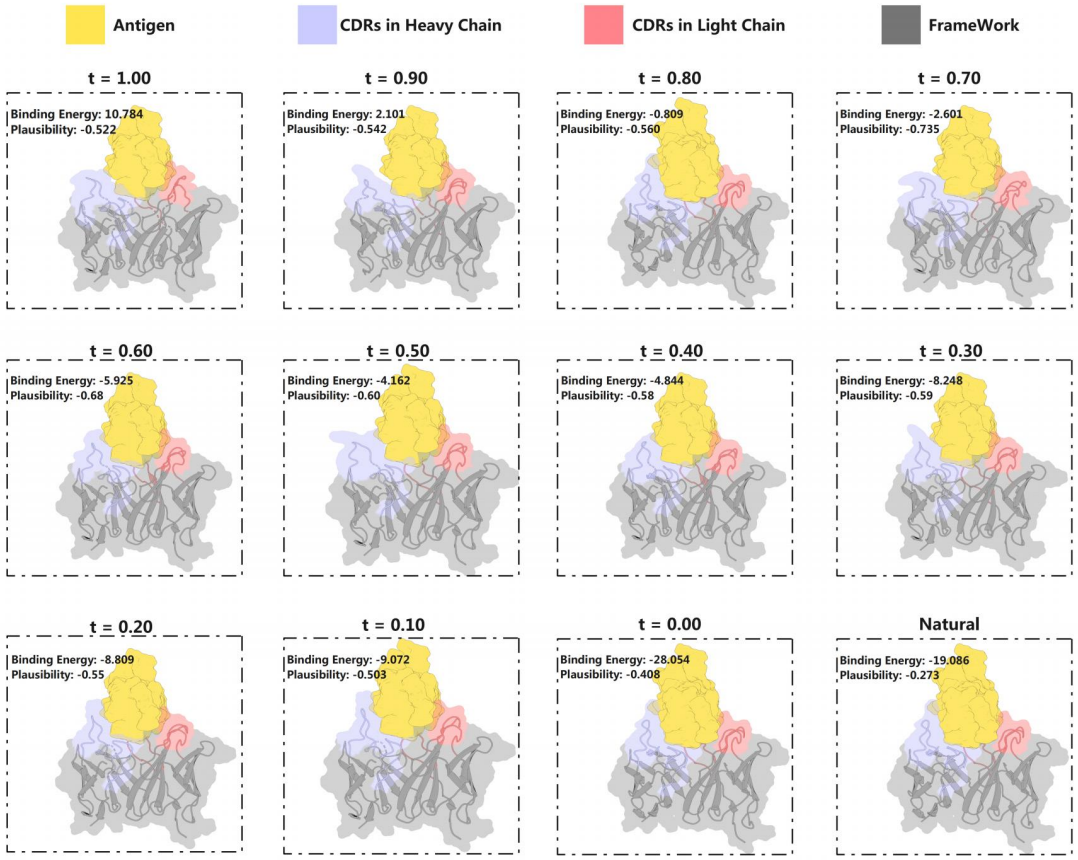


Figure S2. Visualization of generated antibody-antigen complexes during the generative process. The heavy, light, and antigen chains of the antibody-antigen complex (PID:5TLJ) are denoted as D, C, and X.

ICML

International Conference
On Machine Learning

Thanks