



AbODE: Ab initio Antibody Design using Conjoined ODEs

International Conference in Machine Learning (ICML), 2023



Yogesh Verma¹



Markus Heinonen¹



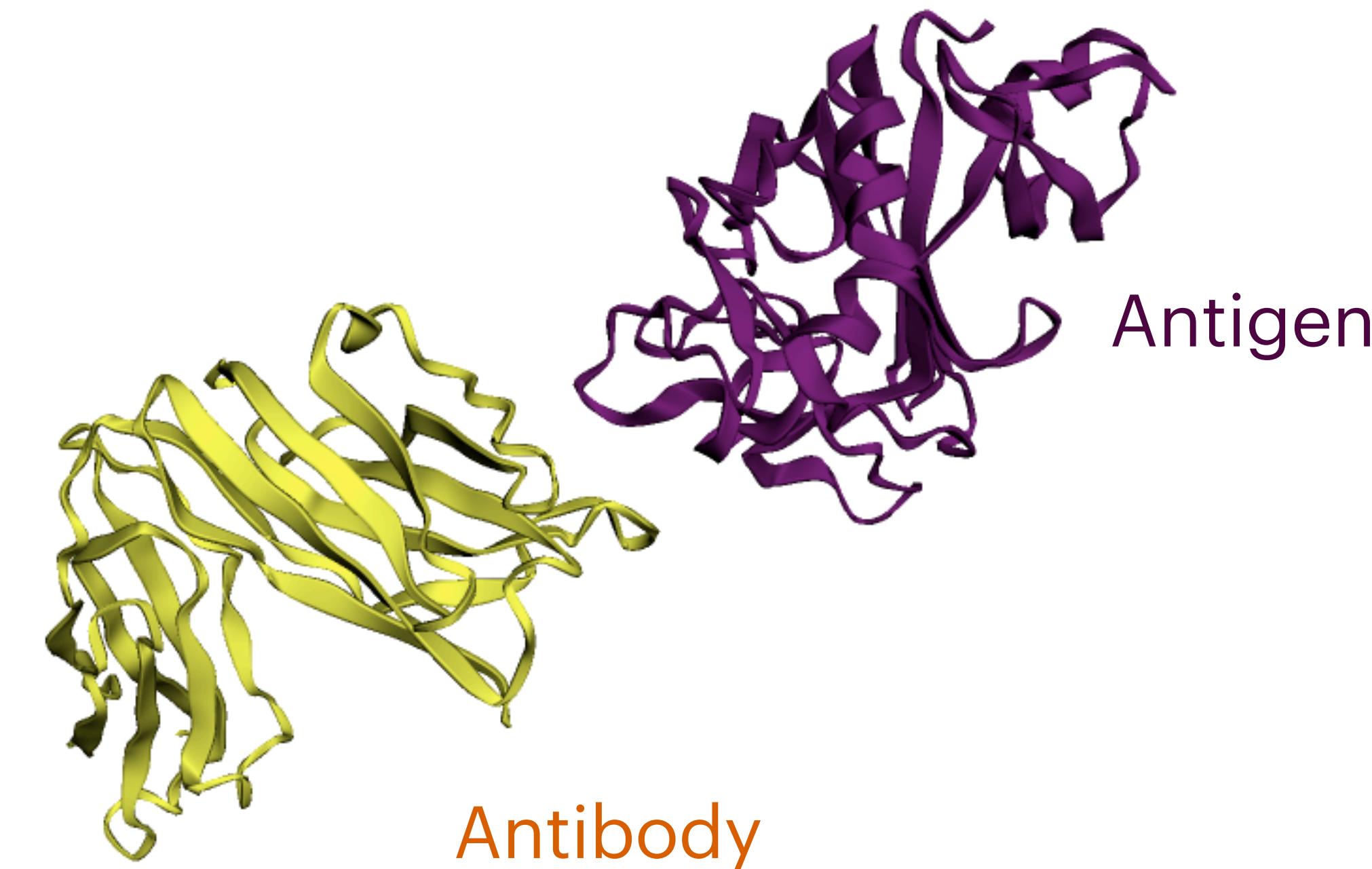
Vikas Garg^{1,2}

¹ Aalto University

² YaiYai Ltd

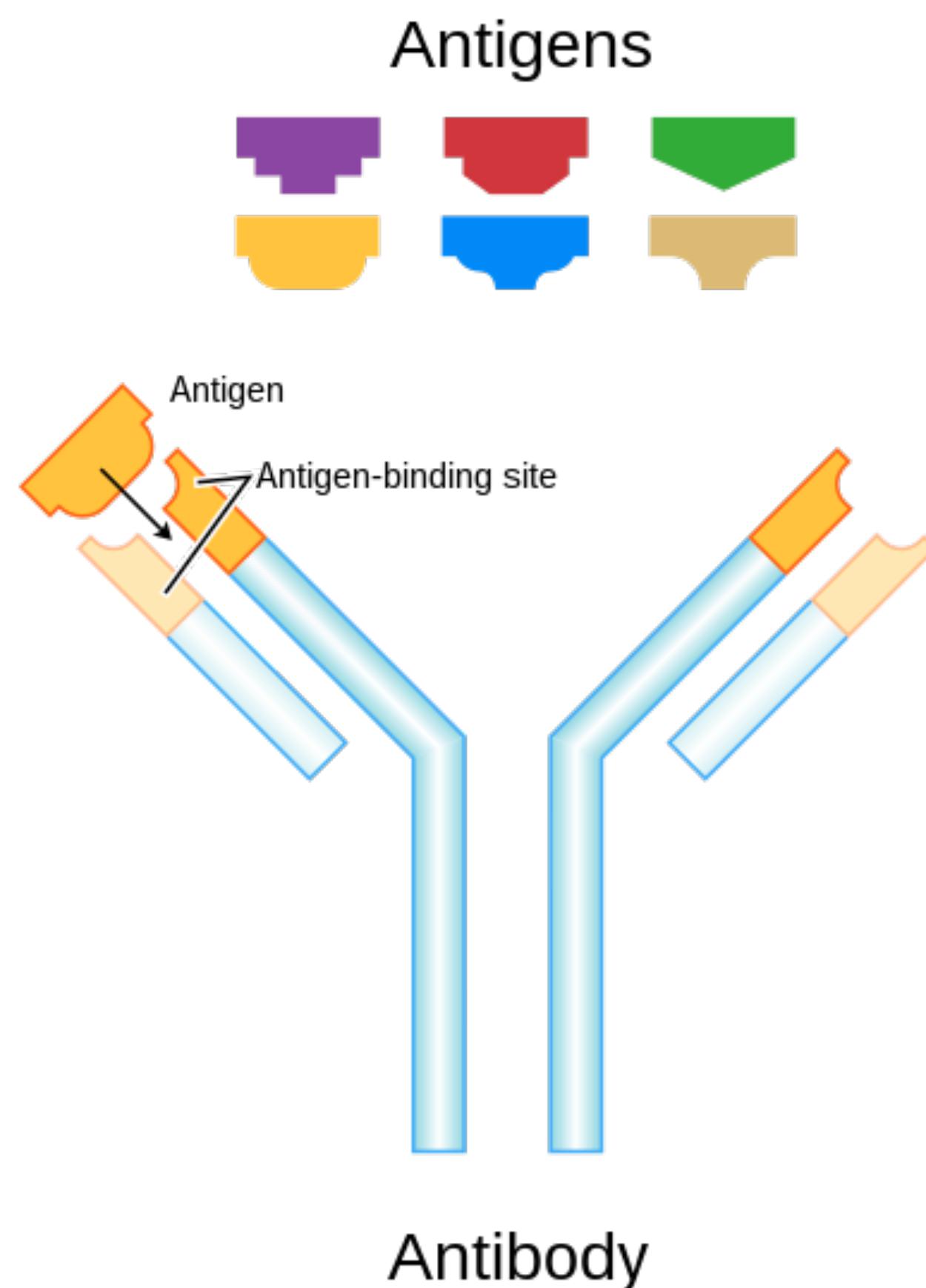
What is an Antibody?

- Antibodies are versatile proteins that bind to pathogens like viruses (Antigens) and stimulate a response.



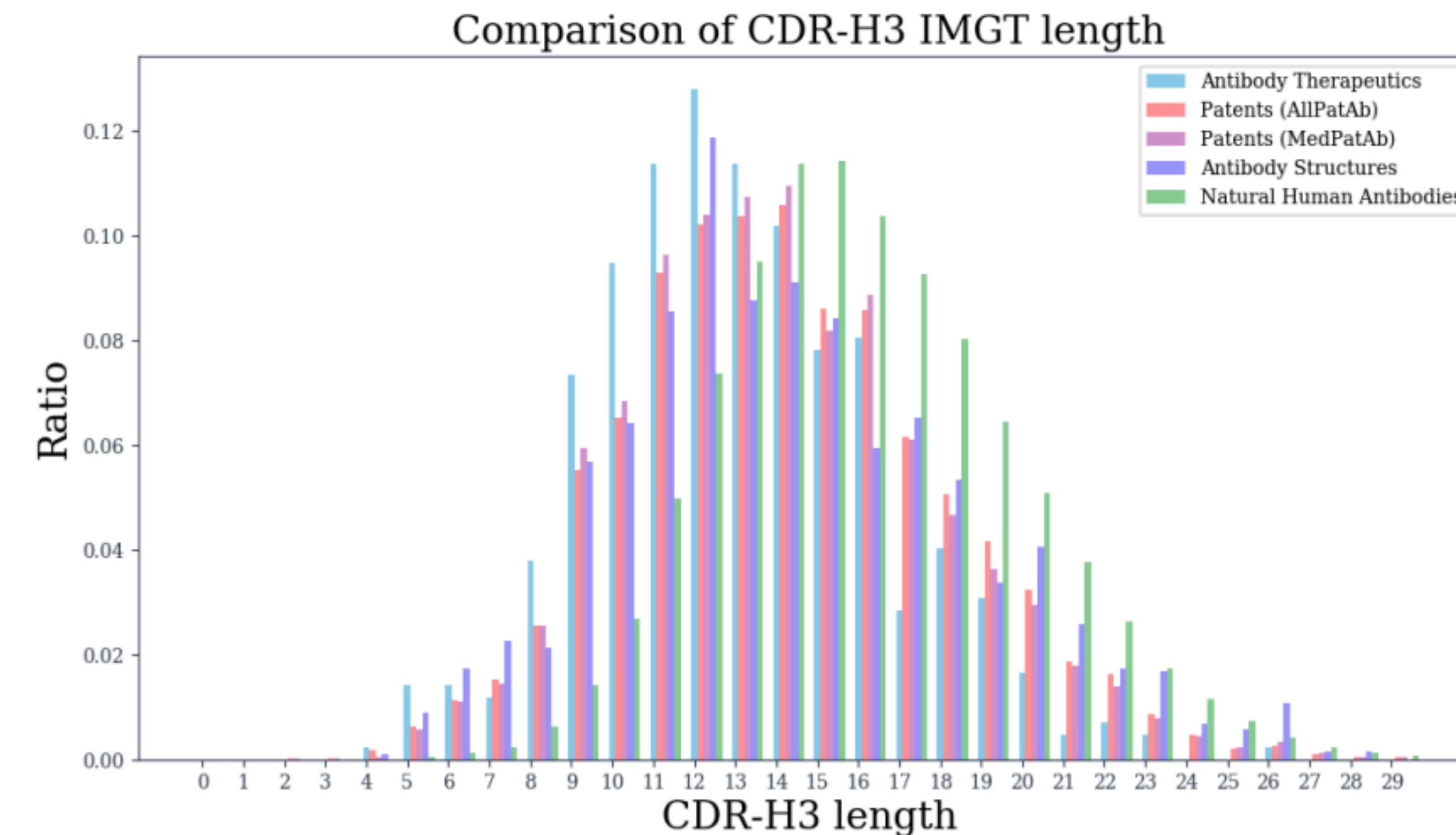
Why to model Antibodies?

- *De novo* generation of new antibodies that target specific *antigens* holds the key to accelerating vaccine discovery

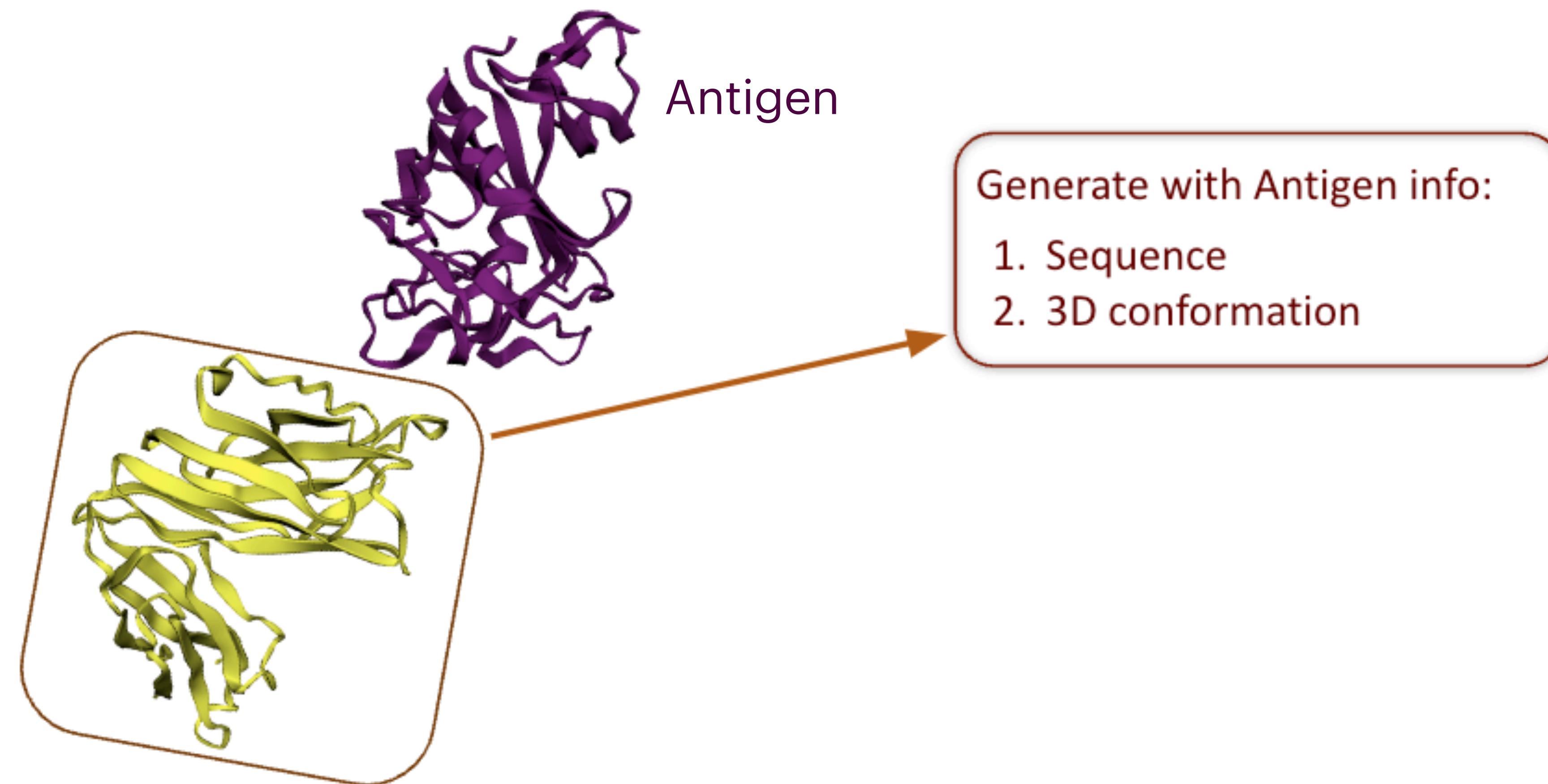


Why to model Antibodies?

- *De novo* generation of new antibodies that target specific *antigens* holds the key to accelerating vaccine discovery
- **Challenge:** Combinatorial search space of over 20^L + Learning $p(\text{Ab}, \text{Ag})$

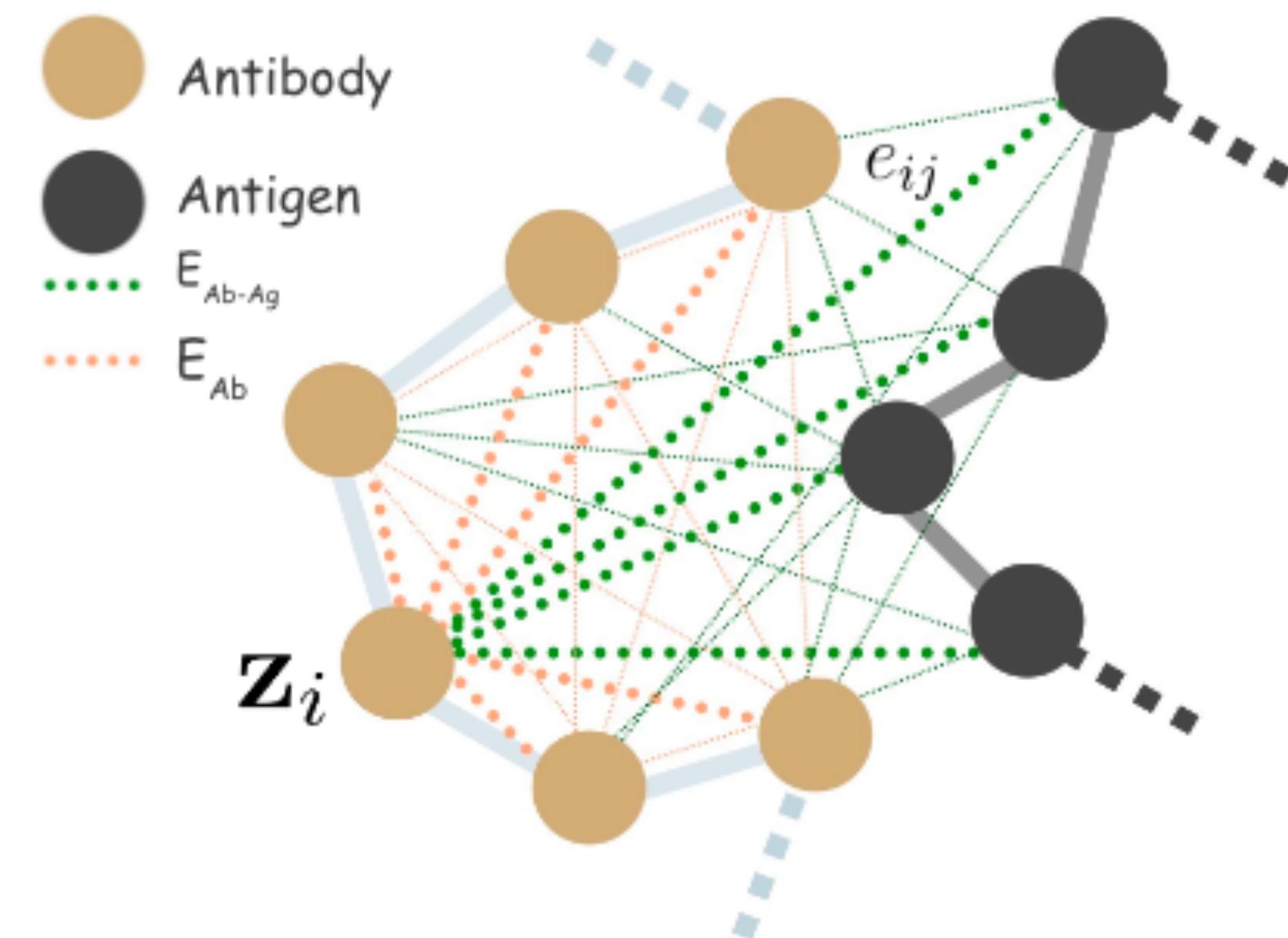


Objective



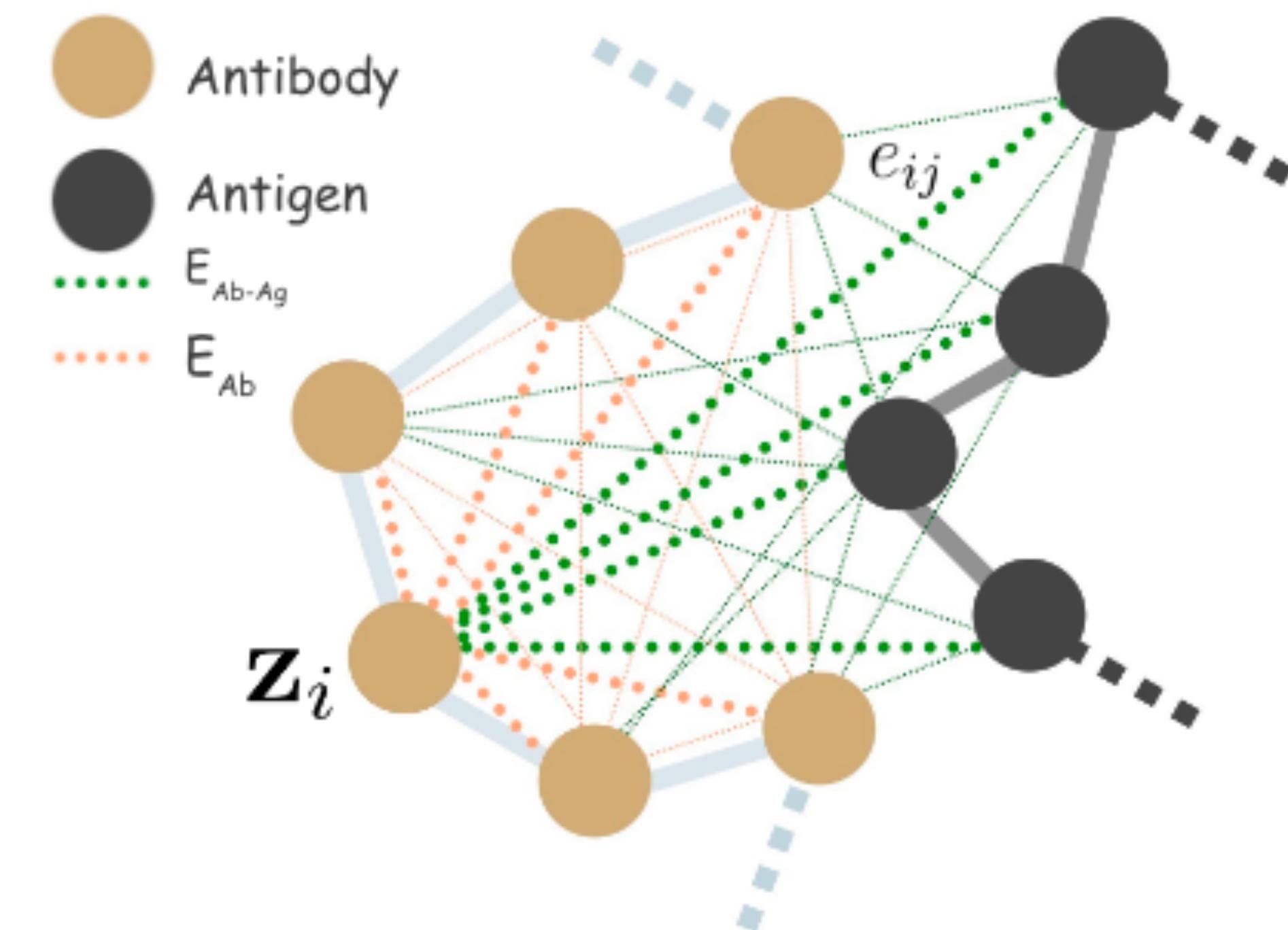
Graph Representation

- We define antigen-antibody complex as a 3D graph $G = (V, E, X)$, where antibody Ab and antigen Ag
 - Vertices $V = (V_{\text{Ab}}, V_{\text{Ag}})$
 - Coordinates $X = (X_{\text{Ab}}, X_{\text{Ag}})$
 - Edges $E = (E_{\text{Ab}}, E_{\text{Ab-Ag}})$



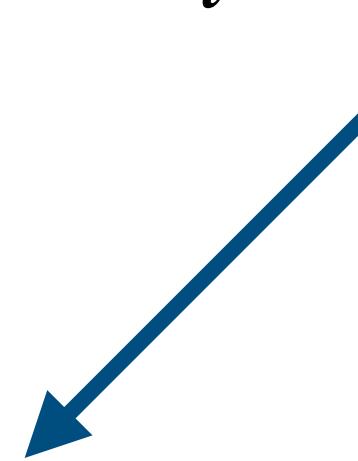
Graph Representation

- We define antigen-antibody complex as a 3D graph $G = (V, E, X)$, where antibody Ab and antigen Ag
 - Vertices $V = (V_{\text{Ab}}, V_{\text{Ag}})$
 - Coordinates $X = (X_{\text{Ab}}, X_{\text{Ag}})$
 - Edges $E = (E_{\text{Ab}}, E_{\text{Ab}-\text{Ag}})$
- Define full connected heterogeneous edges between antibody E_{Ab} residues, and antigens $E_{\text{Ab}-\text{Ag}}$



Graph Representation

$$\mathbf{z}_i = [\mathbf{a}_i, \mathbf{s}_i]$$

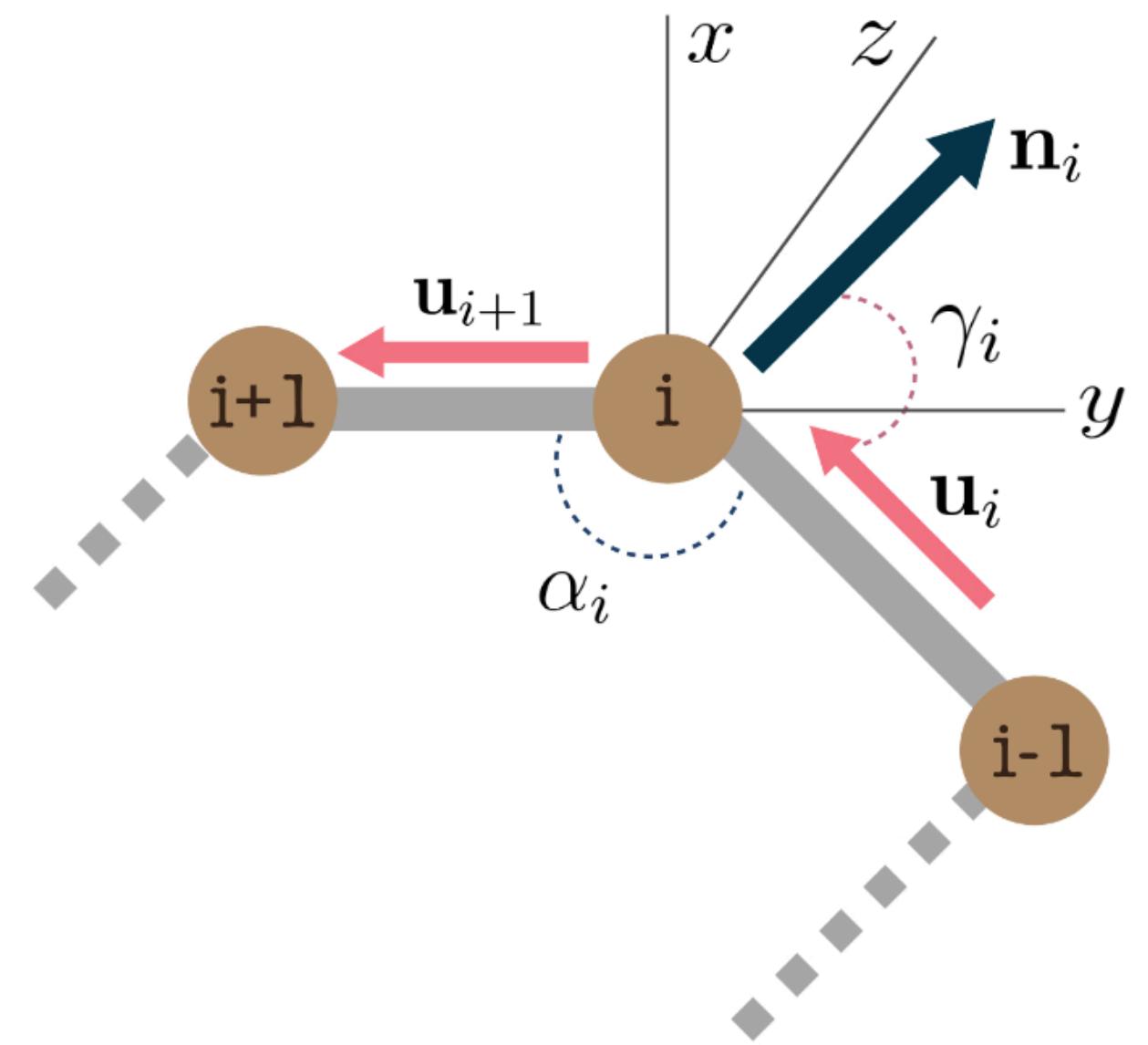


$\mathbf{a}_i \in \mathbb{R}^{20}$: Categorical distribution over
the amino acid labels { Arg, His, ... }

Graph Representation

$$\mathbf{z}_i = [\mathbf{a}_i, \mathbf{s}_i]$$

$\mathbf{a}_i \in \mathbb{R}^{20}$: Categorical distribution over the amino acid labels { Arg, His, ... }



Quaternion type coordinate representation

$$\mathbf{s}_i = (r_i, \alpha_i, \gamma_i)$$

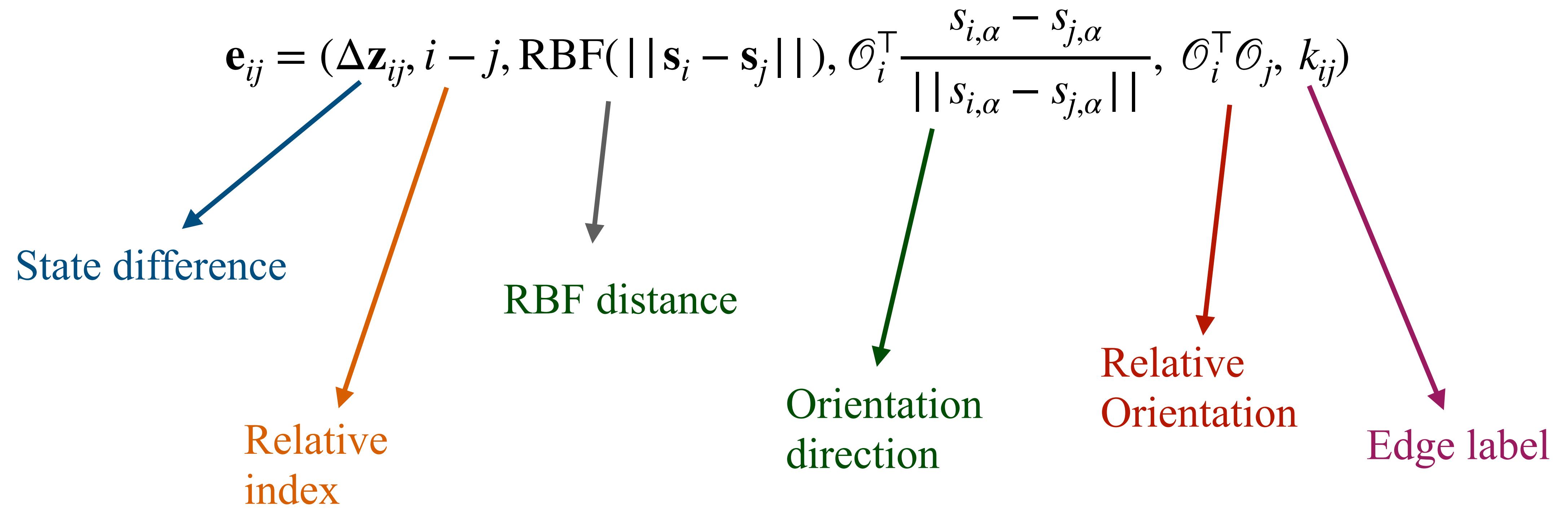
$$r_i = ||\mathbf{u}_i||, \quad \alpha_i = \cos^{-1} \left(\frac{\langle \mathbf{u}_i, \mathbf{u}_{i-1} \rangle}{||\mathbf{u}_i|| \cdot ||\mathbf{u}_{i-1}||} \right)$$

$$\gamma_i = \cos^{-1} \left(\frac{\langle \mathbf{u}_i, \mathbf{n}_i \rangle}{||\mathbf{u}_i|| \cdot ||\mathbf{n}_i||} \right), \quad \mathbf{n}_i = \mathbf{u}_i \times \mathbf{u}_{i-1}$$

$$\mathbf{u}_i = \mathbf{x}_{i+1} - \mathbf{x}_i$$

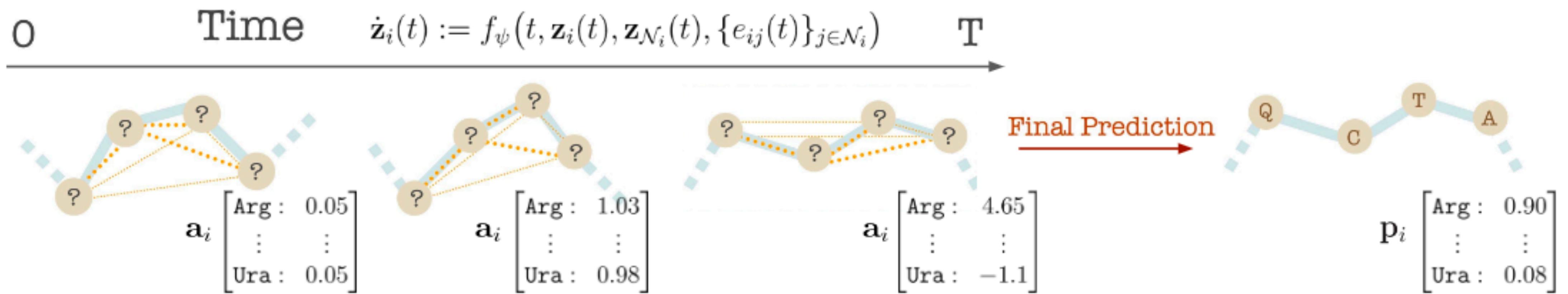
Interactions

- Edge features for higher order information:



Conjoined System of ODEs

- Neural ODEs/PDEs



Conjoined system of ODEs

- We propose to model the distribution of antibody-antigen complexes by a differential graph ODEs $\mathbf{z}(t)$ over time $t \in \mathbb{R}_+$

$$\dot{\mathbf{z}}_i(t) = \frac{\partial \mathbf{z}_i(t)}{\partial t} = f_\psi(t, \mathbf{z}_i(t), \mathbf{z}_{N(i)}(t), \{\mathbf{e}_{ij}(t)\}_j),$$

$$\dot{\mathbf{z}}(t) \triangleq \begin{pmatrix} \dot{\mathbf{z}}_1(t) \\ \vdots \\ \dot{\mathbf{z}}_M(t) \end{pmatrix} = \begin{pmatrix} f_\psi(t, \mathbf{z}_1(t), \mathbf{z}_{N(1)}(t), \{\mathbf{e}_{1j}(t)\}_j) \\ \vdots \\ f_\psi(t, \mathbf{z}_M(t), \mathbf{z}_{N(M)}(t), \{\mathbf{e}_{Mj}(t)\}_j) \end{pmatrix} \longrightarrow \text{Conjoined ODEs}$$

Conjoined system of ODEs

- We propose to model the distribution of antibody-antigen complexes by a differential graph ODEs $\mathbf{z}(t)$ over time $t \in \mathbb{R}_+$

$$\dot{\mathbf{z}}_i(t) = \frac{\partial \mathbf{z}_i(t)}{\partial t} = f_\psi(t, \mathbf{z}_i(t), \mathbf{z}_{N(i)}(t), \{\mathbf{e}_{ij}(t)\}_j),$$

$$\dot{\mathbf{z}}(t) \triangleq \begin{pmatrix} \dot{\mathbf{z}}_1(t) \\ \vdots \\ \dot{\mathbf{z}}_M(t) \end{pmatrix} = \begin{pmatrix} f_\psi(t, \mathbf{z}_1(t), \mathbf{z}_{N(1)}(t), \{\mathbf{e}_{1j}(t)\}_j) \\ \vdots \\ f_\psi(t, \mathbf{z}_M(t), \mathbf{z}_{N(M)}(t), \{\mathbf{e}_{Mj}(t)\}_j) \end{pmatrix} \longrightarrow \text{Conjoined ODEs}$$

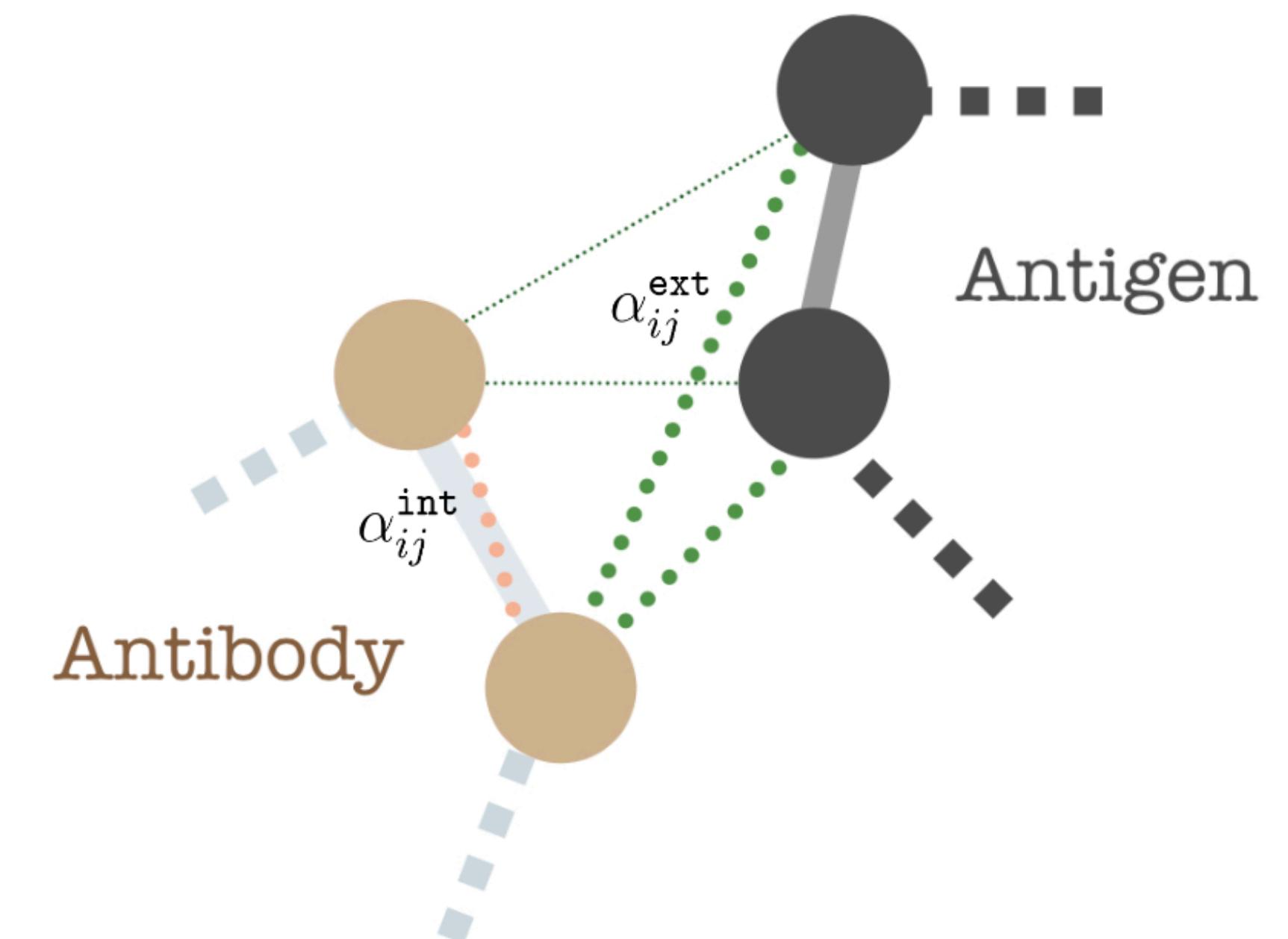
$$\mathbf{z}(T) = \mathbf{z}(0) + \int_0^T \dot{\mathbf{z}}(t) dt \longrightarrow \text{Solving the system}$$

Attention-based differential

- We capture the interactions between the antigen and antibody residues with graph attention with [TransformerConv](#)

$$\alpha_{ij} = \text{softmax} \left(\frac{(\mathbf{W}_3 \mathbf{z}_i)^\top (\mathbf{W}_4 \mathbf{z}_j + \mathbf{W}_6 \mathbf{e}_{ij})}{\sqrt{d}} \right)$$
$$\mathbf{z}'_i = \mathbf{W}_1 \mathbf{z}_i + \sum_{j \in N_{int}(i)} \alpha_{ij}^{\text{int}} (\mathbf{W}_2 \mathbf{z}_j + \mathbf{W}_6 \mathbf{e}_{ij}) + \sum_{j \in N_{ext}(i)} \alpha_{ij}^{\text{ext}} (\mathbf{W}'_2 \mathbf{z}_j + \mathbf{W}'_6 \mathbf{e}_{ij})$$

Internal Contribution External contribution



Training Objective

- We optimise for the data fit of the generated states with loss consists of two components: one for the sequence and another for the structure

$$\mathcal{L} = \mathcal{L}_{\text{seq}} + \mathcal{L}_{\text{structure}}$$

Training Objective

- We optimise for the data fit of the generated states with loss consists of two components: one for the sequence and another for the structure

$$\mathcal{L} = \mathcal{L}_{\text{seq}} + \mathcal{L}_{\text{structure}}$$

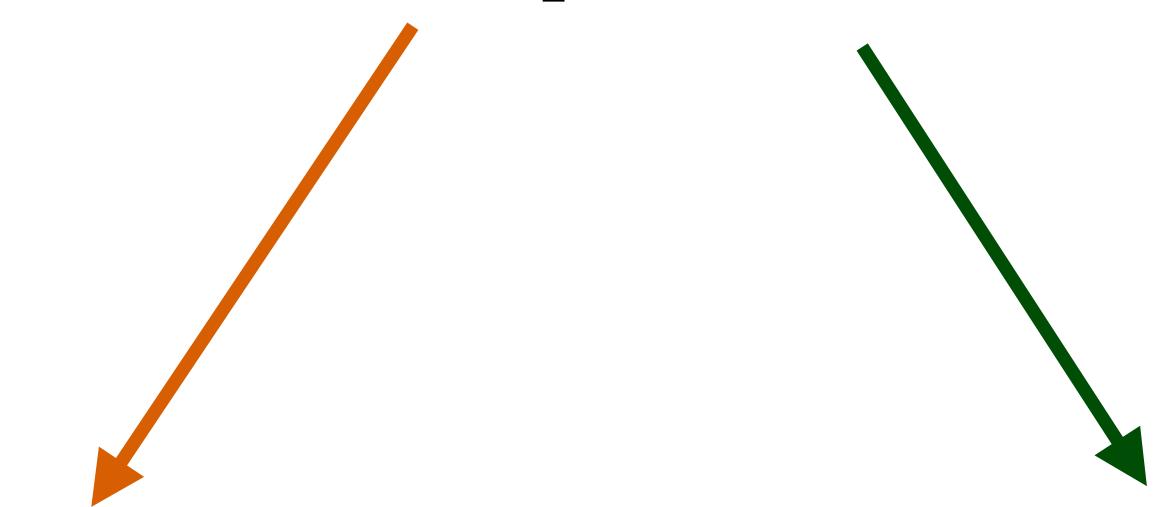


$$\mathcal{L}_{\text{seq}} = \frac{1}{N} \sum_{n=1}^N \frac{1}{M} \sum_{i=1}^{M_i} \text{CE}(\mathbf{a}_{ni}^{\text{true}}, \mathbf{a}_{ni})$$

Training Objective

- We optimise for the data fit of the generated states with loss consists of two components: one for the sequence and another for the structure

$$\mathcal{L} = \mathcal{L}_{\text{seq}} + \mathcal{L}_{\text{structure}}$$



$$\mathcal{L}_{\text{seq}} = \frac{1}{N} \sum_{n=1}^N \frac{1}{M} \sum_{i=1}^{M_i} \text{CE}(\mathbf{a}_{ni}^{\text{true}}, \mathbf{a}_{ni})$$

$$\mathcal{L}_{\text{structure}} = -\frac{1}{N} \sum_{n=1}^N \frac{1}{M} \sum_{i=1}^{M_i} \lambda (\mathcal{L}_{\text{angle}}^{ni} + \mathcal{L}_{\text{radius}}^{ni})$$

Training Objective

- We optimise for the data fit of the generated states with loss consists of two components: one for the sequence and another for the structure, \mathcal{M} : von-mises distribution, \mathcal{N} : normal distribution

$$\mathcal{L} = \mathcal{L}_{\text{seq}} + \mathcal{L}_{\text{structure}}$$

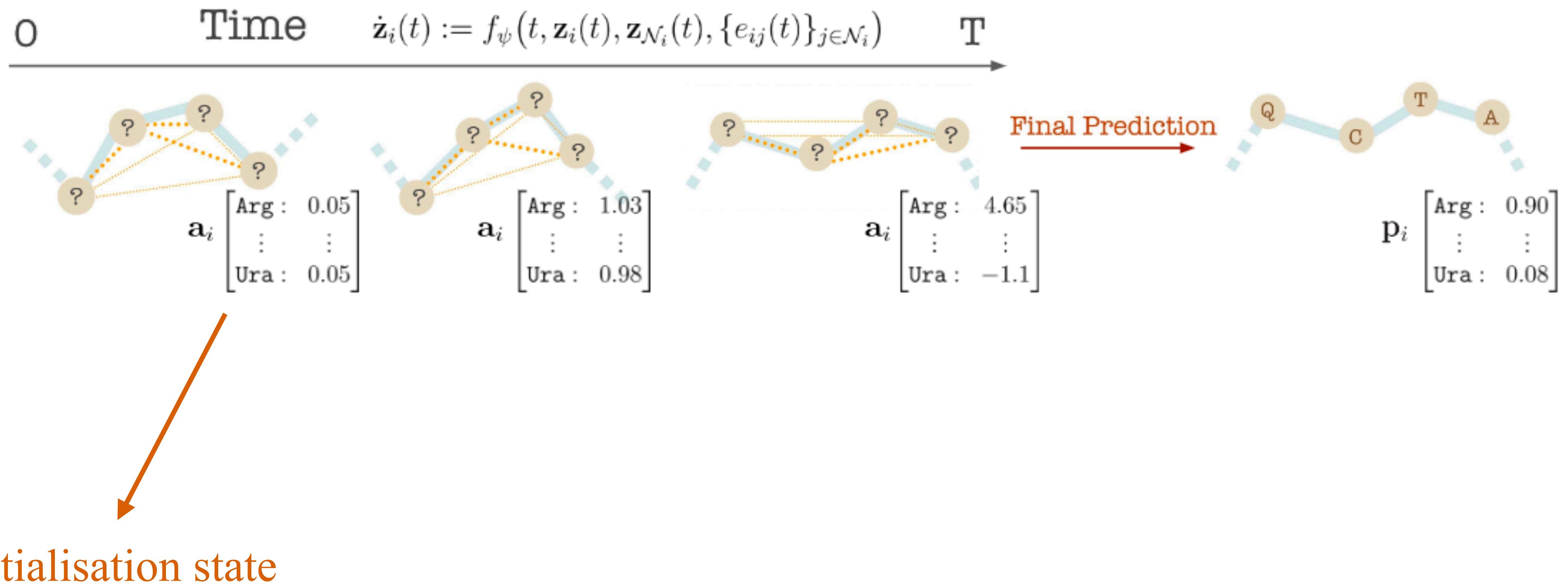
$$\mathcal{L}_{\text{seq}} = \frac{1}{N} \sum_{n=1}^N \frac{1}{M} \sum_{i=1}^{M_i} \text{CE}(\mathbf{a}_{ni}^{\text{true}}, \mathbf{a}_{ni})$$

$$\mathcal{L}_{\text{structure}} = -\frac{1}{N} \sum_{n=1}^N \frac{1}{M} \sum_{i=1}^{M_i} \lambda (\mathcal{L}_{\text{angle}}^{ni} + \mathcal{L}_{\text{radius}}^{ni})$$

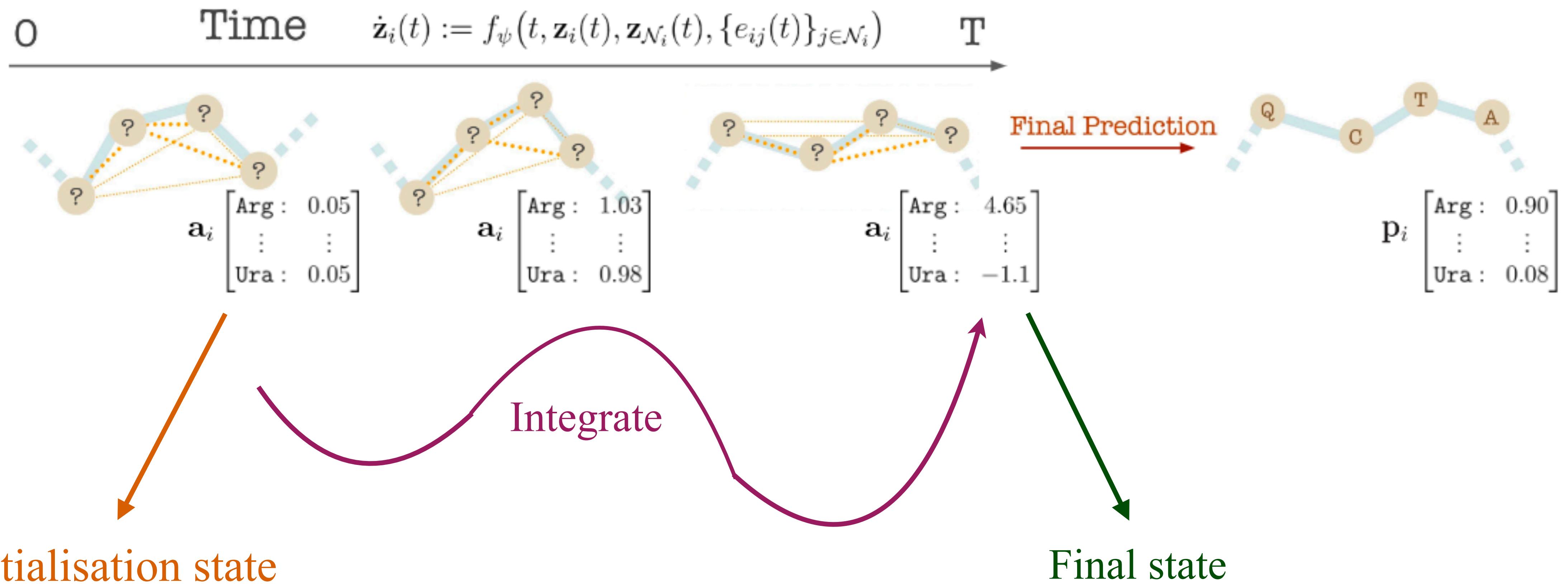
$$\mathcal{L}_{\text{angle}}^{ni} = \sum_k^{\{C_\alpha, C_\gamma, N\}} \sum_{\theta \in \{\alpha, \gamma\}} \log \mathcal{M}(\theta_{ik}^n | \theta_{ik}^{n,\text{true}}, \kappa)$$

$$\mathcal{L}_{\text{radius}}^{ni} = \sum_k^{\{C_\alpha, C_\gamma, N\}} \log \mathcal{N}(r_{ik}^n | r_{ik}^{n,\text{true}}, \sigma_r^2)$$

Sequence and Structure Prediction



Sequence and Structure Prediction



Experiments

- Uncontrolled Antibody sequence and structure generation
 - Generate antibody sequence and structure without any external conditioning
 - PPL : Perplexity (Exponential of negative log likelihood)
 - RMSD: Root Mean Square Deviation by Kabsch Algorithm on C_α spatial features

Experiments

- Uncontrolled Antibody sequence and structure generation
 - Generate antibody sequence and structure without any external conditioning
 - PPL : Perplexity (Exponential of negative log likelihood)
 - RMSD: Root Mean Square Deviation by Kabsch Algorithm on C_α spatial features

Model	CDR-H1		CDR-H2		CDR-H3	
	PPL (↓)	RMSD (↓)	PPL (↓)	RMSD (↓)	PPL (↓)	RMSD (↓)
LSTM	6.79	(N/A)	7.21	(N/A)	9.70	(N/A)
AR-GNN	6.44	2.97	6.86	2.27	9.44	3.63
RefineGNN	6.09	1.18	6.58	0.87	8.38	2.50
AbODE	4.25 ±0.46	0.73 ±0.15	4.32 ±0.31	0.63 ±0.19	6.35 ±0.29	2.01 ±0.13

Experiments

- Antigen conditioned Antibody sequence and structure generation
 - Generate antibody sequence and structure with the antigen conditioning
 - **AAR:** Amino Acid Recovery rate, defined as the overlapping rate between the predicted 1D sequences and the ground truth

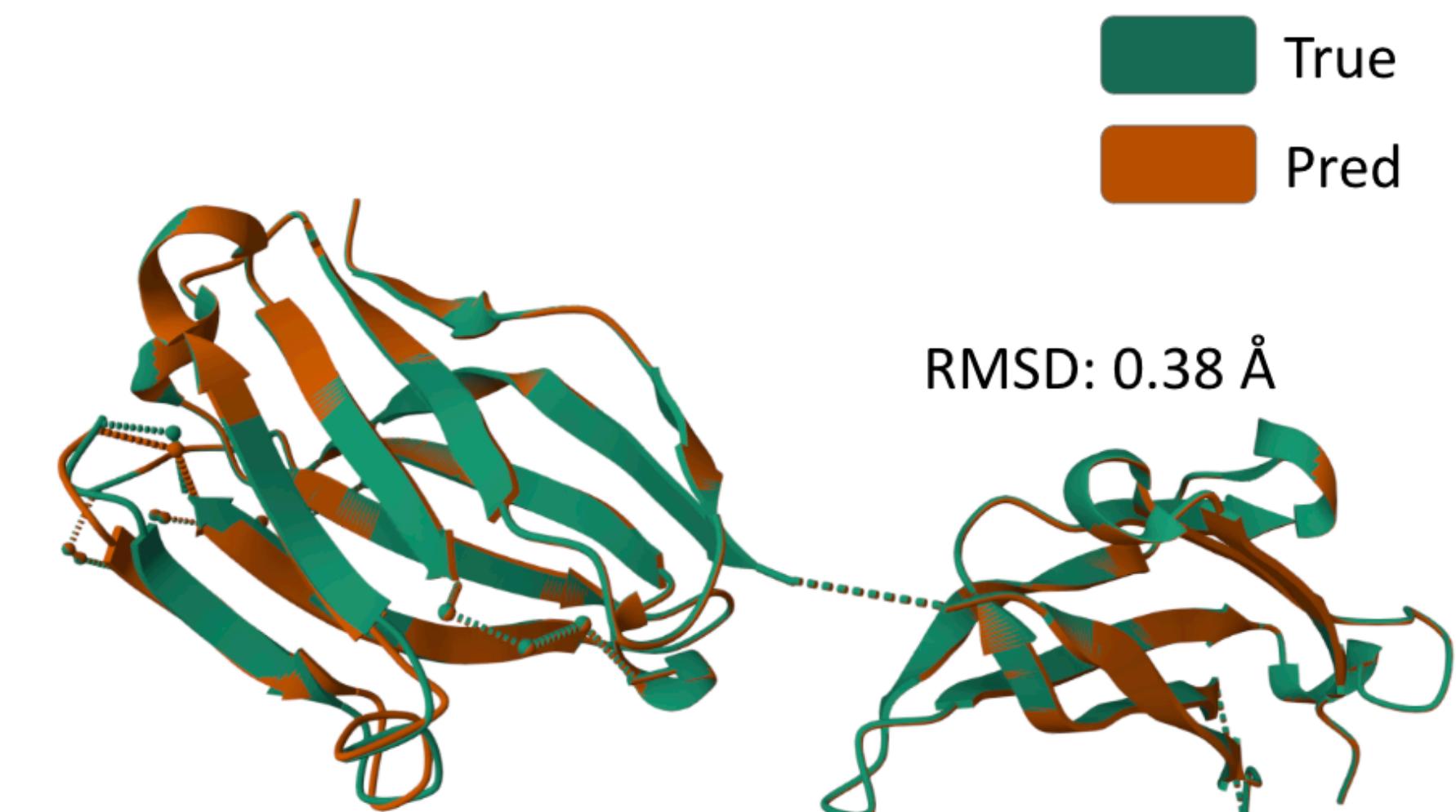
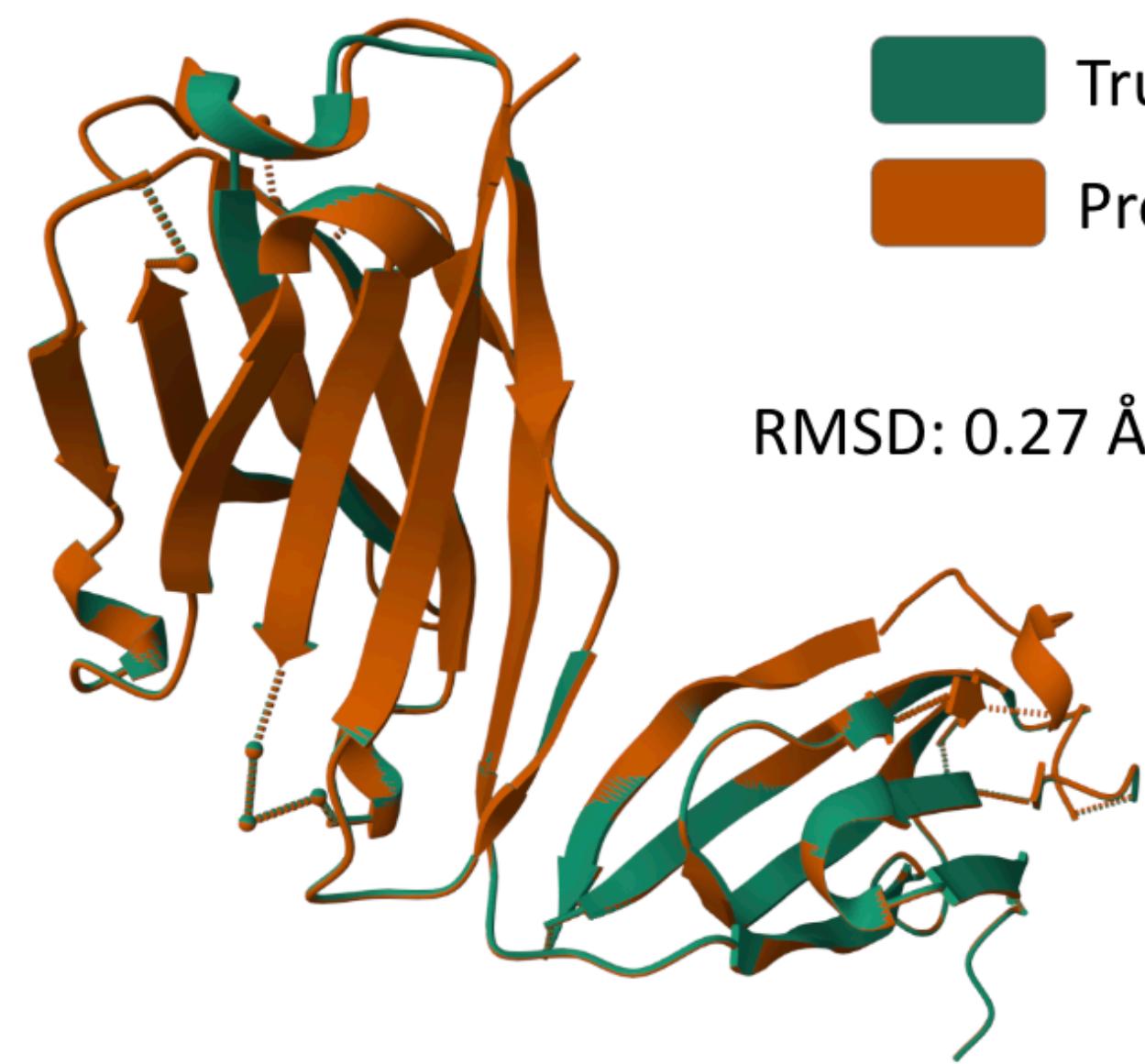
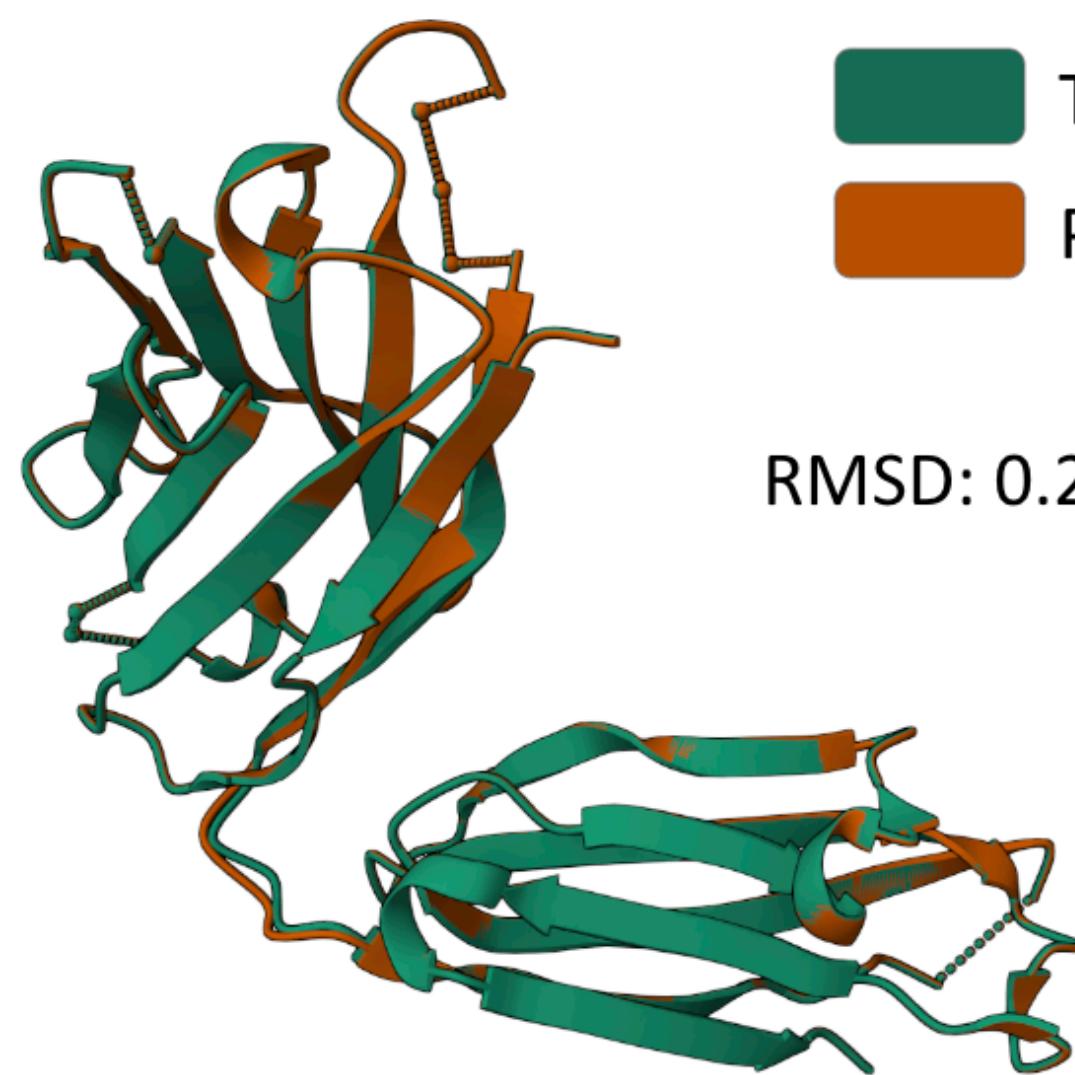
Experiments

- Antigen conditioned Antibody sequence and structure generation
 - Generate antibody sequence and structure with the antigen conditioning
 - **AAR:** Amino Acid Recovery rate, defined as the overlapping rate between the predicted 1D sequences and the ground truth

Model	CDR-H1		CDR-H2		CDR-H3	
	AAR % (\uparrow)	RMSD (\downarrow)	AAR % (\uparrow)	RMSD (\downarrow)	AAR % (\uparrow)	RMSD (\downarrow)
LSTM	40.98 \pm 5.20	(N/A)	28.50 \pm 1.55	(N/A)	15.69 \pm 0.91	(N/A)
C-LSTM	40.93 \pm 5.41	(N/A)	29.24 \pm 1.08	(N/A)	15.48 \pm 1.17	(N/A)
RefineGNN	39.40 \pm 5.56	3.22 \pm 0.29	37.06 \pm 3.09	3.64 \pm 0.40	21.13 \pm 1.59	6.00 \pm 0.55
C-RefineGNN	33.19 \pm 2.99	3.25 \pm 0.40	33.53 \pm 3.23	3.69 \pm 0.56	18.88 \pm 1.37	6.22 \pm 0.59
MEAN	58.29 \pm 7.27	0.98 \pm 0.16	47.15 \pm 3.09	0.95 \pm 0.05	36.38 \pm 3.08	2.21 \pm 0.16
AbODE	70.5 \pm 1.14	0.65 \pm 0.1	55.7 \pm 1.45	0.73 \pm 0.14	39.8 \pm 1.17	1.73 \pm 0.11

Experiments

- Antigen conditioned Antibody sequence and structure generation



Experiments

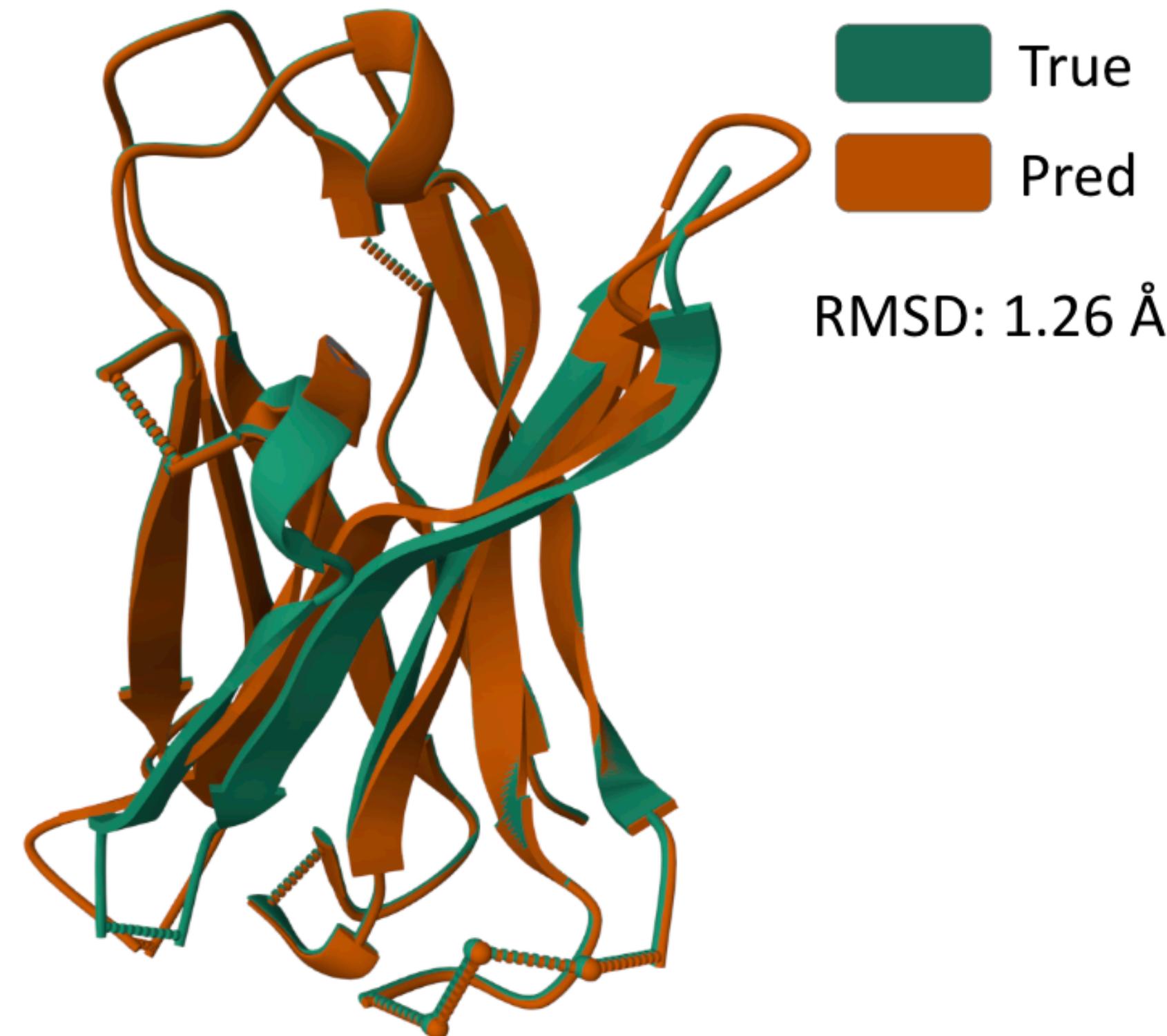
- Antigen-binding CDR-H3 Design
 - Design CDR-H3 that binds to a given antigen, evaluated on 60 diverse complexes selected by RabD

Experiments

- Antigen-binding CDR-H3 Design

- Design CDR-H3 that binds to a given antigen, evaluated on 60 diverse complexes selected by RabD

Model	AAR % (\uparrow)	RMSD (\downarrow)
RosettaAD	22.50	5.52
LSTM	22.36	(N/A)
C-LSTM	22.18	(N/A)
RefineGNN	29.79	7.55
C-RefineGNN	28.90	7.21
MEAN	36.77	1.81
AbODE	39.95 \pm 1.3	1.54 \pm 0.24



Conclusion

- We propose **AbODE**, which models the antibody-antigen complex as a joint graph, and via a system of coupled residue-specific ODEs
- **AbODE** is able to incorporate conditional contextual and spatial information in ODEs tailored for Antibody design

References

- Kabsch, Wolfgang. "A solution for the best rotation to relate two sets of vectors." *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography* 32.5 (1976): 922-923.
- Jin, Wengong, et al. "Iterative refinement graph neural network for antibody sequence-structure co-design." *arXiv preprint arXiv:2110.04624* (2021).
- Adolf-Bryfogle, Jared, et al. "RosettaAntibodyDesign (RAbD): A general framework for computational antibody design." *PLoS computational biology* 14.4 (2018): e1006112.
- Dunbar, James, et al. "SAbDab: the structural antibody database." *Nucleic acids research* 42.D1 (2014): D1140-D1146.
- Graves, Jordan, et al. "A review of deep learning methods for antibodies." *Antibodies* 9.2 (2020): 12
- Chen, Ricky TQ, et al. "Neural ordinary differential equations." *Advances in neural information processing systems* 31 (2018).
- Fischman, Sharon, and Yanay Ofran. "Computational design of antibodies." *Current opinion in structural biology* 51 (2018): 156-162.
- Ingraham, John, et al. "Generative models for graph-based protein design." *Advances in neural information processing systems* 32 (2019).
- Jing, Bowen, et al. "Learning from protein structure with geometric vector perceptrons." *arXiv preprint arXiv:2009.01411* (2020).
- Kong, Xiangzhe, Wenbing Huang, and Yang Liu. "Conditional antibody design as 3d equivariant graph translation." *arXiv preprint arXiv:2208.06073* (2022)