Antibody-Antigen Docking and Design via Hierarchical Equivariant Refinement
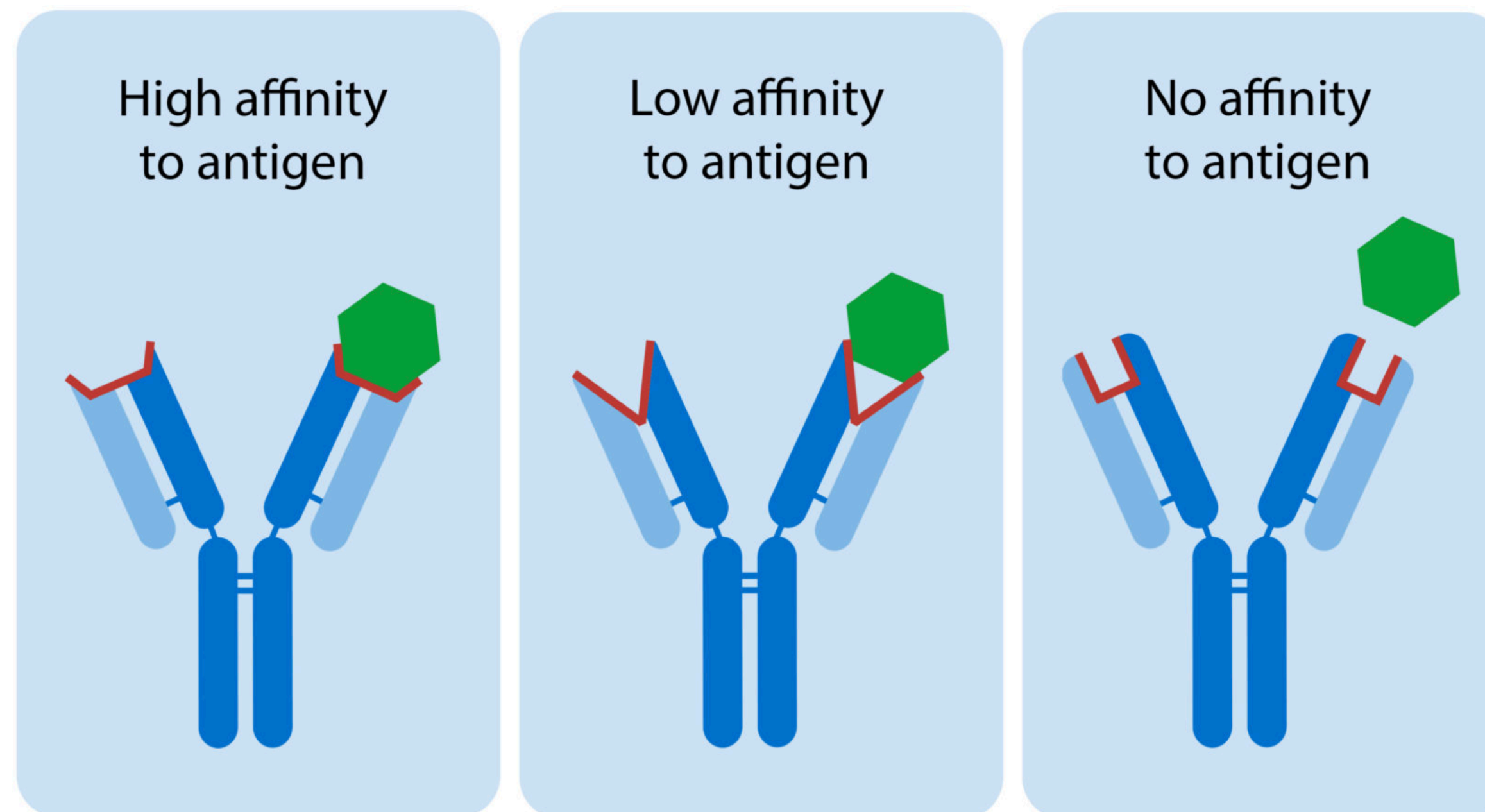
Wengong Jin (Eric and Wendy Schmidt Center, Broad Institute)

Regina Barzilay, Tommi Jaakkola (MIT CSAIL)

Antibody-Antigen Binding

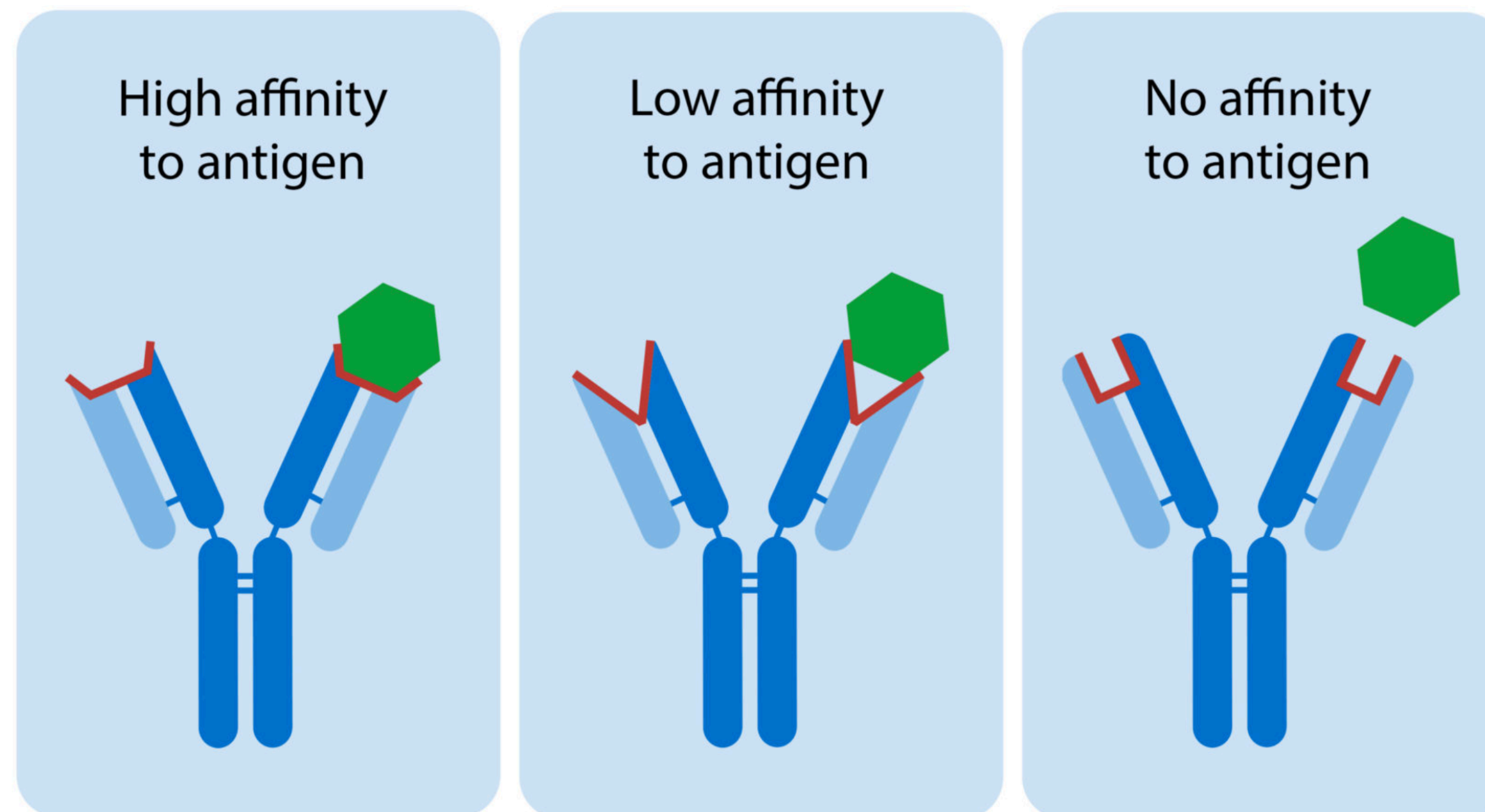## A fundamental problem in immunology and therapeutic design

- Antibodies bind to antigens (e.g., viral proteins) and trigger immune response

Antibody-Antigen Binding

A fundamental problem in immunology and therapeutic design

- Antibodies bind to antigens (e.g., viral proteins) and trigger immune response

- The binding affinity is governed by the 3D binding interface where antibody residues (paratope) closely interact with antigen residues (epitope).

Antibody-Antigen Binding
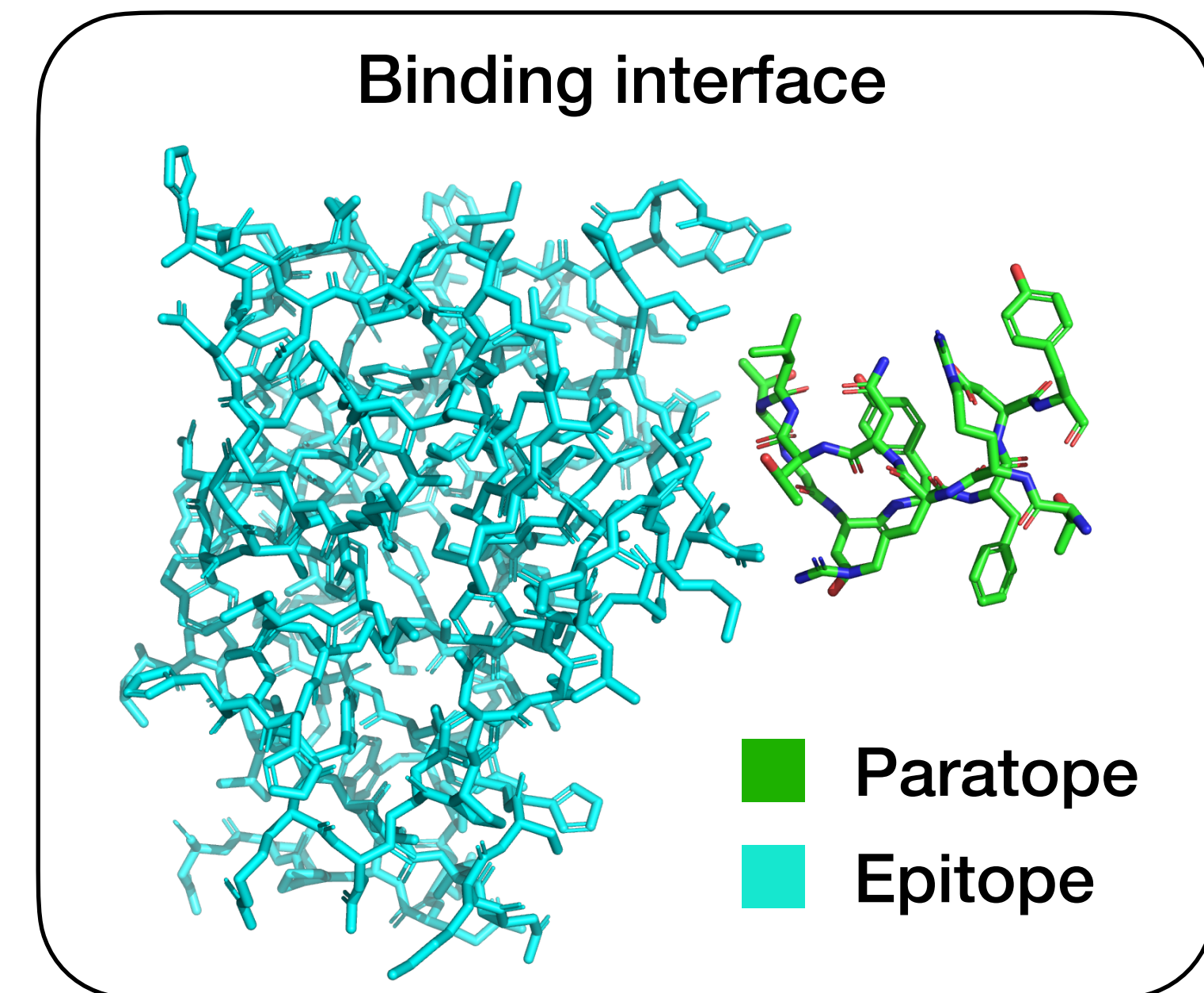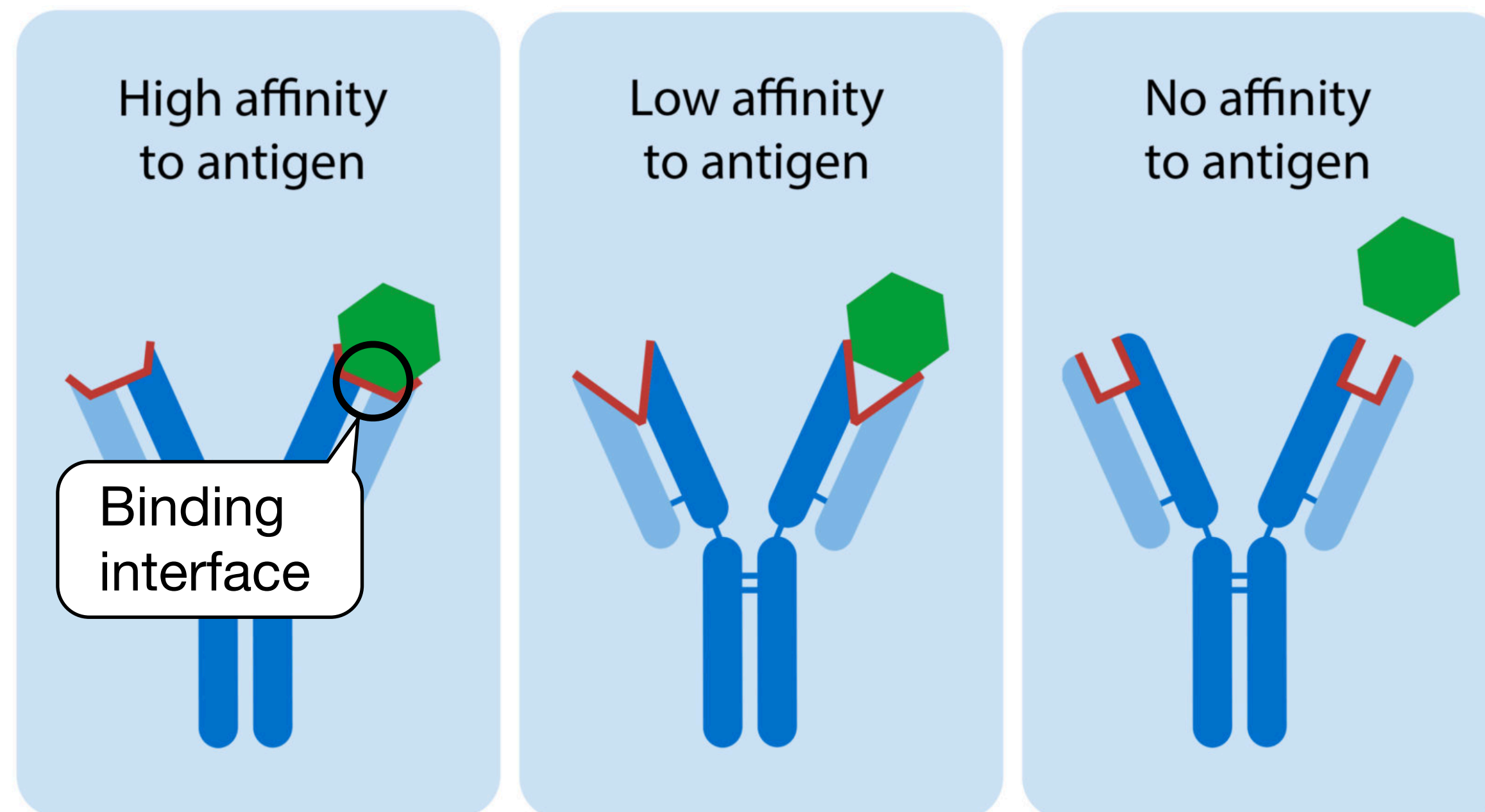## A fundamental problem in immunology and therapeutic design

- Antibodies bind to antigens (e.g., viral proteins) and trigger immune response

- The binding affinity is governed by the 3D binding interface where antibody residues (paratope) closely interact with antigen residues (epitope).

High affinity
to antigen

Binding
interface

Low affinity
to antigen

No affinity
to antigen

Binding interface

Paratope

Epitope

2

Antibody-Antigen Binding
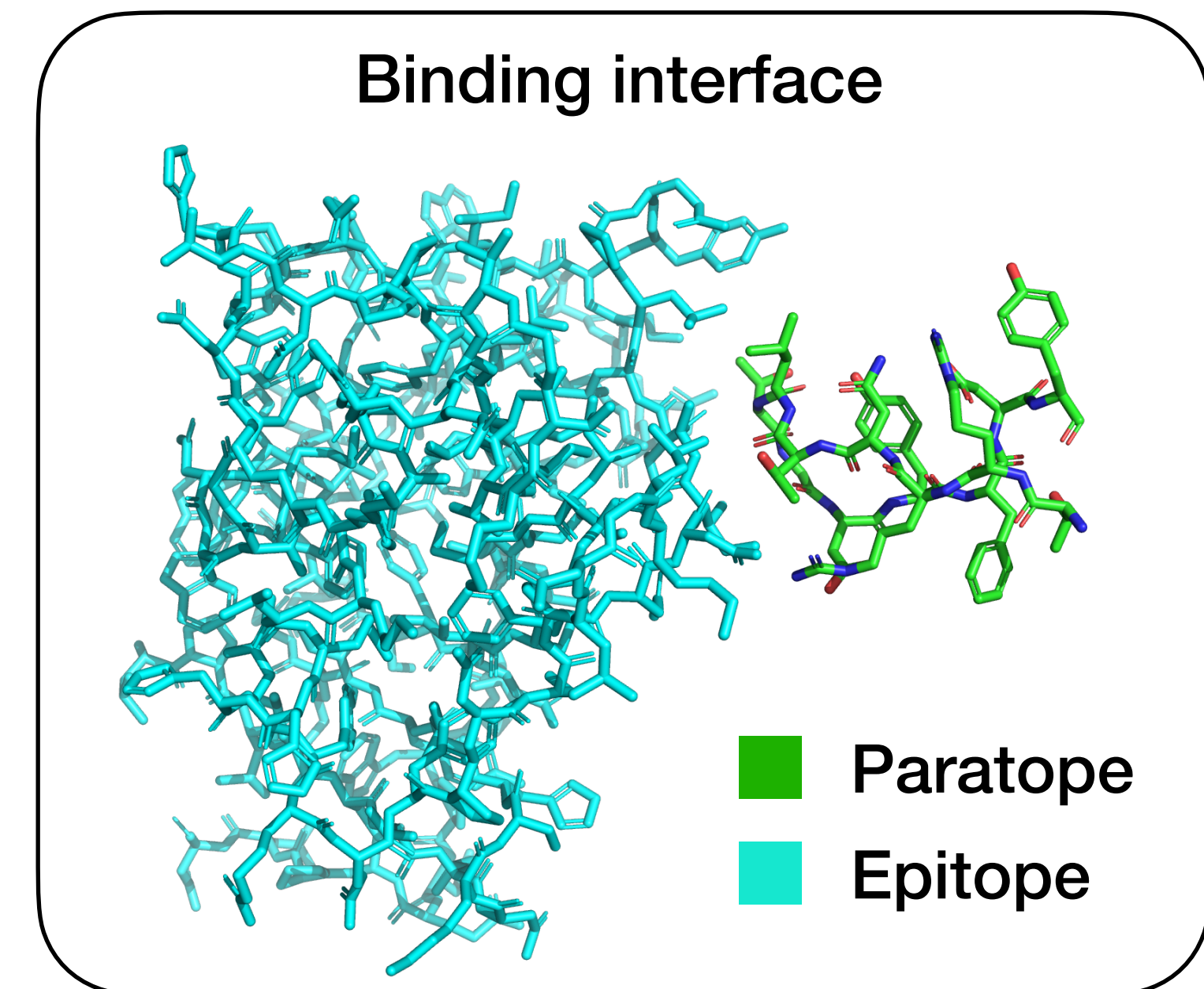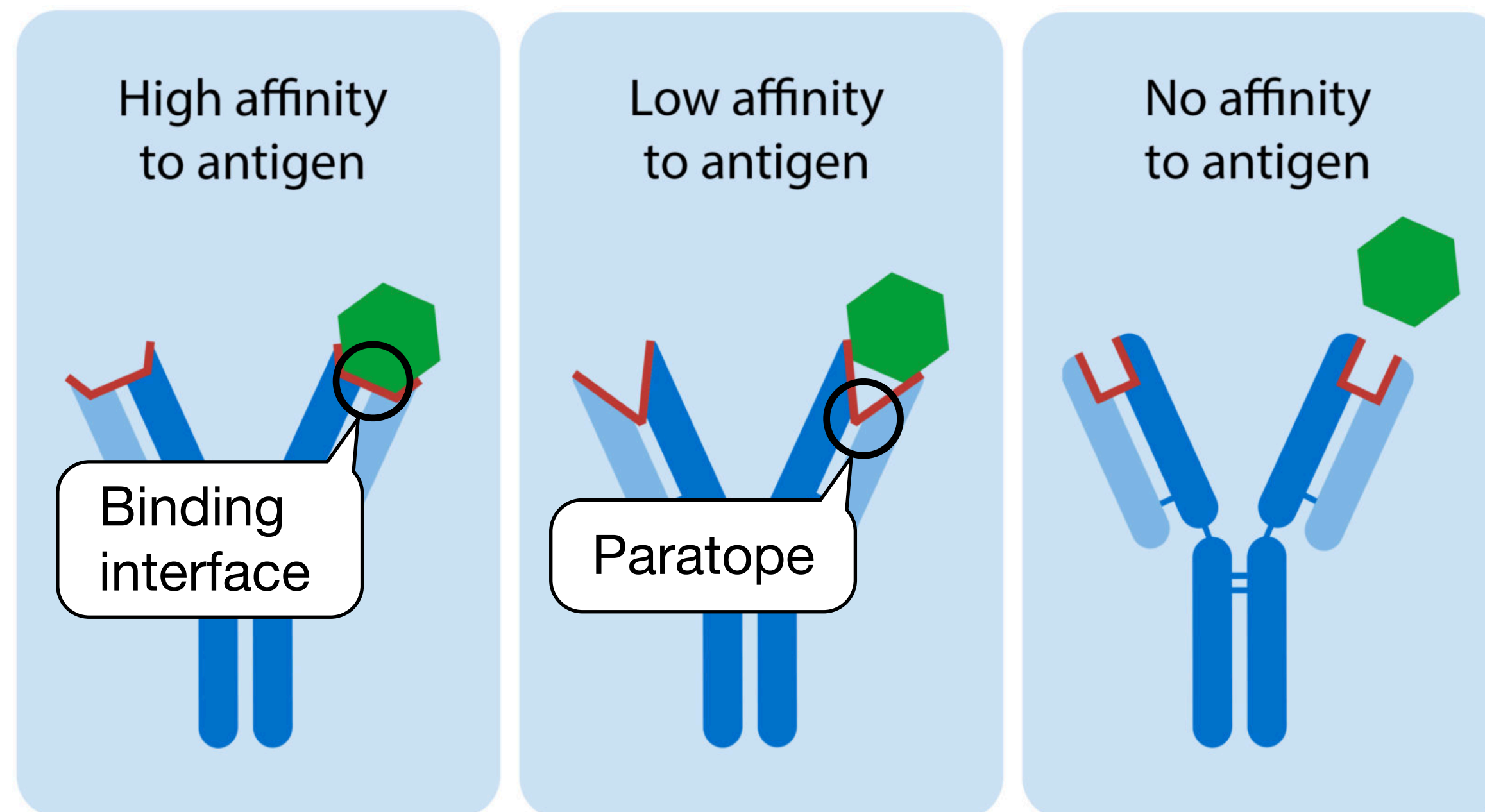## A fundamental problem in immunology and therapeutic design

- Antibodies bind to antigens (e.g., viral proteins) and trigger immune response

- The binding affinity is governed by the 3D binding interface where antibody residues (paratope) closely interact with antigen residues (epitope).

Antibody-Antigen Binding

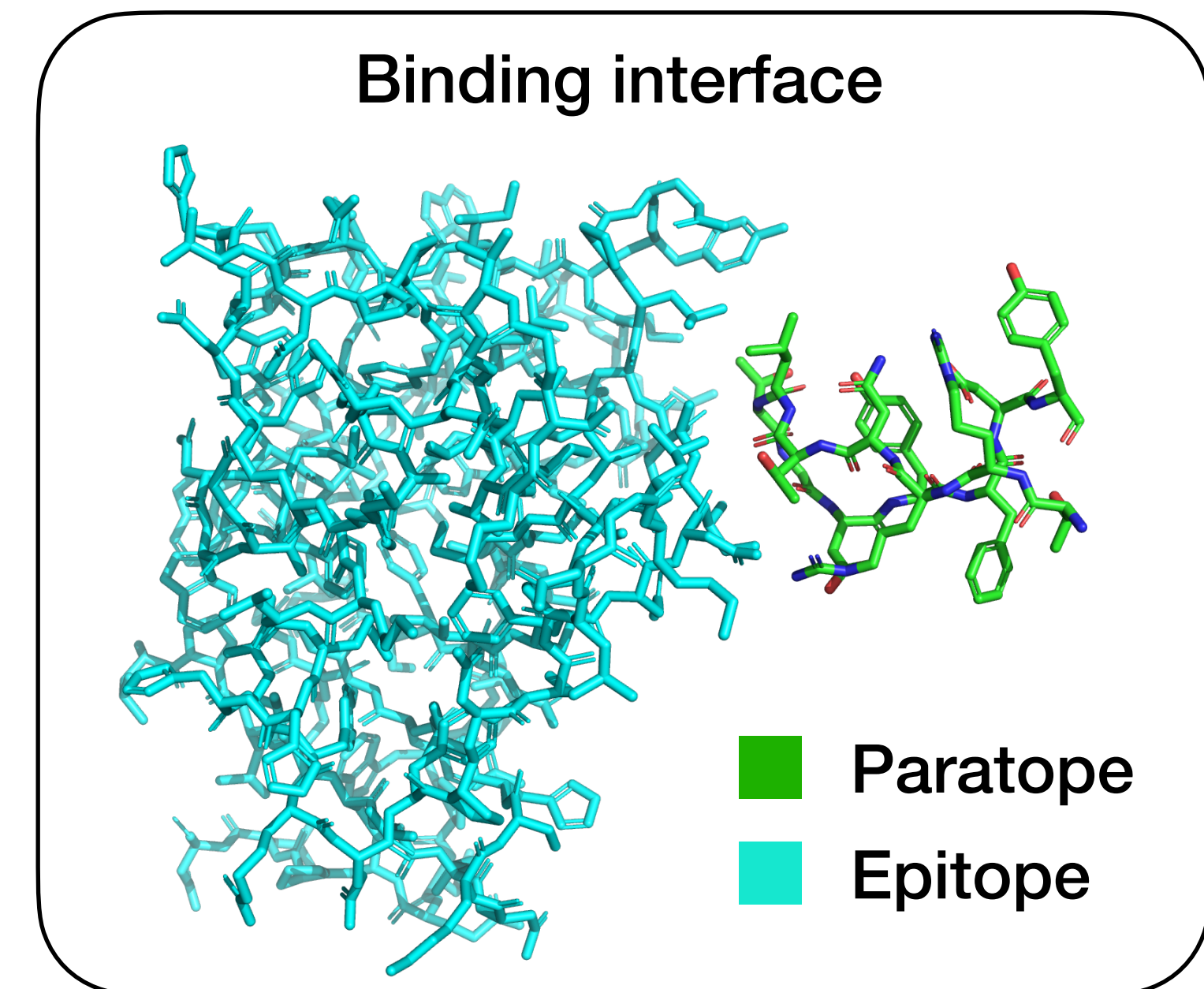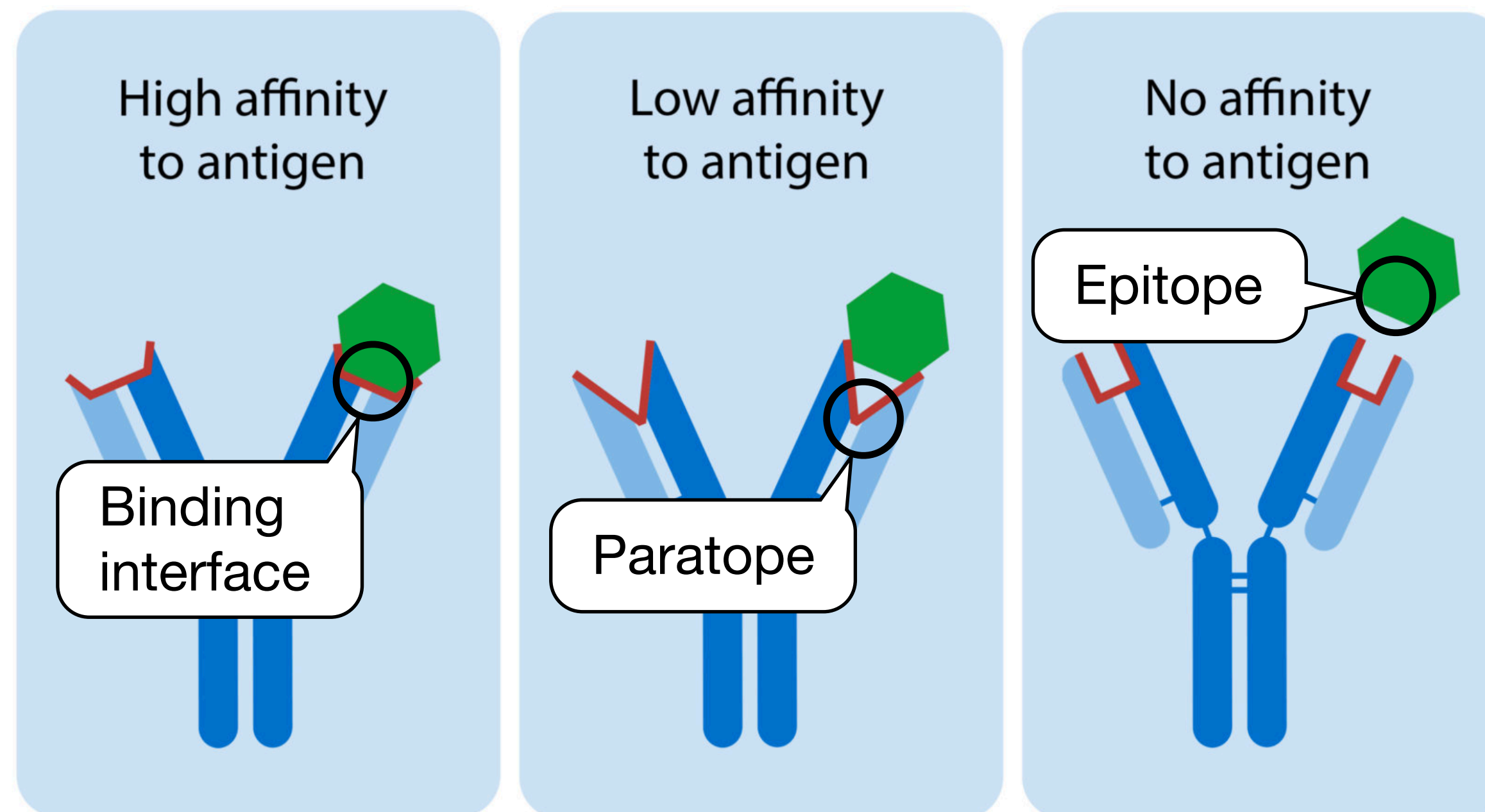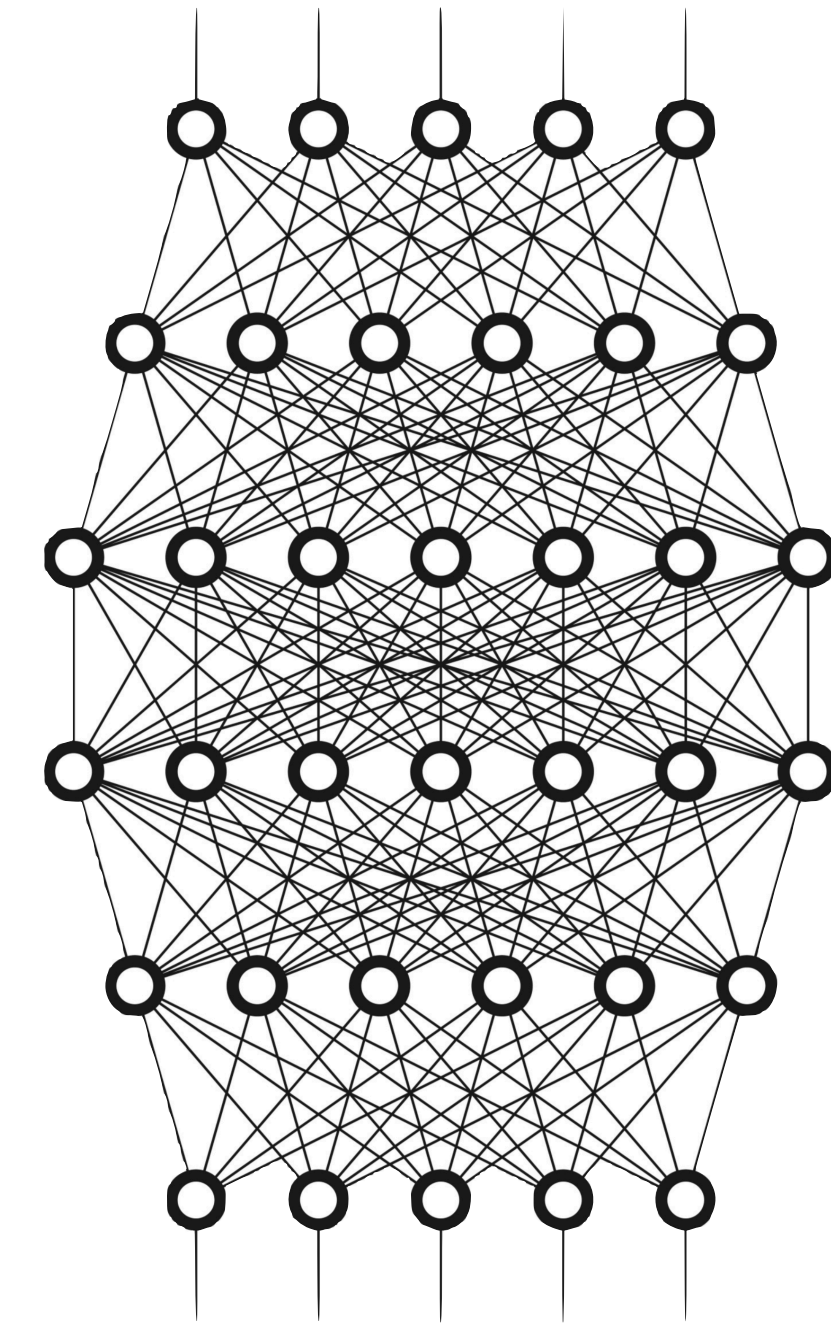## A fundamental problem in immunology and therapeutic design

- Antibodies bind to antigens (e.g., viral proteins) and trigger immune response

- The binding affinity is governed by the 3D binding interface where antibody residues (paratope) closely interact with antigen residues (epitope).

Antibody-Antigen Binding

A tale of two problems: docking and generation

Binding affinity

Antibodies

Antibody-Antigen Binding
A tale of two problems: docking and generation

- **Docking**: given an antibody paratope sequence and epitope 3D structure, can we predict whether/how they bind?

  - An accurately docked paratope allows us to easily calculate its binding affinity

Binding affinity

Dock

Antibodies

Antibody-Antigen Binding
A tale of two problems: docking and generation

- **Docking**: given an antibody paratope sequence and epitope 3D structure, can we predict whether/how they bind?

  - An accurately docked paratope allows us to easily calculate its binding affinity

- **Generation**: can we generate a paratope with high binding affinity to a given epitope?

  - This inverse problem is difficult due to a combinatorial search space of 20^N

Binding affinity
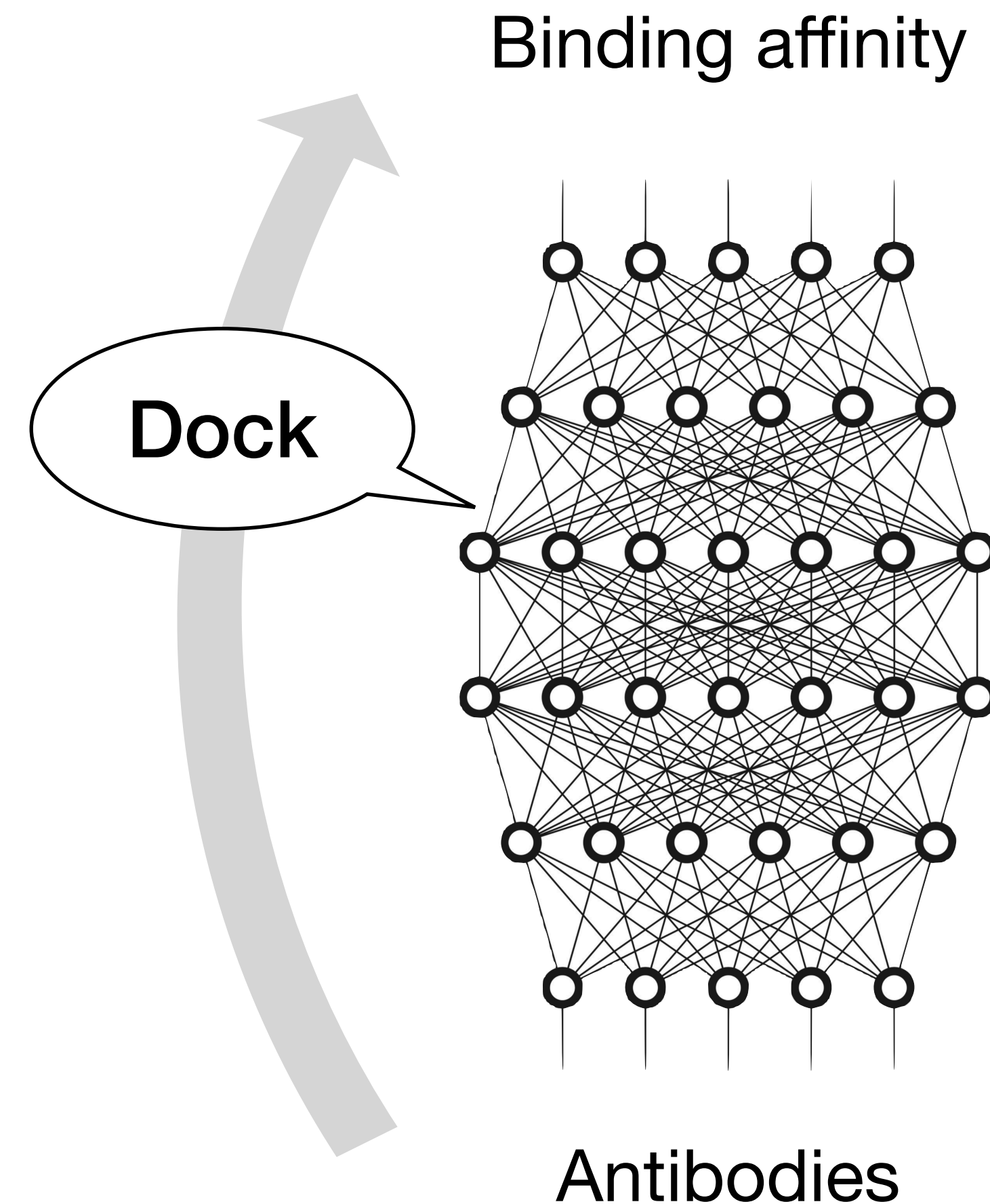
Dock

Design

Antibodies

Antibody-Antigen Binding
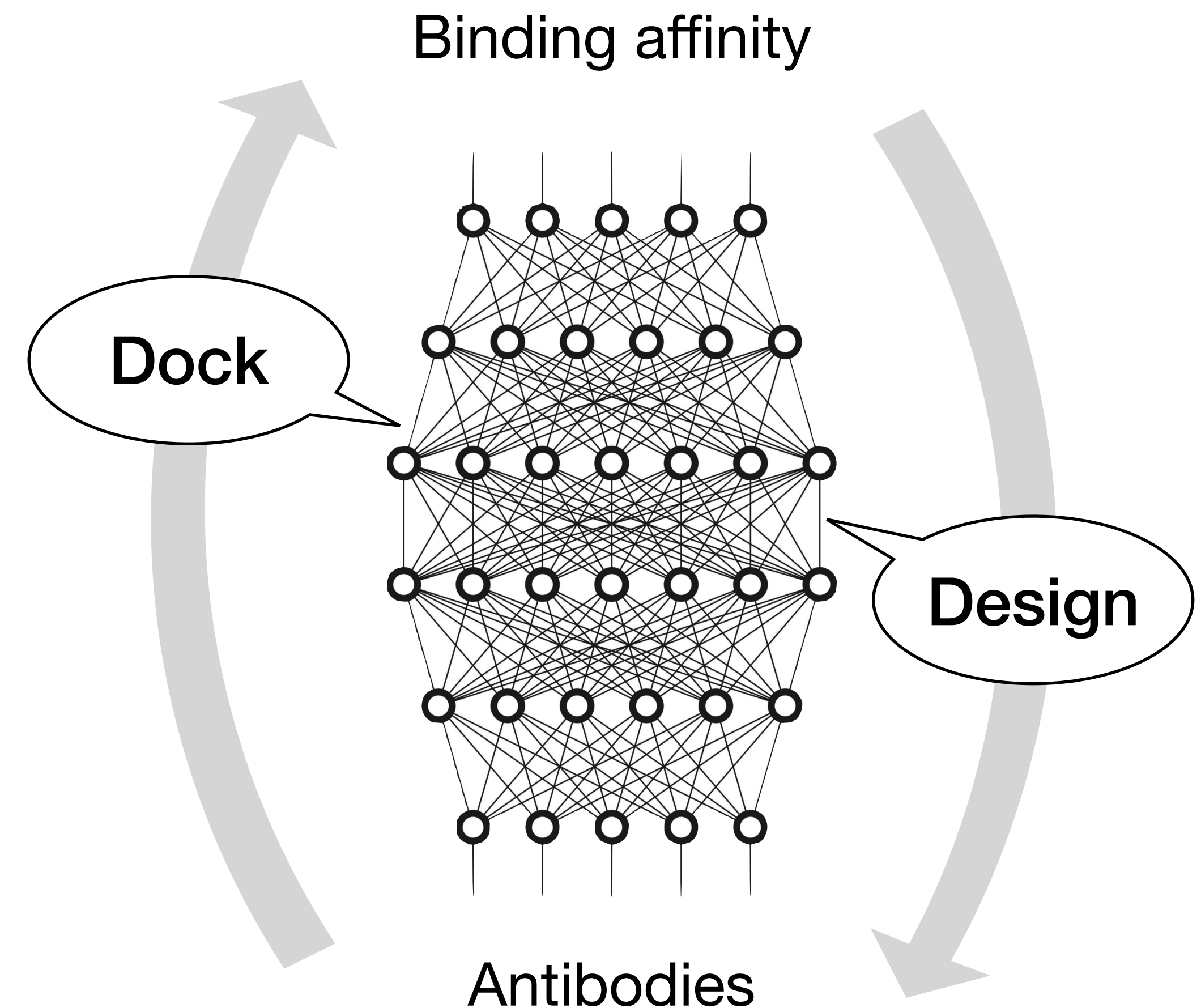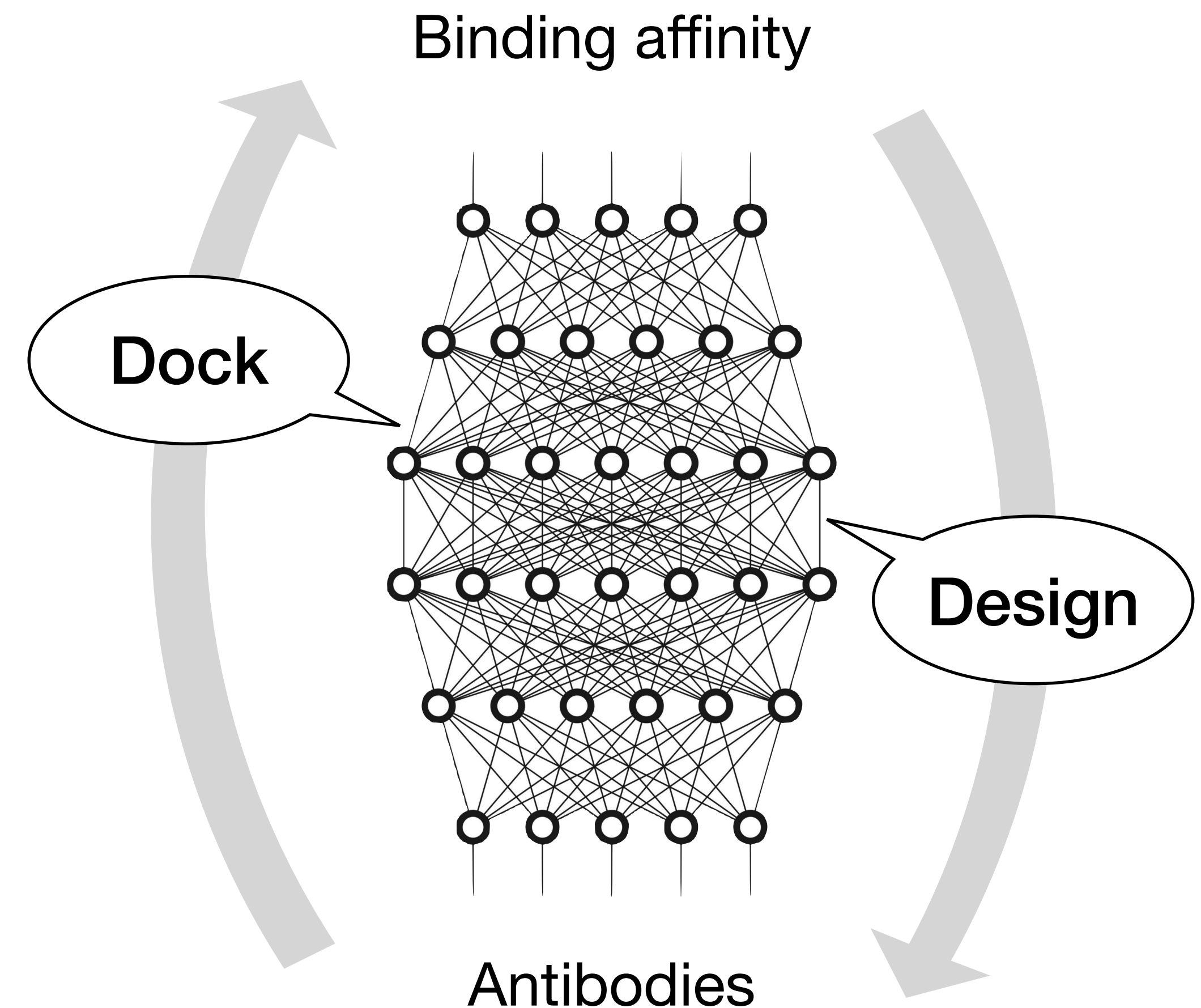## A tale of two problems: docking and generation

- **Docking**: given an antibody paratope sequence and epitope 3D structure, can we predict whether/how they bind?

  - An accurately docked paratope allows us to easily calculate its binding affinity

- **Generation**: can we generate a paratope with high binding affinity to a given epitope?

  - This inverse problem is difficult due to a combinatorial search space of 20^N

- Goal: design a unified model for both tasks

Binding affinity

Dock

Design

Antibodies

Antibody-Antigen Binding
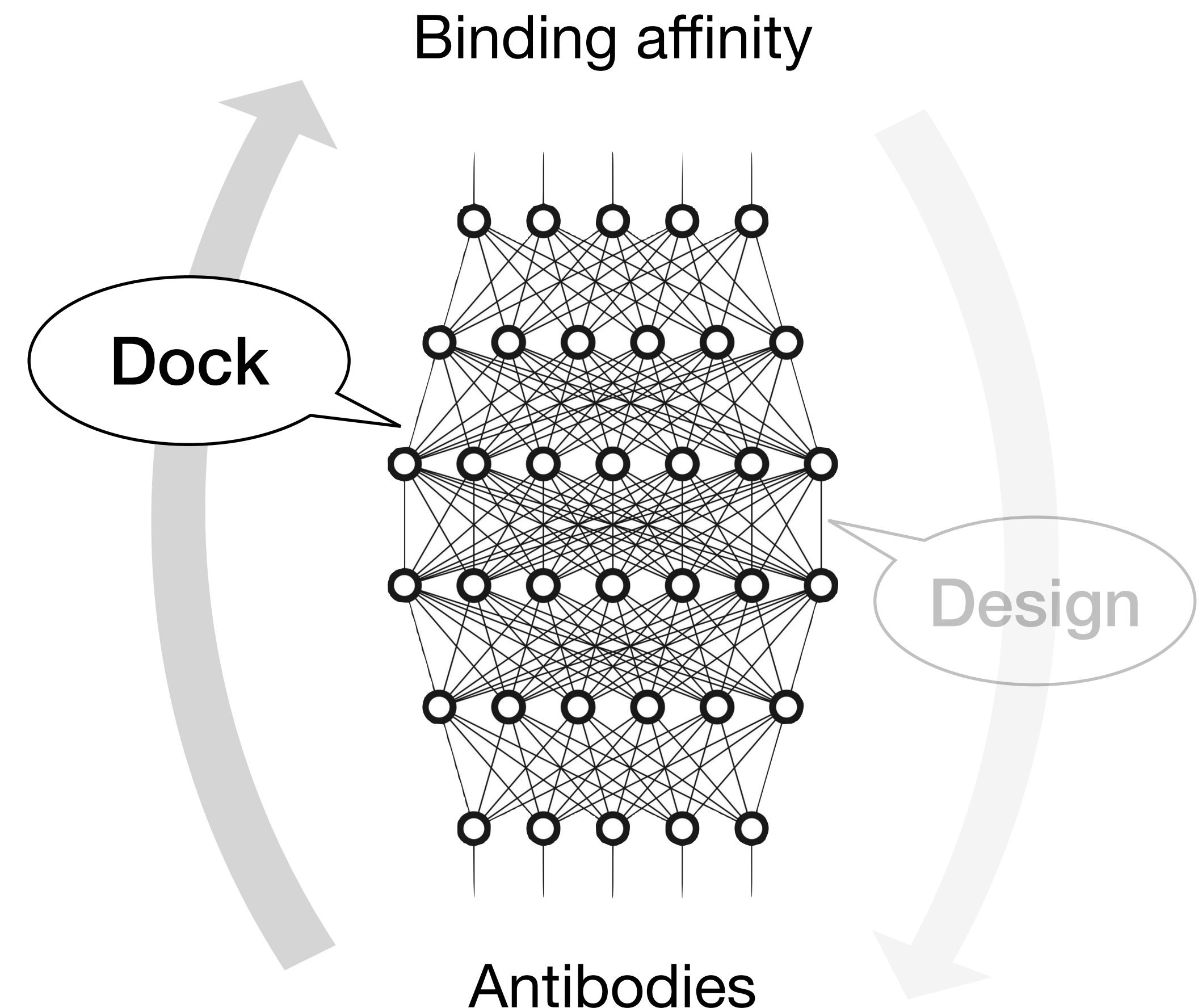## A tale of two problems: docking and generation

- **Docking**: given an antibody paratope sequence and epitope 3D structure, can we predict whether/how they bind?

  - An accurately docked paratope allows us to easily calculate its binding affinity

- **Generation**: can we generate a paratope with high binding affinity to a given epitope?

  - This inverse problem is difficult due to a combinatorial search space of 20^N

- Goal: design a unified model for both tasks

Binding affinity

Dock

Design

Antibodies

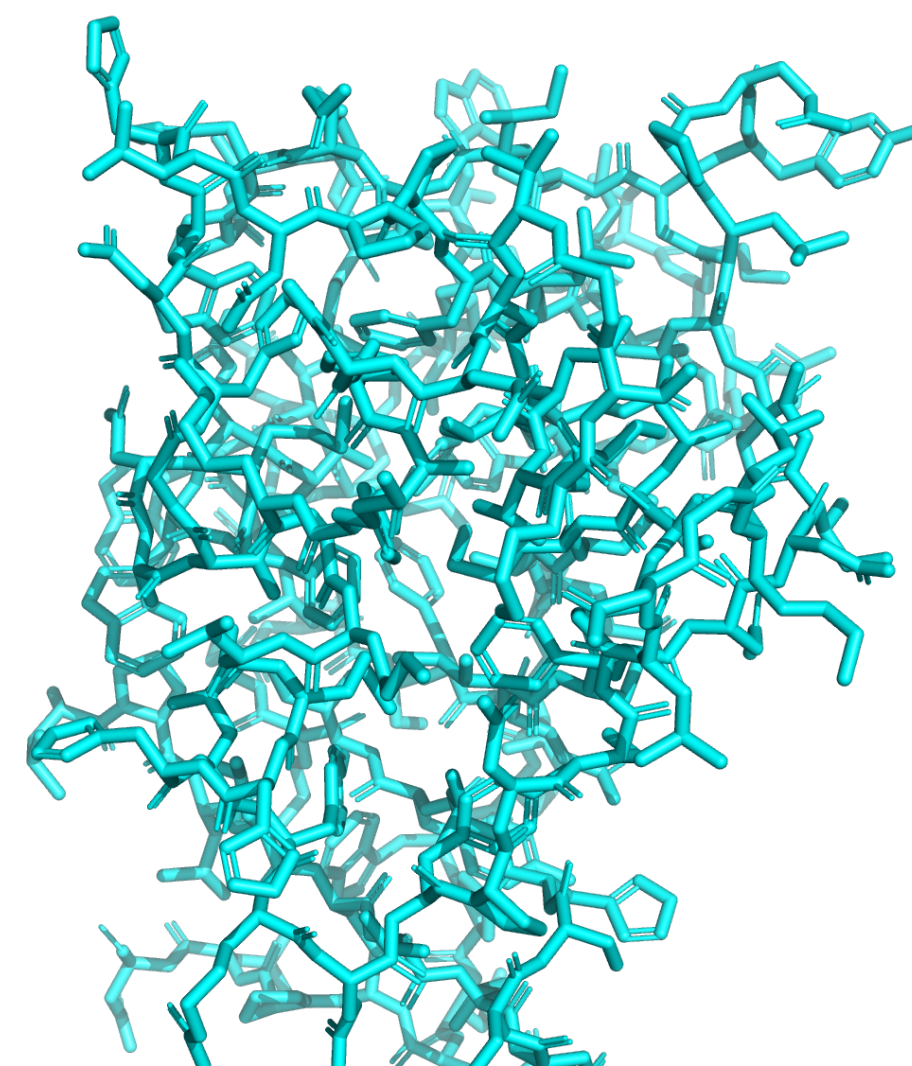# Antibody-Antigen Docking

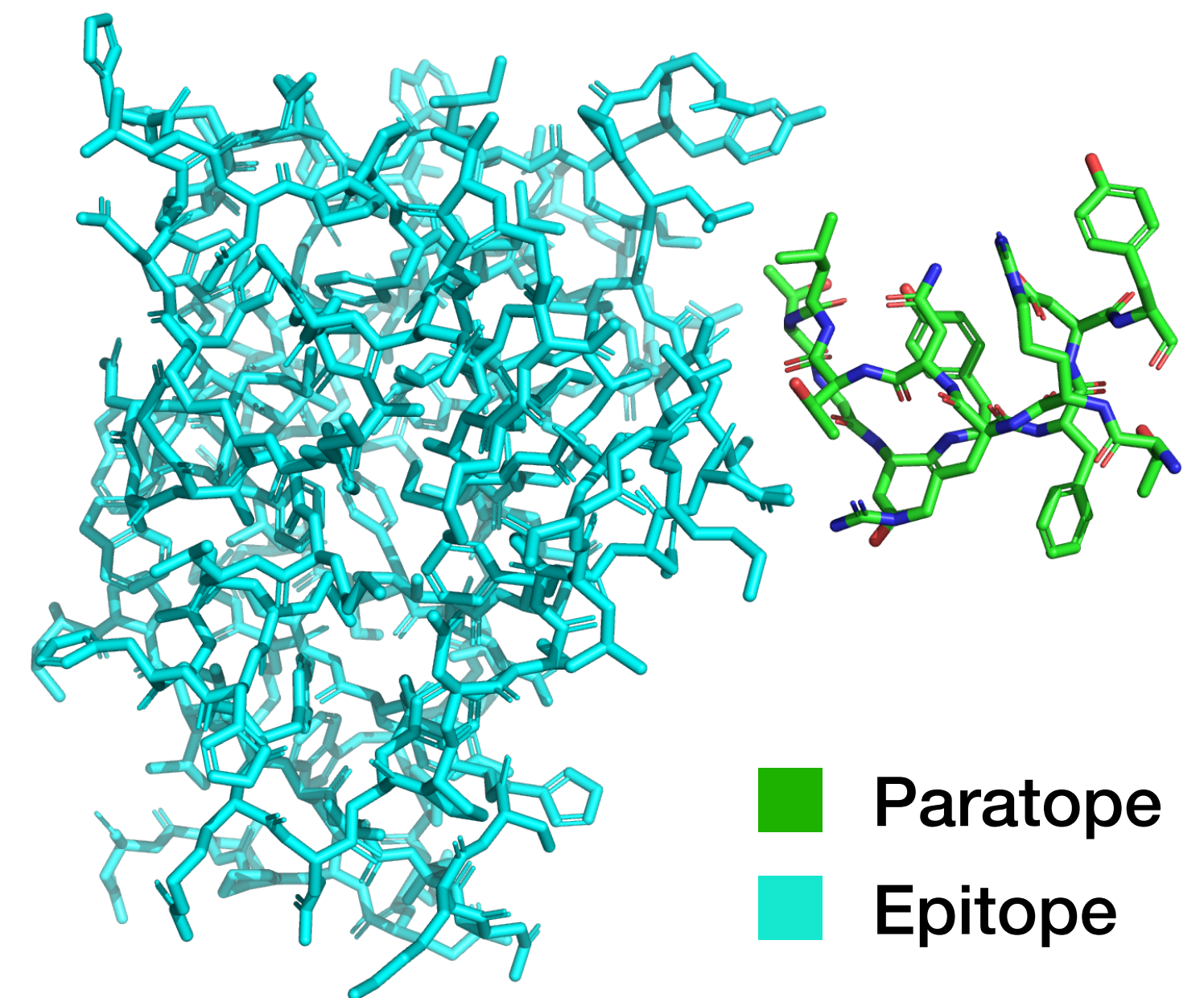## Formulation: point cloud completion

Paratope: RNTLTGDYFDY

+

Epitope (point cloud)

Dock

Paratope
Epitope

Antibody-Antigen Docking
Formulation: point cloud completion

- Given a given epitope points and a paratope sequence, a model needs to put paratope atoms at the right location

Paratope: RNTLTGDYFDY

+

Epitope (point cloud)

Dock

Paratope
Epitope

Hierarchical Equivariant Refinement Network
Main innovation: hierarchical equivariance

- Simultaneously fold and dock a paratope onto a given epitope

- We update coordinates iteratively by computing <u>forces</u> between residues/atoms

- Motivation: force is equivariant under rotation and translation

Hierarchical Equivariant Refinement Network
Main innovation: hierarchical equivariance

- Simultaneously fold and dock a paratope onto a given epitope

- We update coordinates iteratively by computing <u>forces</u> between residues/atoms

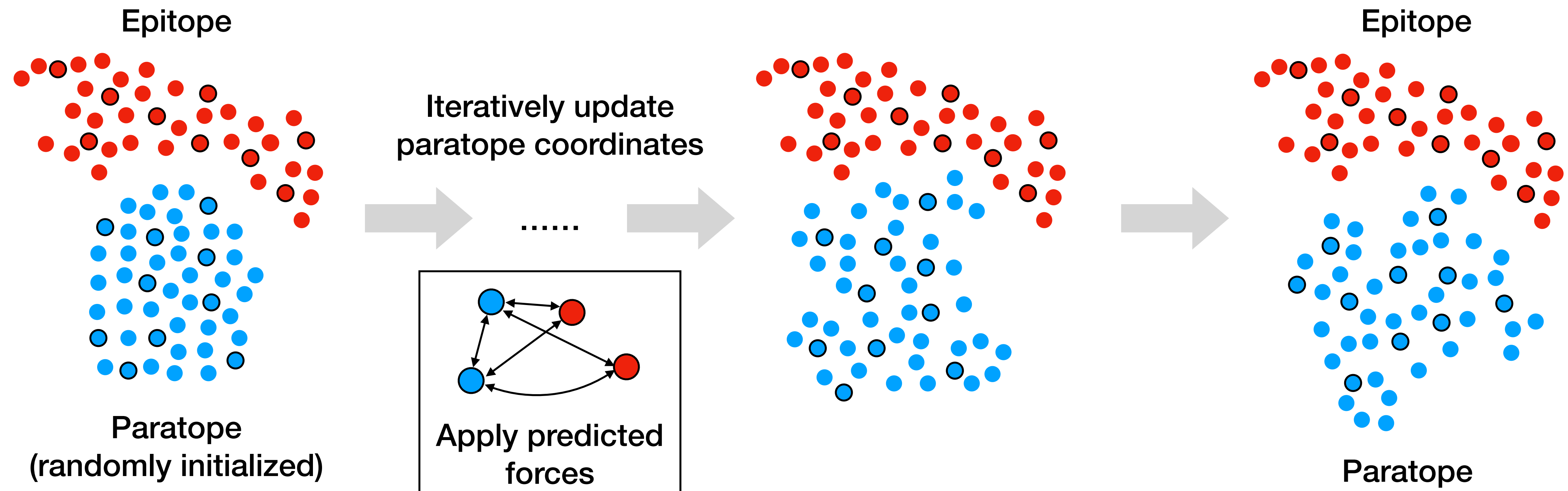- Motivation: force is equivariant under rotation and translation

Hierarchical Equivariant Refinement Network
Main innovation: hierarchical equivariance

Epitope

Next state $\mathcal{G}^{(t+1)}_{a,b}$

Paratope

Hierarchical Equivariant Refinement Network

Main innovation: hierarchical equivariance

- Step 1: Encode residues and atoms by a hierarchical message passing network

Epitope

Hierarchical encoder

$h(b_{j,l})$

$h(a_{i,k})$

Paratope

Next state $\mathcal{G}^{(t+1)}_{a,b}$

Hierarchical Equivariant Refinement Network
Main innovation: hierarchical equivariance

- Step 1: Encode residues and atoms by a hierarchical message passing network

- Step 2: Compute residue-level force between C_α atoms for global backbone update

Epitope

Hierarchical encoder

Residue-level force

Next state $\mathscr{G}_{a,b}^{(t+1)}$

$h(b_{j,l})$

$h(a_{i,k})$

$x(b_{k,1})$

$x_{i,1}^{(t)}$

$g_{i,k}^{(t)}$

$g_{i,j}^{(t)}$

$x_{j,1}^{(t)}$

(C_α atoms)

Paratope

6

Hierarchical Equivariant Refinement Network

Main innovation: hierarchical equivariance

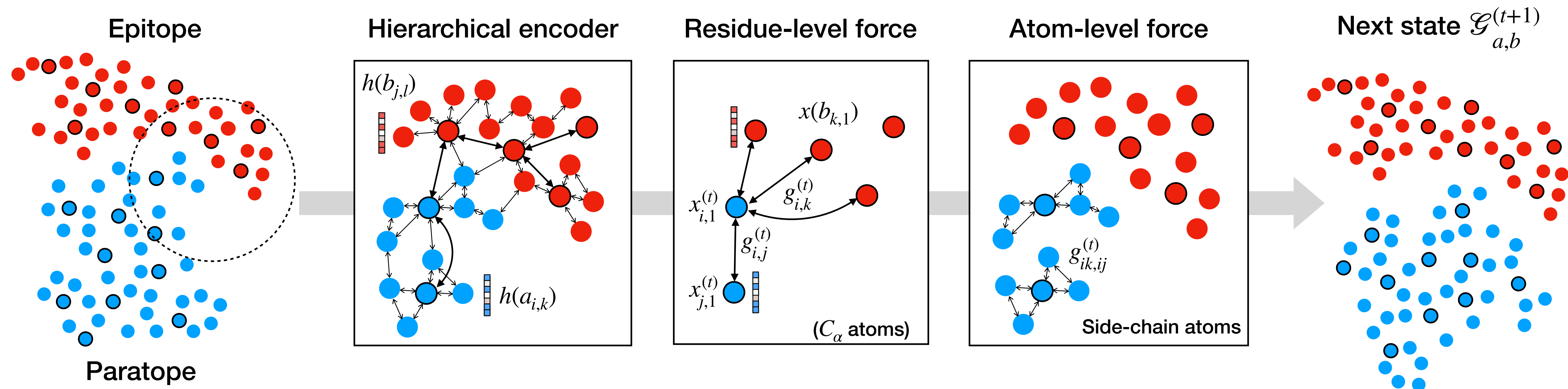- Step 1: Encode residues and atoms by a hierarchical message passing network

- Step 2: Compute residue-level force between C_α atoms for global backbone update

- Step 3: Compute atom-level forces within each residue for local side-chain update

Epitope

Paratope

Hierarchical encoder

$h(b_{j,l})$

$h(a_{i,k})$

Residue-level force

$x(b_{k,1})$

$g_{i,k}^{(t)}$

$x_{i,1}^{(t)}$

$g_{i,j}^{(t)}$

$x_{j,1}^{(t)}$

(C_α atoms)

Atom-level force

$g_{ik,ij}^{(t)}$

Side-chain atoms

Next state $\mathscr{G}_{a,b}^{(t+1)}$

6

Antibody-Antigen Binding

## A tale of two problems: docking and generation

- **Docking**: given an antibody paratope sequence and epitope 3D structure, can we predict whether/how they bind?

  - An accurately docked paratope allows us to easily calculate its binding affinity

- **Generation**: can we generate an paratope with high binding affinity to a given epitope?

  - This inverse problem is difficult due to a combinatorial search space

Binding affinity

Dock

Design

Antibodies
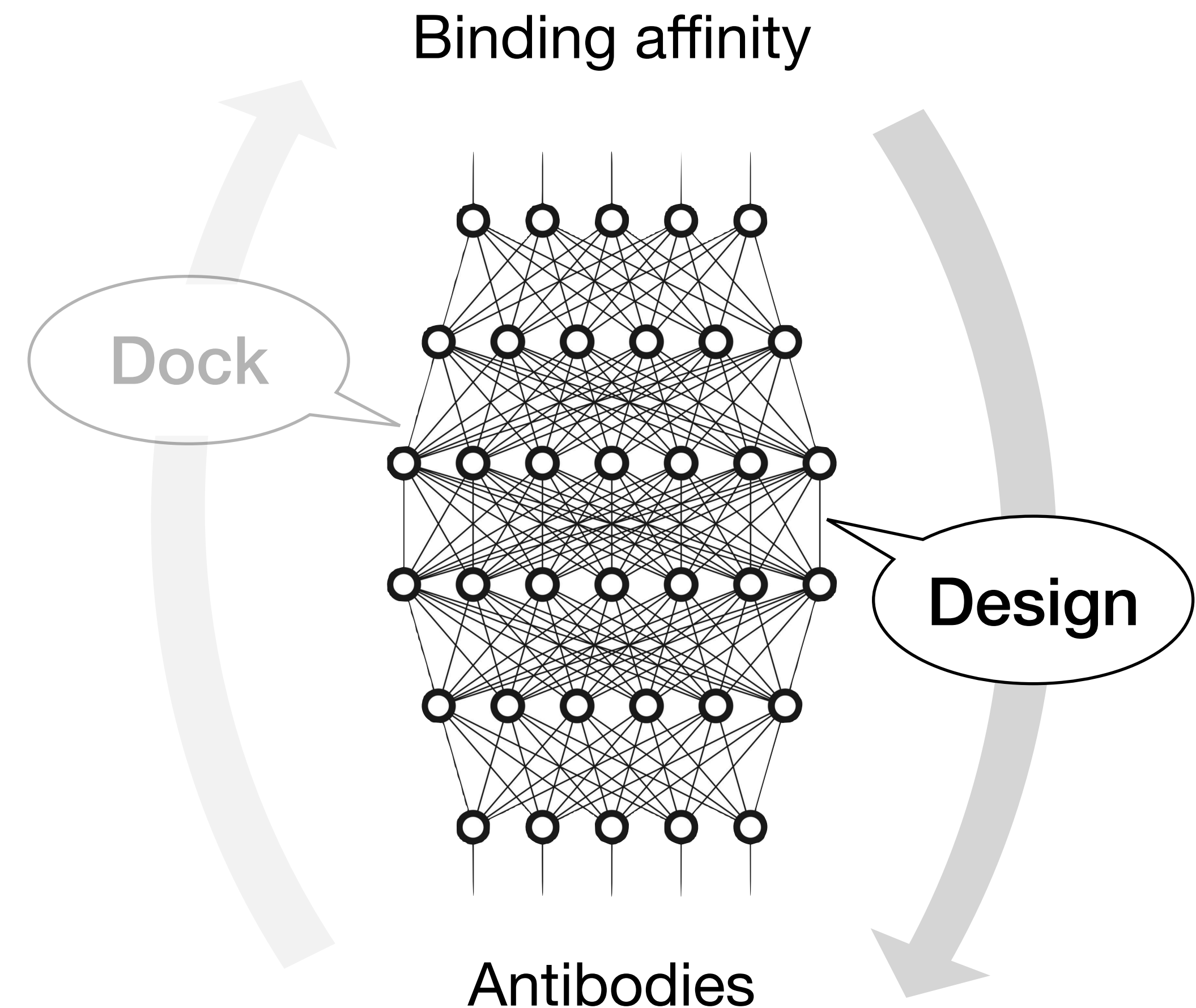
Epitope-specific Antibody Generation
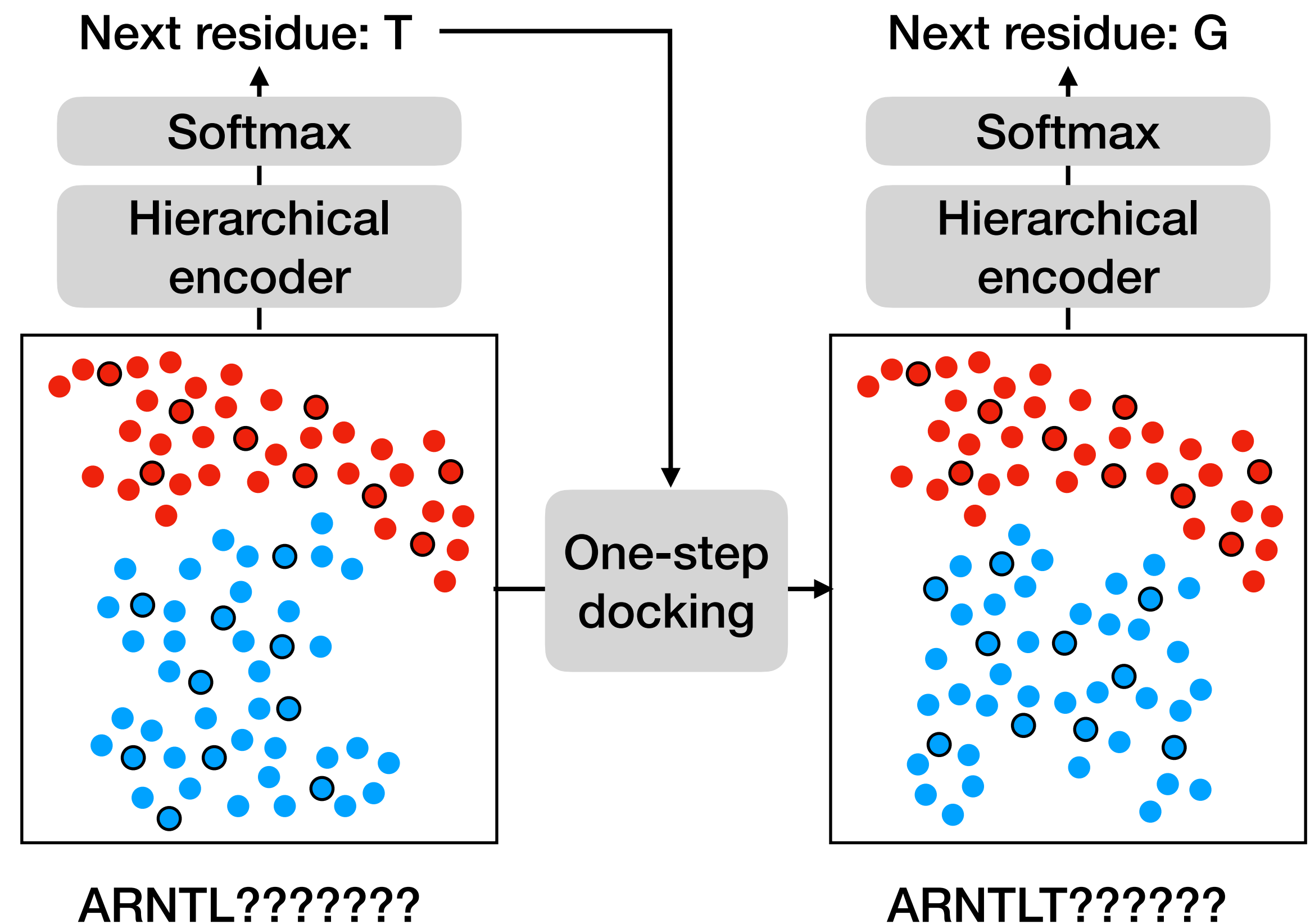## Formulation: point cloud completion

- Similar to docking, the generation task can also be viewed as point cloud completion

- We only need to slightly adapt the docking workflow for paratope generation

Generate

Epitope

Paratope: RNTLTGDYFDY
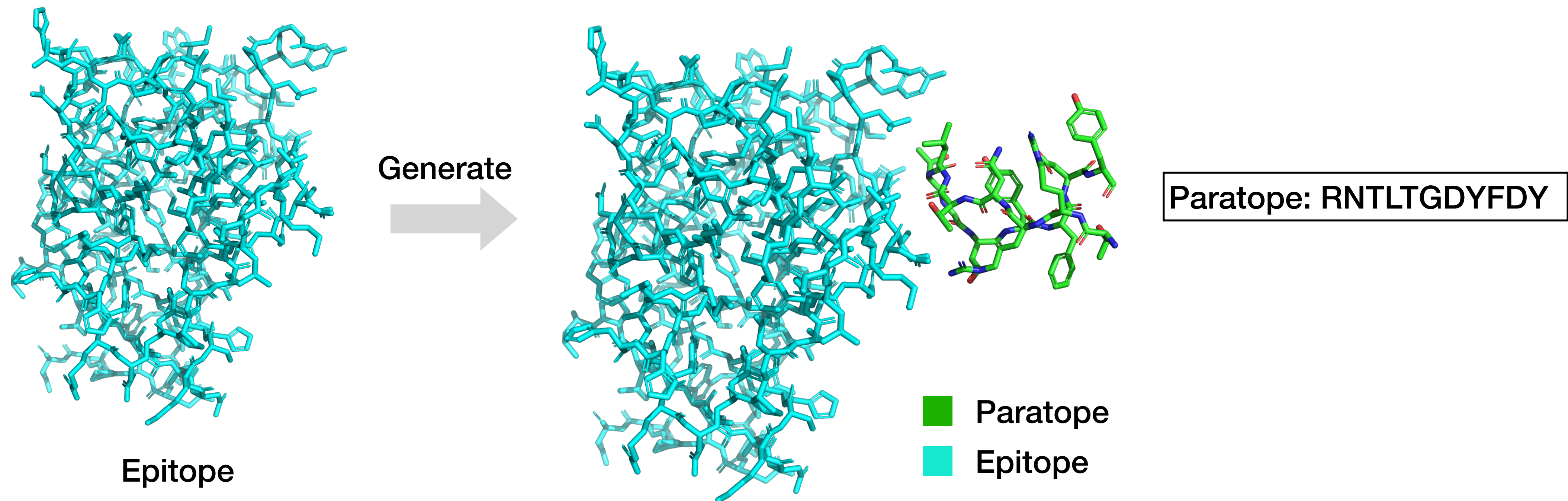
Paratope
Epitope

HERN Paratope Generation Procedure

Main innovation: geometric hidden state representation

- HERN generates paratope autoregressively by adding one residue at a time

- Different from standard RNNs, each HERN hidden state is a paratope-epitope complex rather than a partial paratope sequence

- In each generation step, we use the docking model to infer the geometric hidden state from a partial paratope sequence

Next residue: T

| Softmax |
| Hierarchical encoder |

ARNTL??????

One-step docking

Next residue: G

| Softmax |
| Hierarchical encoder |

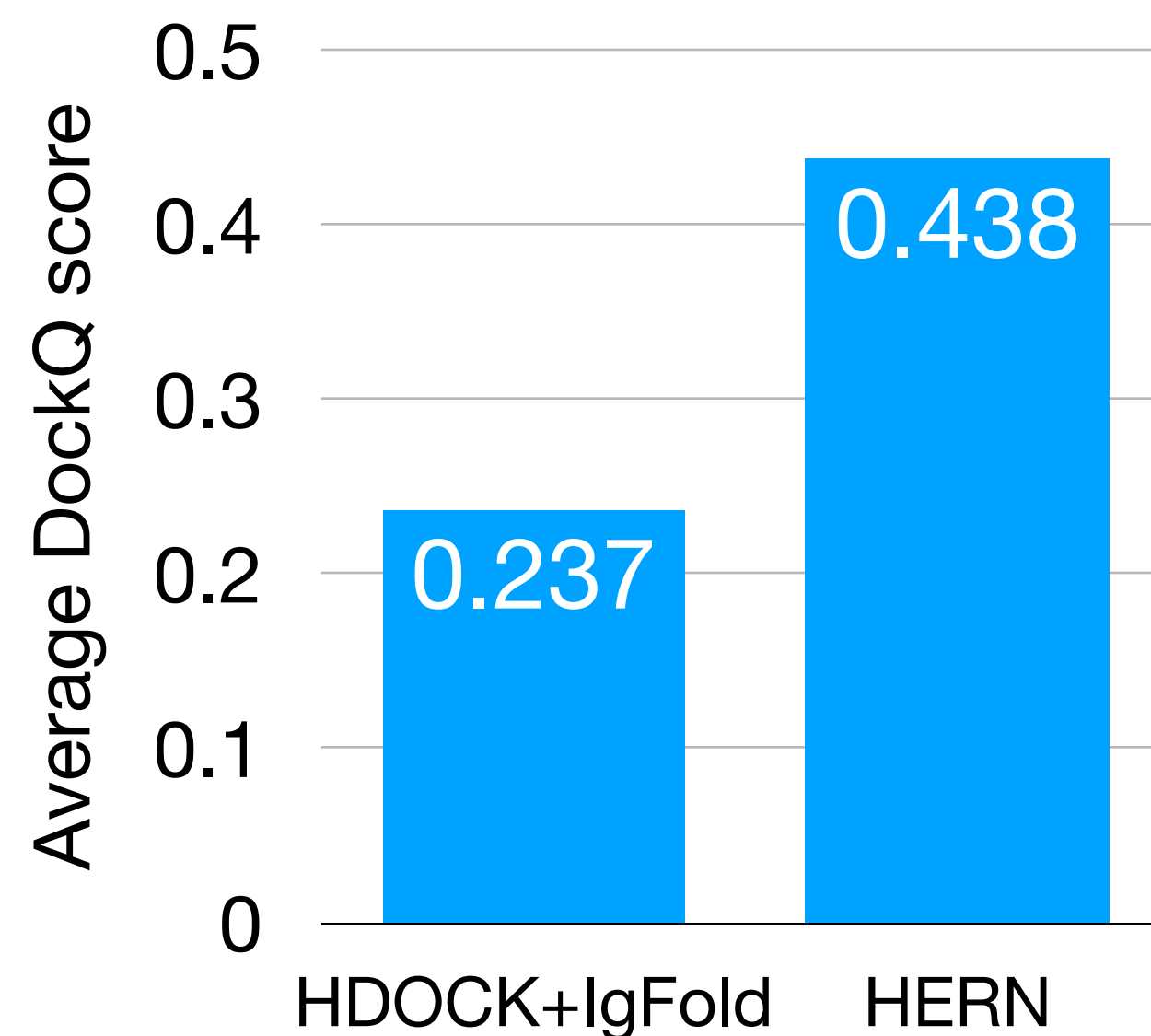ARNTLT?????

Evaluation: Docking Performance
Does our docked structure agree with the ground truth?

- Baseline: we fold a paratope sequence using IgFold (Ruffolo & Gray, 2022) and dock it using the HDOCK model (Yan et al., 2020)

- We report the DockQ metric (Basu & Wallner 2016) over a test set of 60 antibody-antigen complexes (Adolf-Bryfogle et al., 2018)

Evaluation: Docking Performance

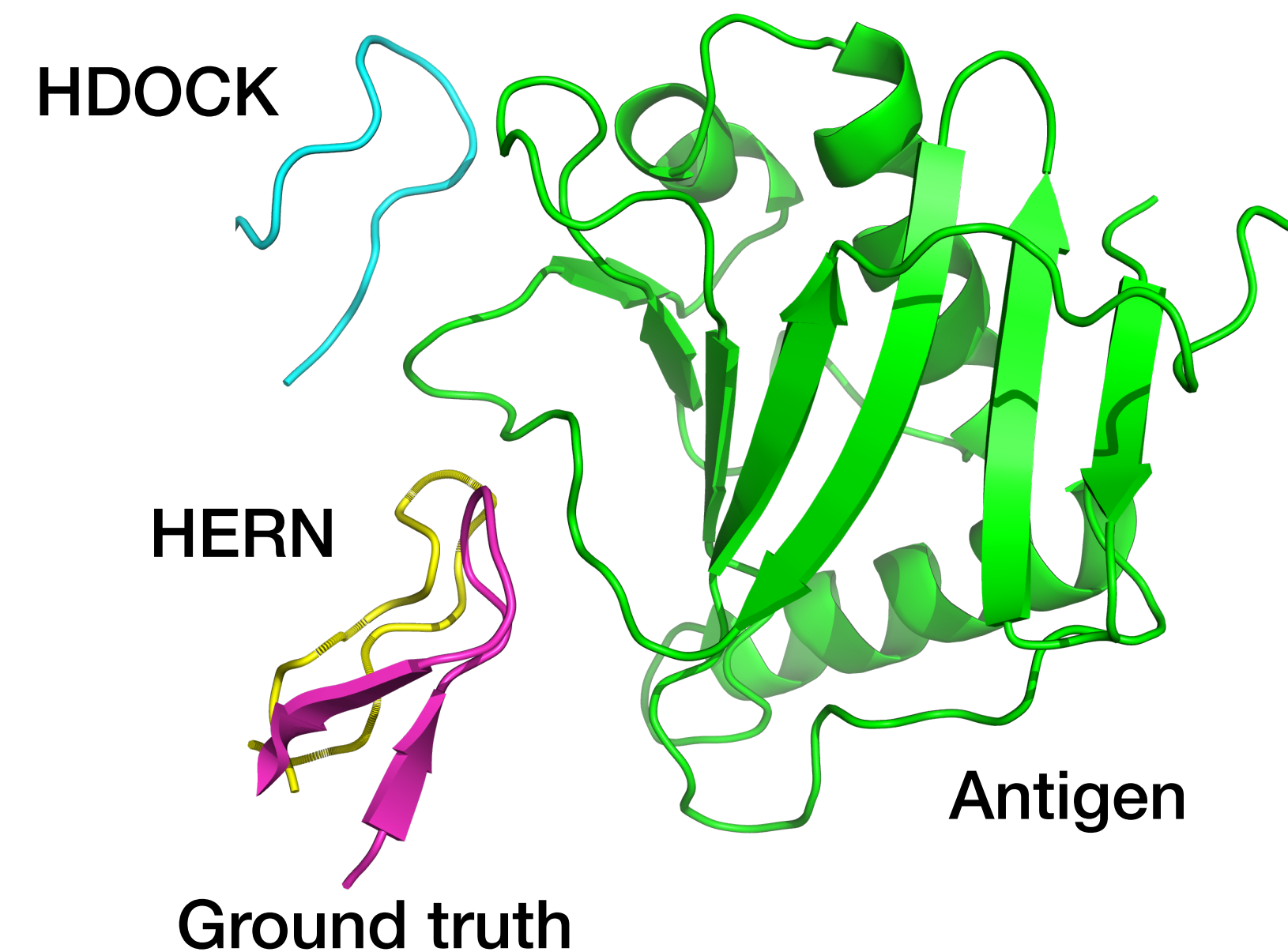## Does our docked structure agree with the ground truth?

- Baseline: we fold a paratope sequence using IgFold (Ruffolo & Gray, 2022) and dock it using the HDOCK model (Yan et al., 2020)

- We report the DockQ metric (Basu & Wallner 2016) over a test set of 60 antibody-antigen complexes (Adolf-Bryfogle et al., 2018)

Evaluation: Docking Performance

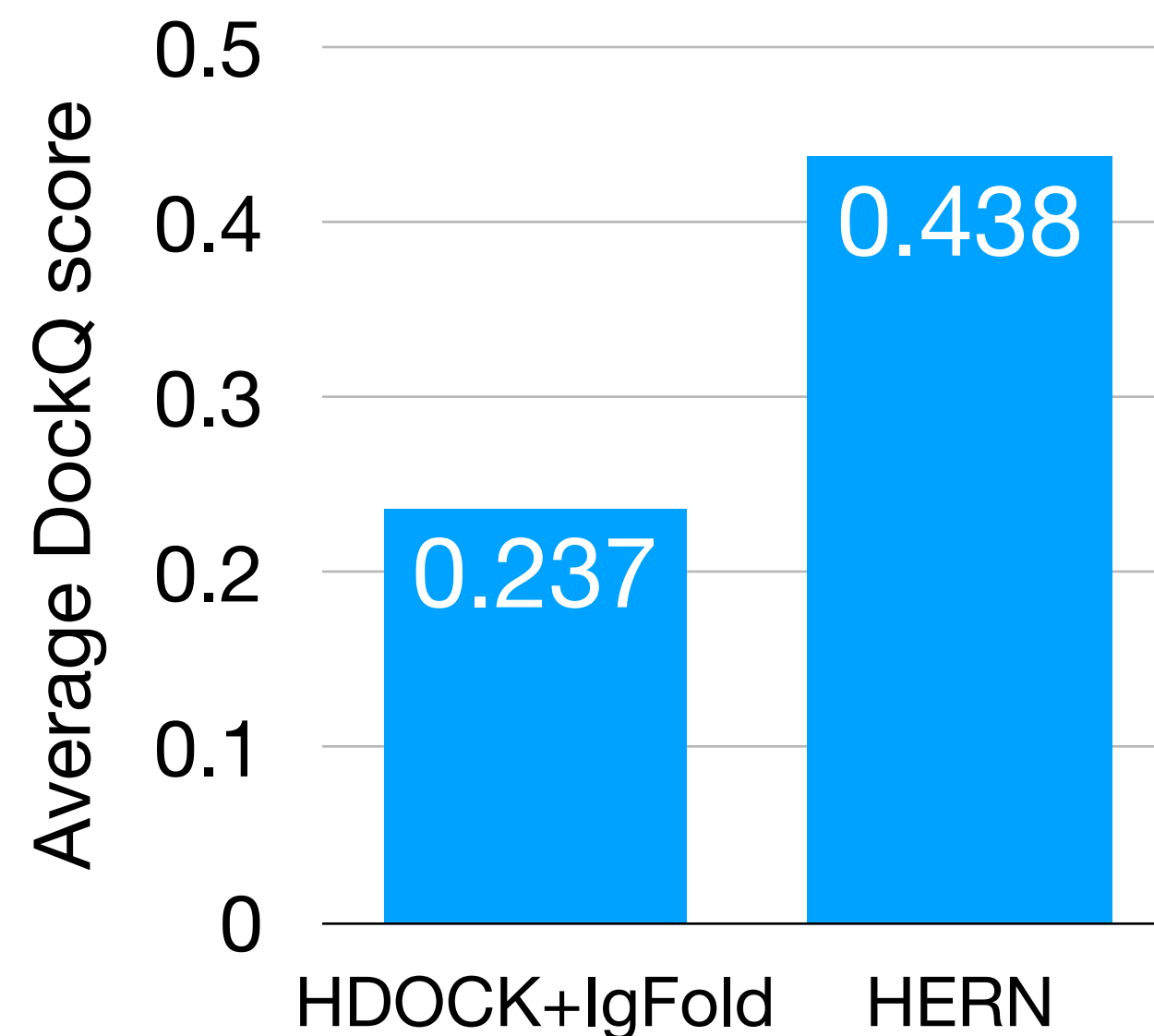## Does our docked structure agree with the ground truth?

- Baseline: we fold a paratope sequence using IgFold (Ruffolo & Gray, 2022) and dock it using the HDOCK model (Yan et al., 2020)

- We report the DockQ metric (Basu & Wallner 2016) over a test set of 60 antibody-antigen complexes (Adolf-Bryfogle et al., 2018)
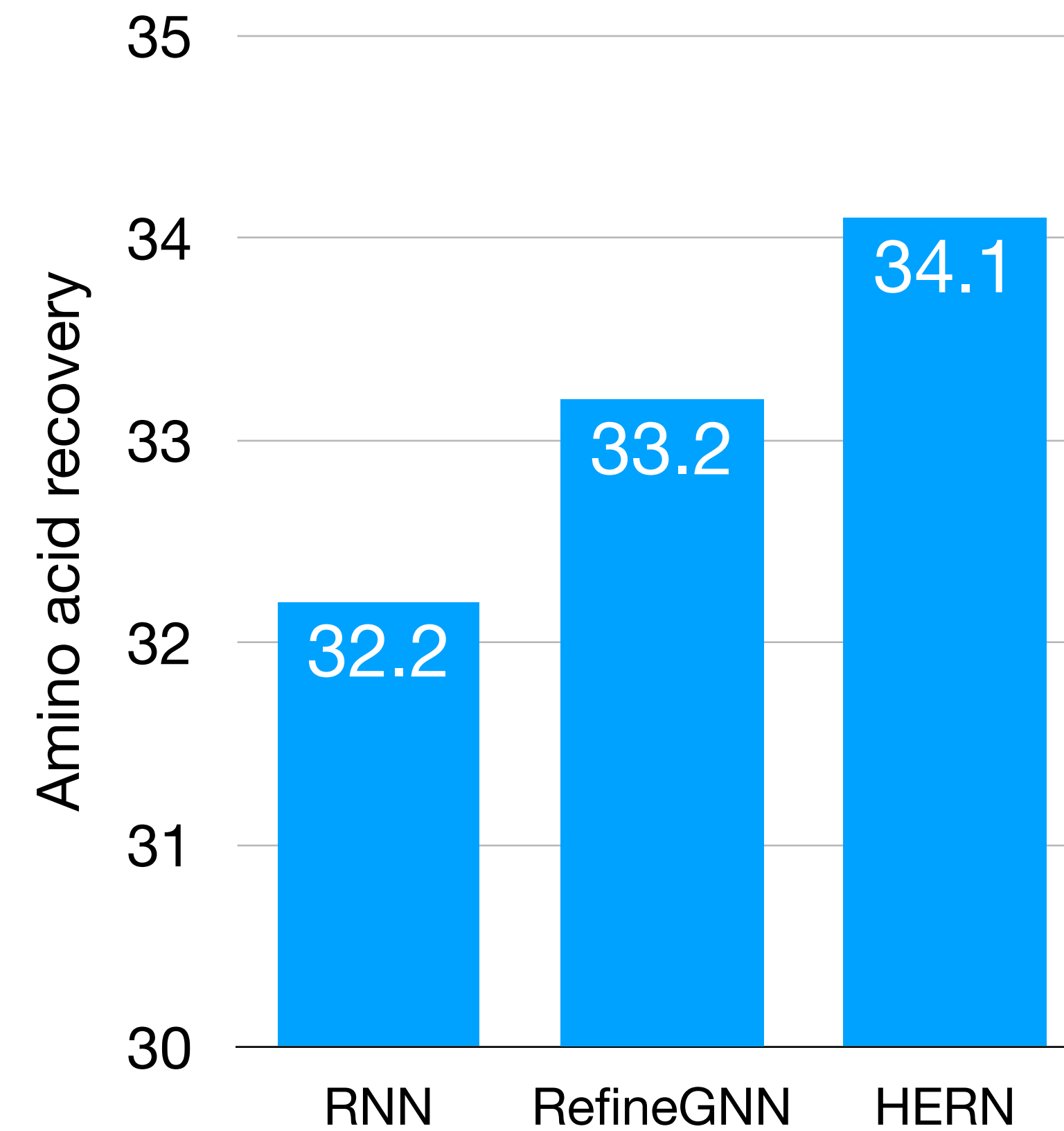
Evaluation: Generation Performance

Is our generate paratope close to the ground truth?

- Baseline 1: sequence-based RNN

- Baseline 2: structure-based RefineGNN (Jin et al., 2021)

- Both models employ a GNN-based epitope encoder and attention layers to <u>condition</u> on the epitope

- We report the amino acid recovery (AAR) metric on the same test set

- AAR = percentage of residues matching the ground truth

Evaluation: Generation Performance

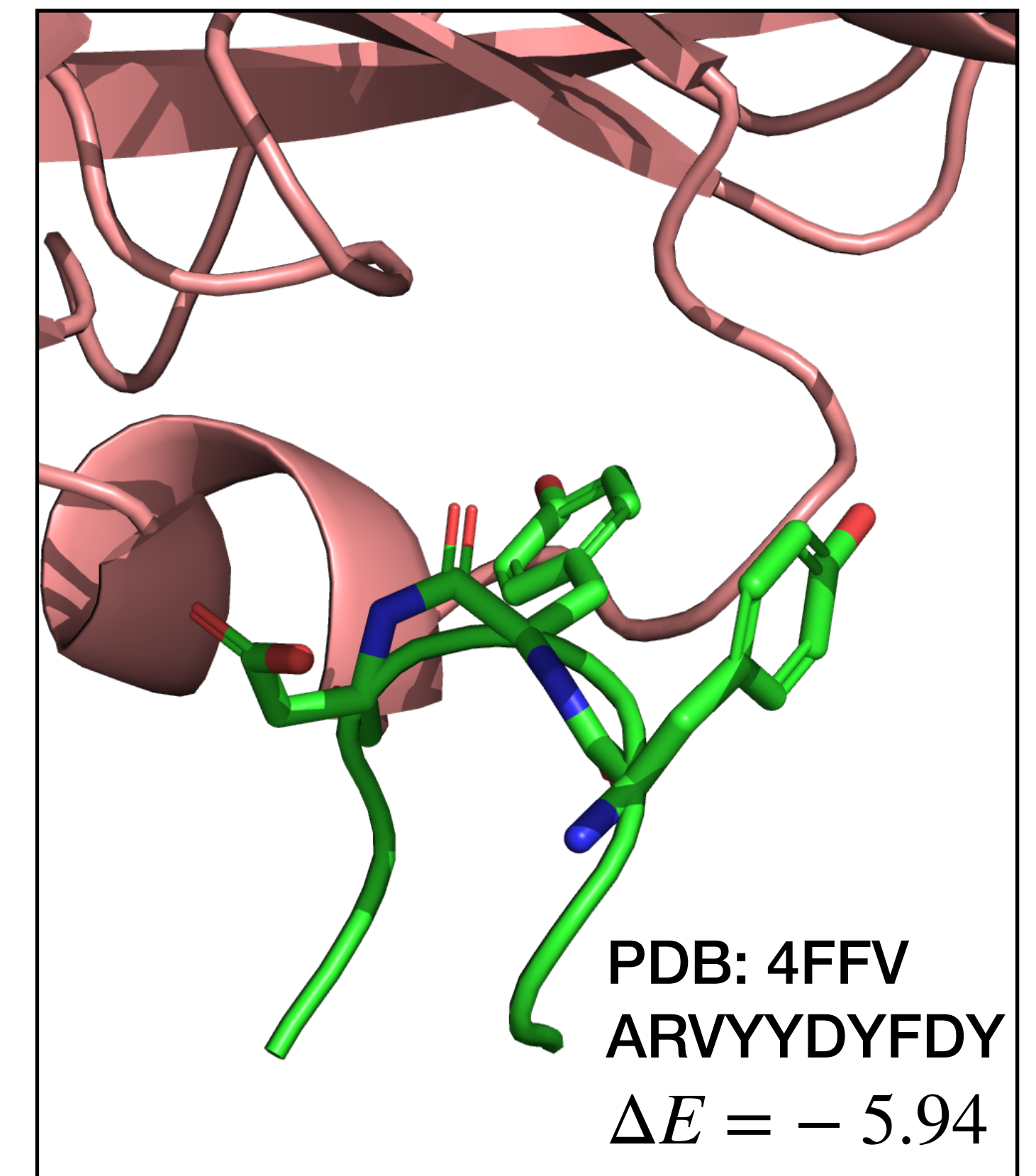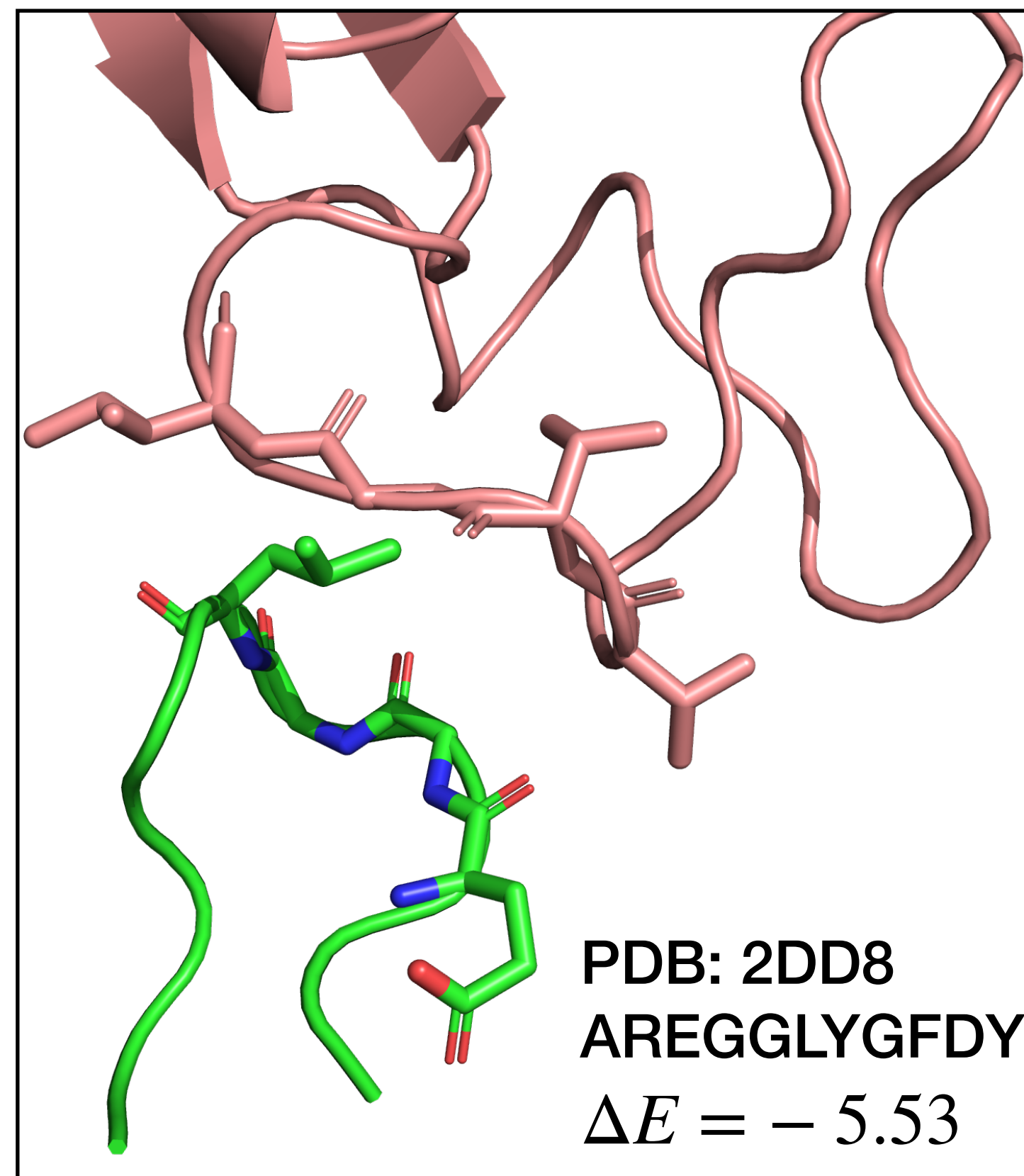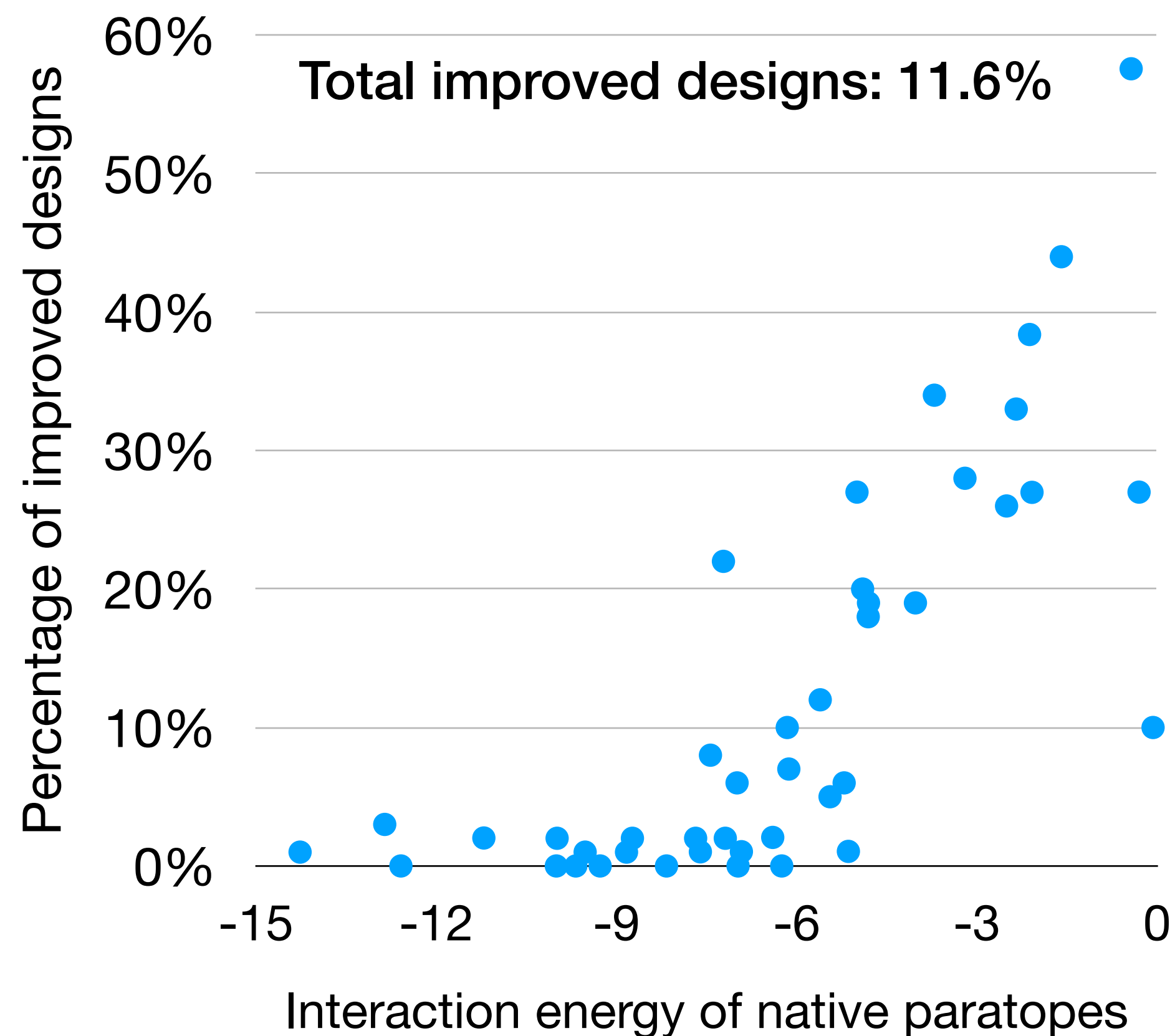## Is our generate paratope close to the ground truth?

- Baseline 1: sequence-based RNN

- Baseline 2: structure-based RefineGNN (Jin et al., 2021)

- Both models employ a GNN-based epitope encoder and attention layers to <u>condition</u> on the epitope

- We report the amino acid recovery (AAR) metric on the same test set

- AAR = percentage of residues matching the ground truth

Evaluation: Generation Performance

Can we discover new paratopes better than native binders?

- Calculate interaction energy of generated paratopes and compare with native ones

Total improved designs: 11.6%

Percentage of improved designs (y-axis): 0%, 10%, 20%, 30%, 40%, 50%, 60%

Interaction energy of native paratopes (x-axis): -15, -12, -9, -6, -3, 0

PDB: 2DD8
AREGGLYGFDY
$\Delta E = -5.53$

PDB: 4FFV
ARVYYDYFDY
$\Delta E = -5.94$

Summary and Acknowledgement

We propose HERN for paratope docking and design, whose key innovations are:

- Simultaneous folding and docking v.s. rigid-body docking

- Hierarchical equivariance via residue/atom-level force update

- Geometric hidden states for sequence generation

Generate

Dock

Paratope: RNTLTGDYFDY

■ Paratope
■ Epitope

Summary and Acknowledgement

We propose HERN for paratope docking and design, whose key innovations are:

- Simultaneous folding and docking v.s. rigid-body docking

- Hierarchical equivariance via residue/atom-level force update

- Geometric hidden states for sequence generation

Generate

Dock

Paratope: RNTLTGDYFDY

■ Paratope
■ Epitope

13