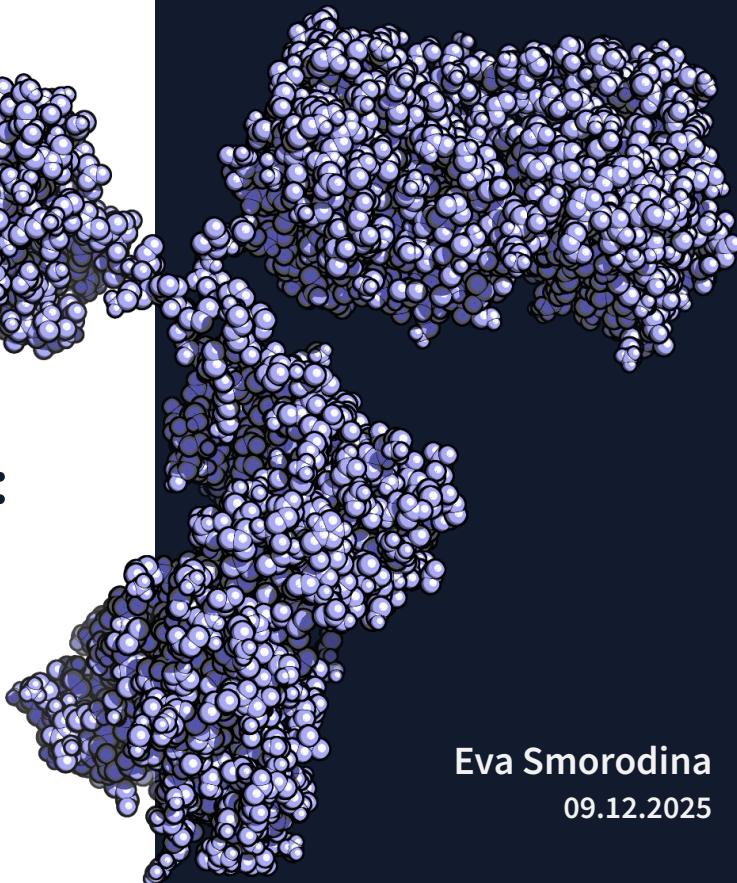


Computational structural biology: how to analyze and visualize molecules in 3D

Introduction to PyMOL



Eva Smorodina
09.12.2025



ribes.ev@gmail.com



eva-smorodina



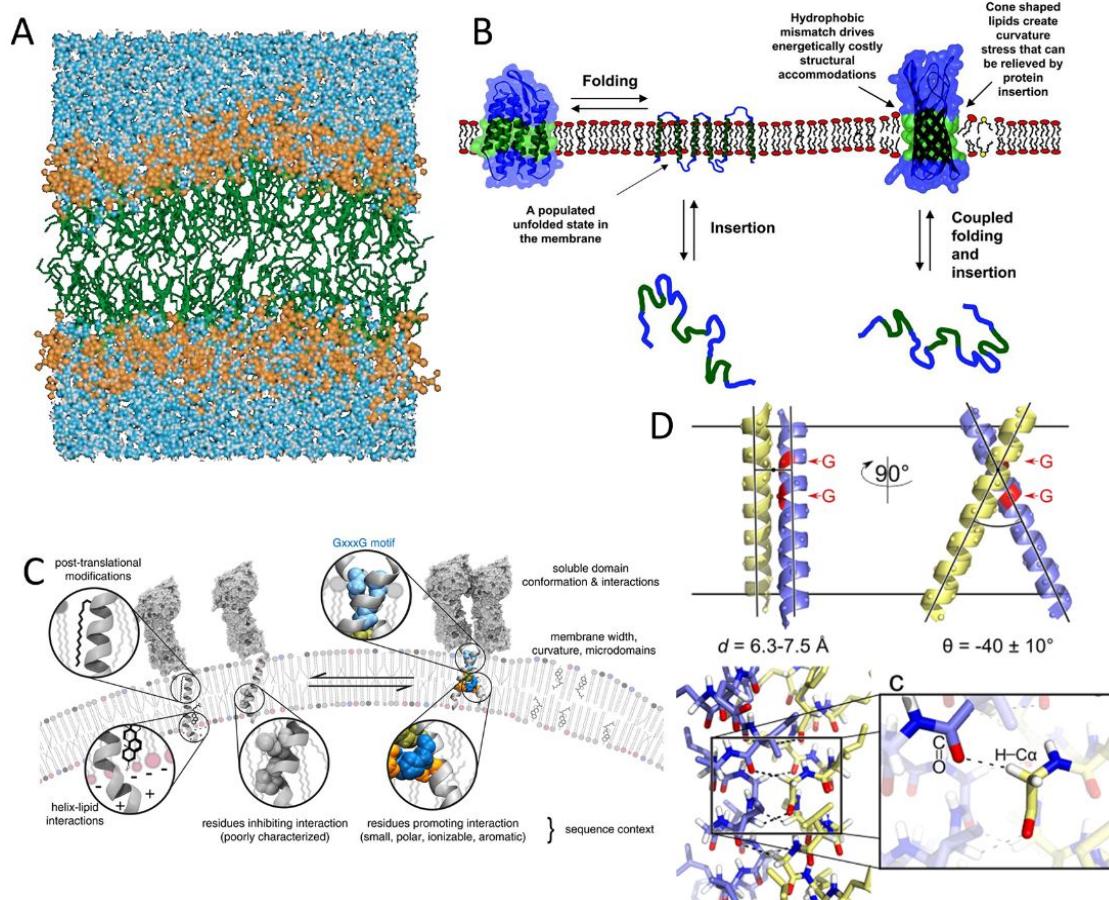
evasmorodina.com

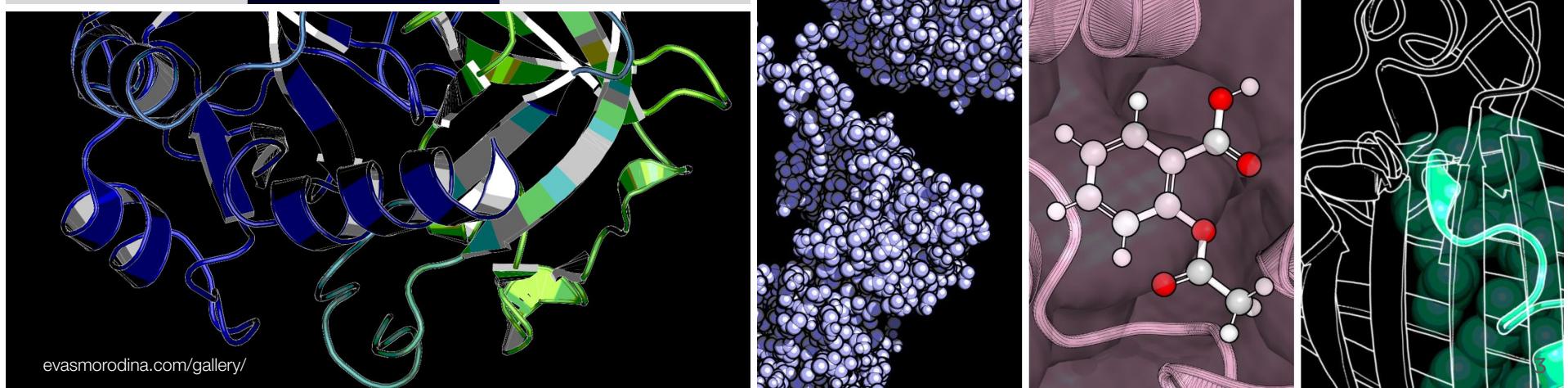
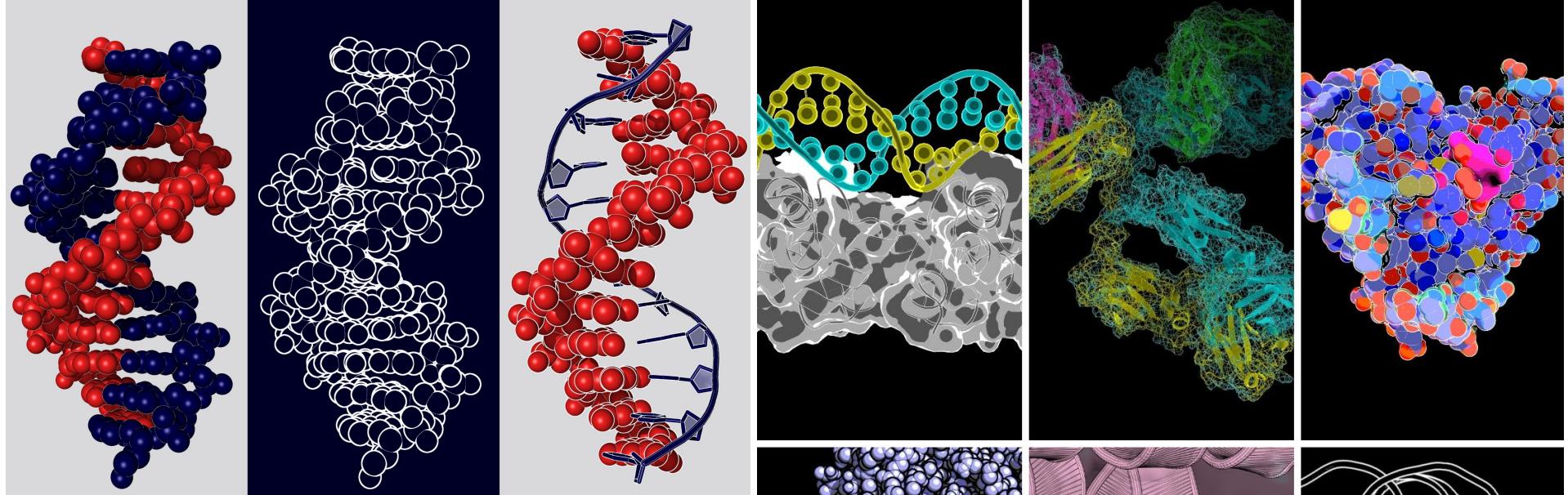
What is computational structural biology?

Computational structural biology:

- interdisciplinary field (biology + chemistry + math + computer science + physics)
- analysis, modeling and prediction of 3D structures of biomolecules (proteins, nucleic acids, etc.)
- helps to understand and explain molecular functions (PPIs, ligand-protein binding, etc.)

Molecular modeling is part of structural bioinformatics that represent and simulate the properties and behavior of molecules.





Why do we need molecular structures at all?

Central dogma of molecular biology:
sequence → **structure** → function

Challenges in experimental structure determination
usually associated with:

- Flexibility/dynamics/disorderness/conformational changes (IDPs)
- Heterogeneity/large size/complexity (protein/RNA/DNA complexes)
- Hydrophobicity/membrane association (ion channels)
- Transient interactions/non-covalent contacts (enzymes)

+ **Experimental approaches (X-ray, cryo-EM, NMR)**
are time- and resource-consuming, taking months
and sometimes years of effort

Specific examples

- **Understanding molecular functions:** observing the precise arrangement of atoms in enzyme binding sites (catalytic triad) → how they catalyze chemical reactions
- **Drug design and development:** protein structures → how ligands interact with target molecules → more effective drugs (HIV protease inhibitors)
- **Revealing disease mechanisms:** structures of virus proteins (SARS-CoV2's spike) → antiviral drug development (mRNA vaccines)
- **Learning biological interactions:** how receptors bind to signaling molecules (insulin) → cellular communication pathways → developing therapies (diabetes treatment)
- **Evolutionary relationships:** comparing molecular structures across species (cytochrome c proteins) → evolutionary history and relatedness
- **Engineering, nano- and biotechnology:** molecular structures → design of materials and devices for specific functions (liposomal nanocarriers)

Applications of molecular modeling

Things to remember about structures:

- Many sequences, but much fewer structures
- Similar sequences can have different structures
- Different sequences can have similar structures
- Structural data are more complex and less investigated than sequence data

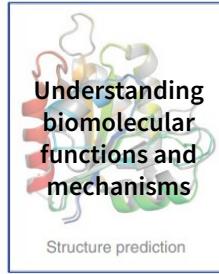
→ Molecular modeling can be useful in all of the cases

<p>Understanding biomolecular functions and mechanisms</p> <p>Structure prediction</p>	<p>Personalized medicine</p> <p>Protein–protein docking</p>	<p>Drug discovery in development</p> <p>Ligand docking</p>	<p>Rational design of vaccines and antibodies</p> <p>Loop modeling</p>	<p>Biotechnology and protein engineering</p> <p>Biomineral surface docking</p>
<p>Design of proteins and protein–protein interactions</p> <p>Protein design</p>	<p>Development of in silico methods to complement in vitro studies</p> <p>Modeling with experimental data</p>	<p>Tasks</p>	<p>Symmetric assemblies</p>	<p>Non-canonical chemistries</p>
<p>Antibodies</p>	<p>Membrane proteins</p>	<p>RNA and DNA</p>	<p>Peptides</p>	<p>Carbohydrates</p>

Applications of molecular modeling

Main techniques:

- **Structure prediction**
 - Molecular docking
 - Chemoinformatics
 - Molecular modelling
 - Protein design
 - Molecular dynamics
- + Structural analysis
+ Structure visualization
+ Programming
+ Data analysis
+ Machine learning

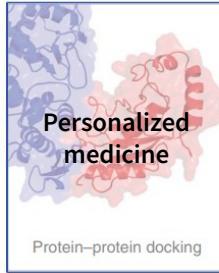


- **Structure prediction** is a computational method used to determine the 3D structure of biomolecules (such as proteins, RNA, DNA, lipids, carbohydrates) based on their sequences
- It's essential for understanding protein structures and hence their functions and roles in biological processes, offering a more efficient alternative to time- and resource-intensive experimental methods

Applications of molecular modeling

Main techniques:

- Structure prediction
 - **Molecular docking**
 - Chemoinformatics
 - Molecular modelling
 - Protein design
 - Molecular dynamics
- + Structural analysis
+ Structure visualization
+ Programming
+ Data analysis
+ Machine learning

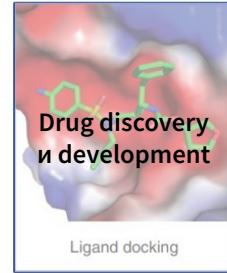


- **Protein-protein docking** is a computational method used to predict the interaction and binding conformation between two proteins (or macromolecules in general)
- It's essential for understanding protein interactions and designing drugs, as it helps in identifying how proteins bind to each other, thereby influencing biological functions and potential therapeutic targets

Applications of molecular modeling

Main techniques:

- Structure prediction
 - **Molecular docking**
 - **Chemoinformatics**
 - Molecular modelling
 - Protein design
 - Molecular dynamics
- + Structural analysis
+ Structure visualization
+ Programming
+ Data analysis
+ Machine learning

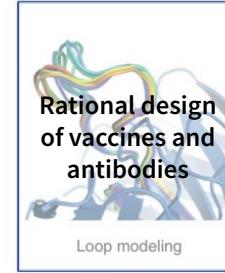


- **Small molecule docking** is a computational method employed in drug discovery to predict the binding conformation and affinity of small molecules (ligands) with biological targets (e. g. proteins)
- It enables the identification of potential drug candidates and optimization of their interactions with target proteins, aiding in the development of new medications
- **Chemoinformatics** is a multidisciplinary field that combines principles of chemistry, computer science, and information technology to analyze and interpret chemical data
- It enables the efficient storage, retrieval, and analysis of chemical information, helping to find new drug candidates, optimize lead compounds, and predict their properties and activities

Applications of molecular modeling

Main techniques:

- Structure prediction
 - Molecular docking
 - Chemoinformatics
 - **Molecular modelling**
 - Protein design
 - Molecular dynamics
- + Structural analysis
+ Structure visualization
+ Programming
+ Data analysis
+ Machine learning



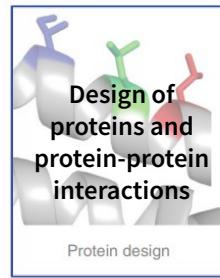
- **Loop modeling** is a computational method used to predict the 3D structure of flexible regions (e.g. loops), in proteins, which play critical roles in protein function and interactions (such as antibody-antigen binding)
- It's essential for understanding protein folding and dynamics, designing novel proteins with specific functionalities, and optimizing drug molecules to enhance their binding affinity and specificity

Applications of molecular modeling

Main techniques:

- Structure prediction
- Molecular docking
- Chemoinformatics
- Molecular modelling
- **Protein design**
- Molecular dynamics

+ Structural analysis
+ Structure visualization
+ Programming
+ Data analysis
+ Machine learning



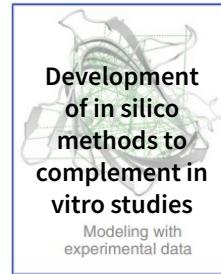
- **Protein design** involves either modifying existing proteins or constructing entirely new ones to achieve specific structures and functions
- It is crucial for biotechnology, medicine, and materials science, enabling the development of custom-made enzymes, therapeutics, and biomaterials tailored to specific needs and applications

Applications of molecular modeling

Main techniques:

- Structure prediction
- Molecular docking
- Chemoinformatics
- Molecular modelling
- Protein design
- **Molecular dynamics**

- + Structural analysis
- + Structure visualization
- + Programming
- + Data analysis
- + Machine learning

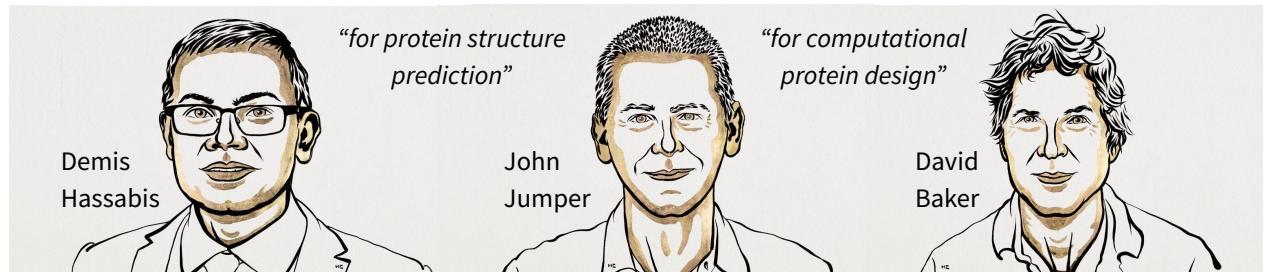


- **Biomolecular modeling with experimental data** integrates computational techniques with experimental measurements to simulate and understand complex biomolecular systems
- It uncovers molecular mechanisms and interactions, designs novel therapeutics, complements experimental observations, and guides further computational studies through experimental verification

Structure prediction and protein design have recently gained the most attention

The Nobel Prize in Chemistry

2024: for protein structure prediction (Demis Hassabis, John Jumper) and computational protein design (David Baker)



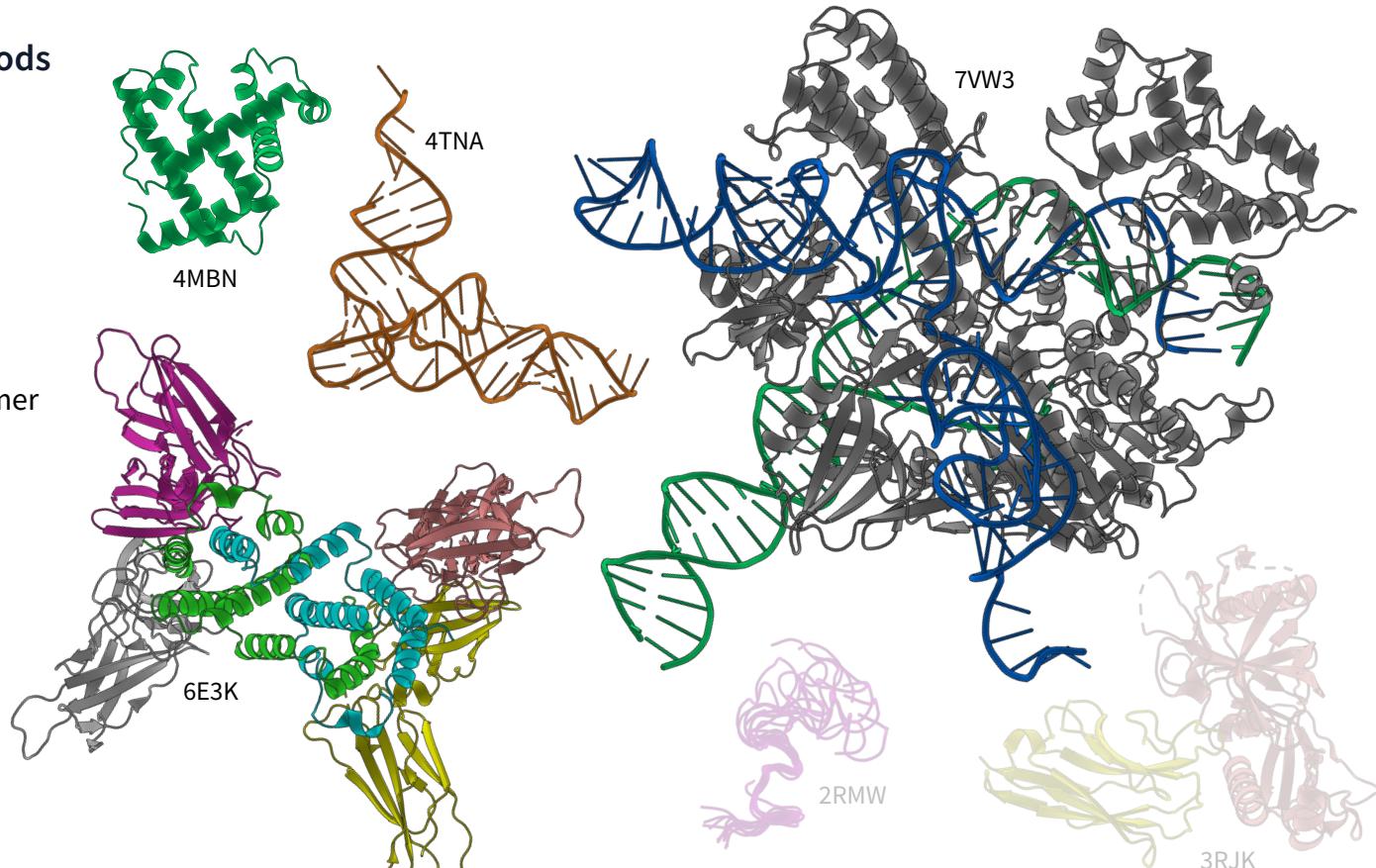
Types of methods for structure prediction and protein design

	Physics-based	Data-driven (AI)
Foundation	Physical laws and energy functions	Empirical data and statistical patterns
Accuracy	High, but depends on parameters	Data-dependent, high with good datasets
Speed	Computationally and time consuming	Slow and expensive for training, fast for prediction
Generalizability	Can predict novel systems	Limited by the scope of the data
Setup	Simulations of each step of the process	Output prediction based on learned patterns
Limitations	<ul style="list-style-type: none">• template and scoring function quality• high computational cost• limited sampling	<ul style="list-style-type: none">• dependency of data quality and size• hard-to-explain black box nature• generalization issues

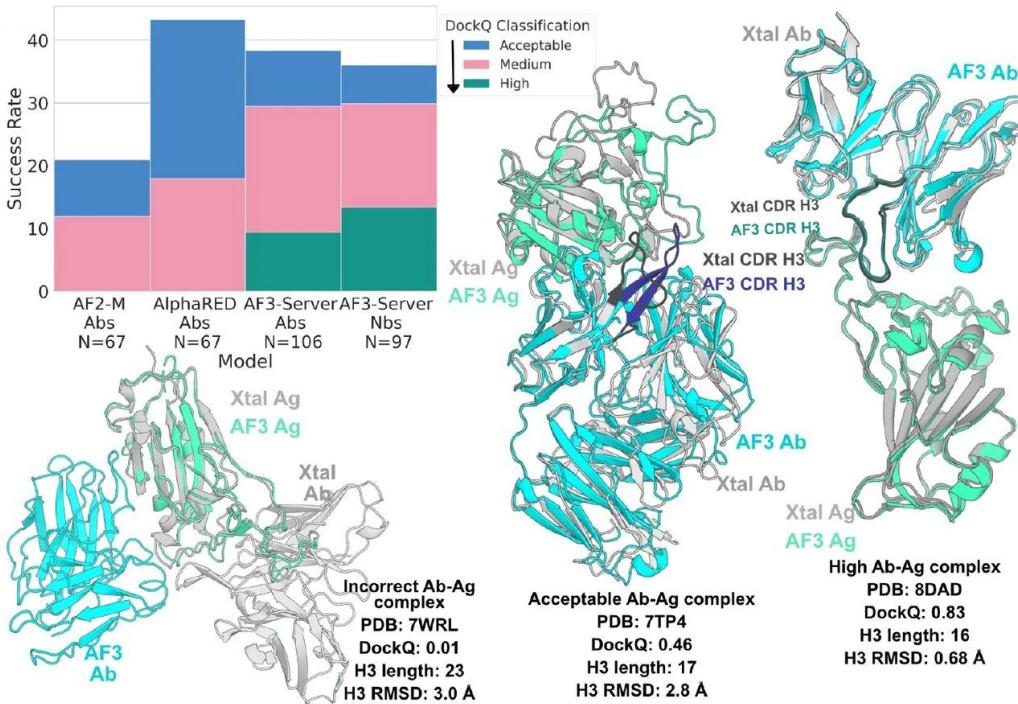
Biomolecular structures and corresponding structure prediction tools

Structure prediction methods

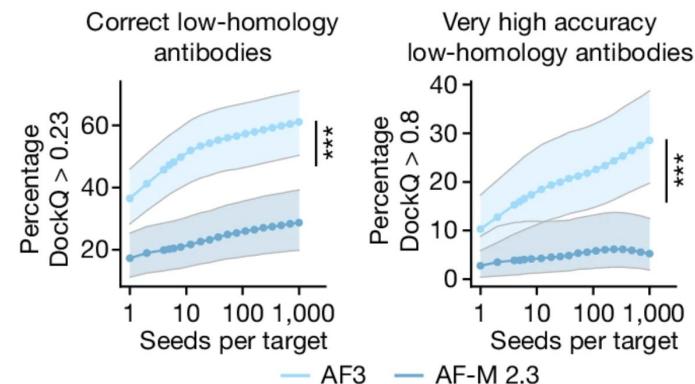
- **Monomer protein**
 - AlphaFold2
 - RoseTTAFold
 - OpenFold
 - trRosetta
- **Multimer protein**
 - AlphaFold-multimer
- **Monomer non-protein**
 - RhoFold
 - trRosettaRNA
 - Deep DNAshape
- **Multimer non-protein**
 - AlphaFold3
 - Boltz-1
 - Chai-1
- **Special cases**



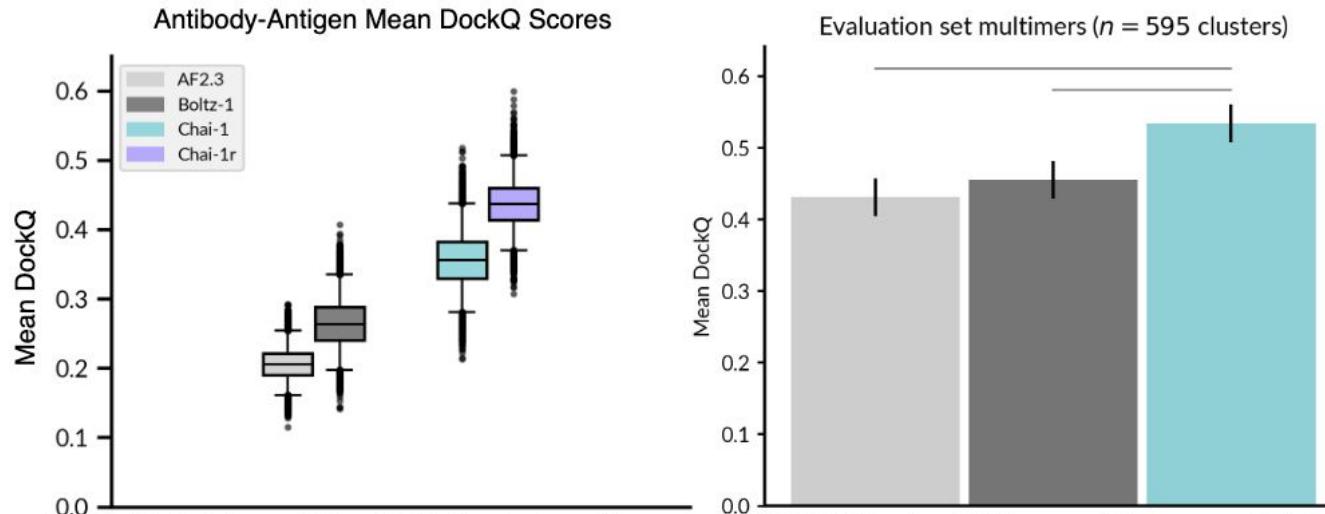
AlphaFold3



Only 60% of antibody-antigen pairs tested have DockQ > 0.23 → 40% are still expected to fail even after 1000 seeds with confidence ranking



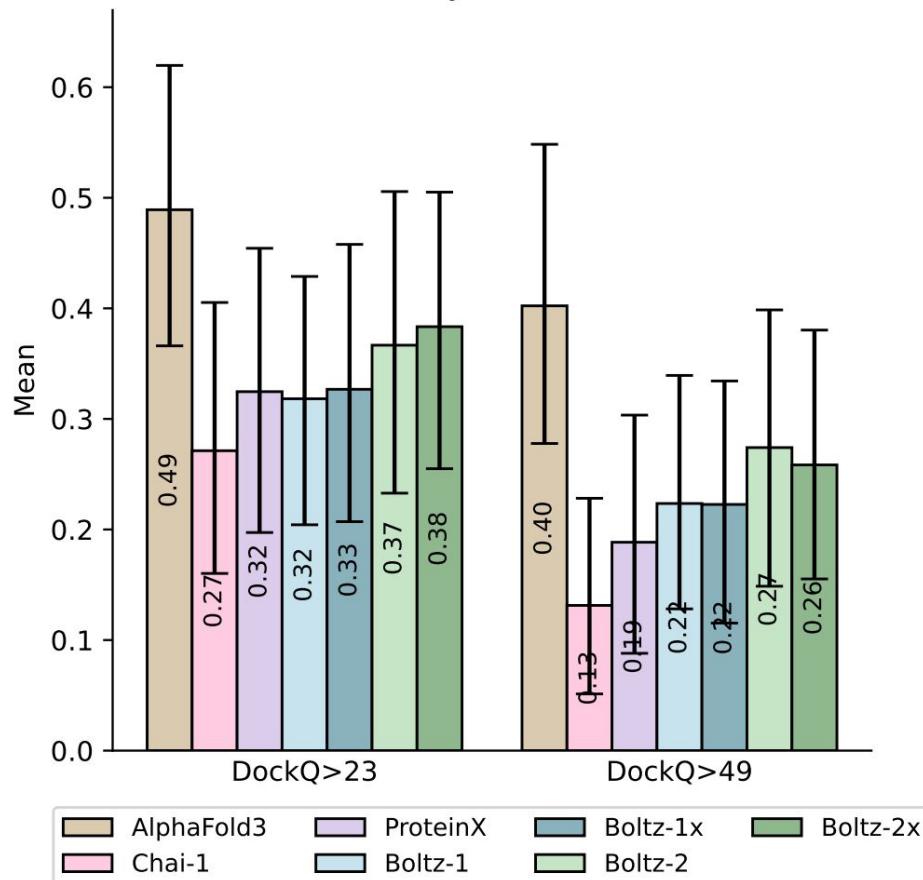
Chai-1 might outperform Boltz in antibody–antigen modeling, but the results are still very similar



Mean DockQ of interface predictions. Chai-1r takes 4 epitope residues as a restraint. AlphaFold3 is not available for commercial use, so we use Boltz-1, an AF3 reimplementations from Genesis Therapeutics and MIT. As Boltz-1 has not released a confidence model, we evaluate all methods based on average DockQ across 5 samples. We restrict to complexes that fit into memory of all models. Error bars are calculated with 10,000 bootstrap samples.

(A) **Performance on 165 Antibody-Antigen interfaces from 73 PDB IDs that were released after the training data cutoff of all models.** DockQ was computed for each interface, and then clustered by 40% Antigen sequence similarity ($n=60$ total clusters). (B) **Performance across 595 low-homology protein-protein interface clusters (40% and 100% sequence identity for proteins and peptides)** derived from 1273 interfaces across 665 PDBs which appeared after the Chai-1 training data cutoff. Performance of AF2.3 & Boltz-1 may be overestimated, as there is no guarantee these interfaces are low homology compared to their training splits.

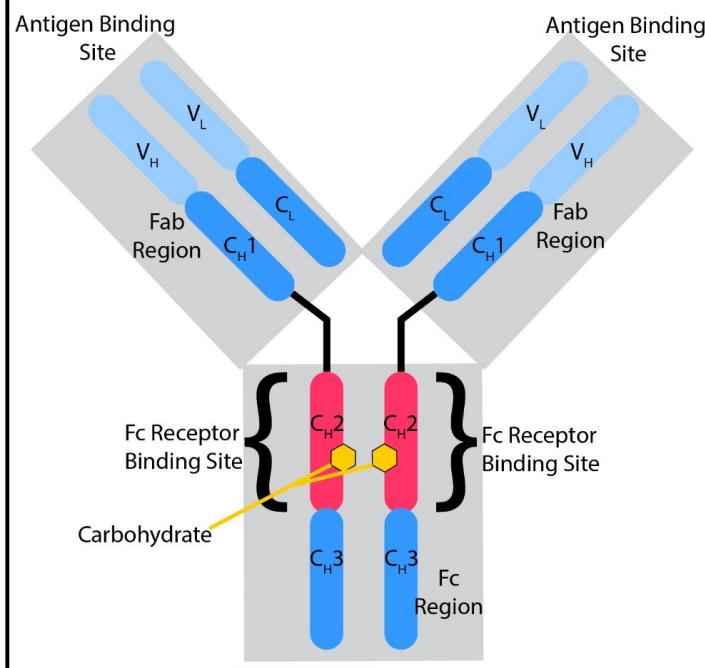
Antibody Benchmark



Antibodies have large and complex structure that is hard to investigate in silico

Figure 1: Antibody Structure

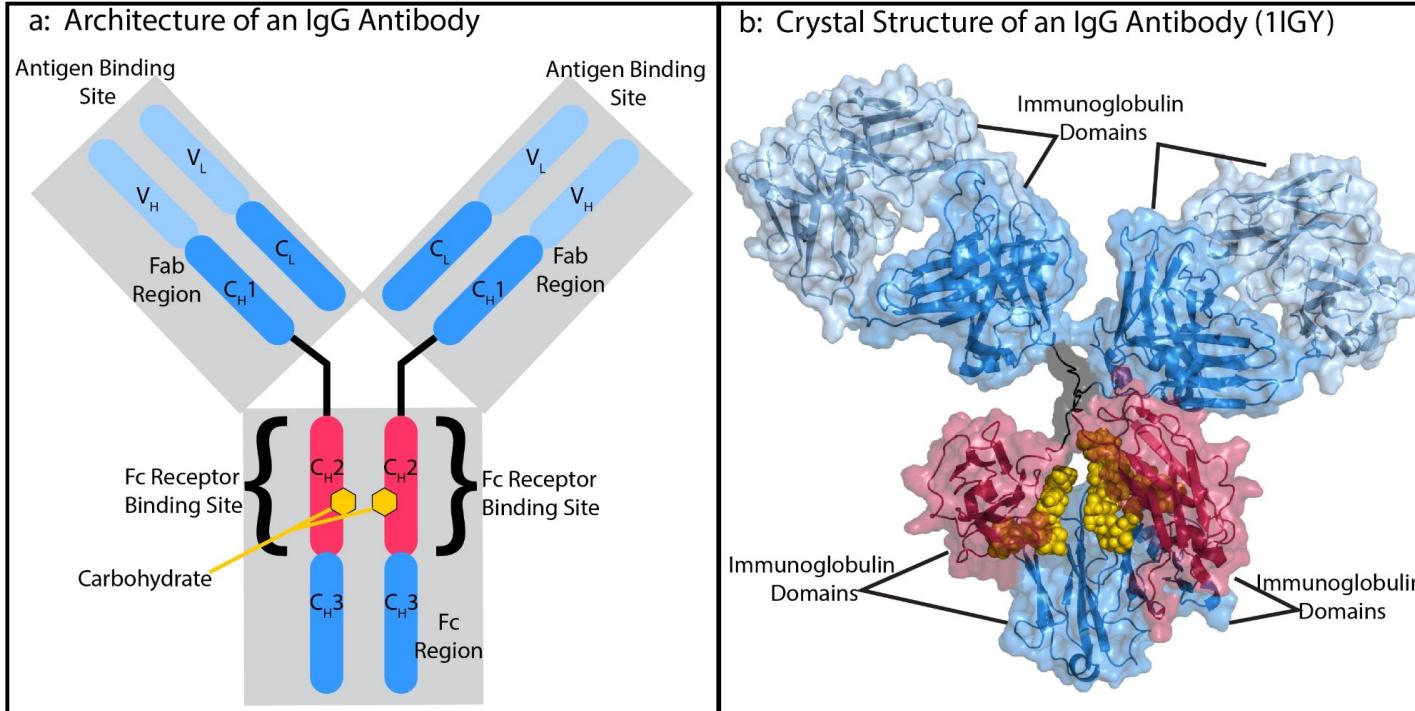
a: Architecture of an IgG Antibody



Antibody structure from a textbook

Antibodies have large and complex structure that is hard to investigate in silico

Figure 1: Antibody Structure



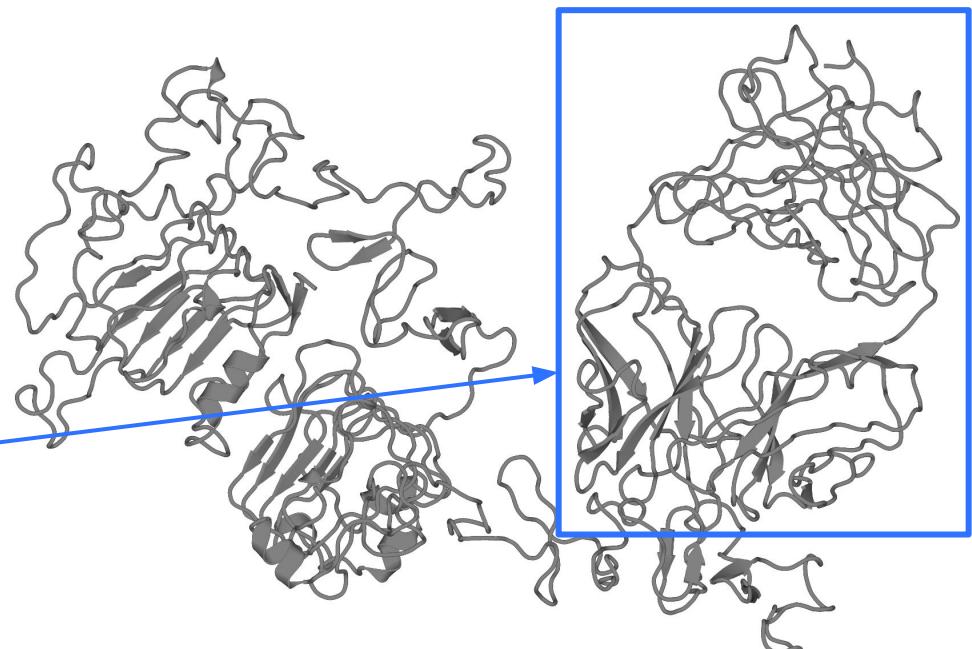
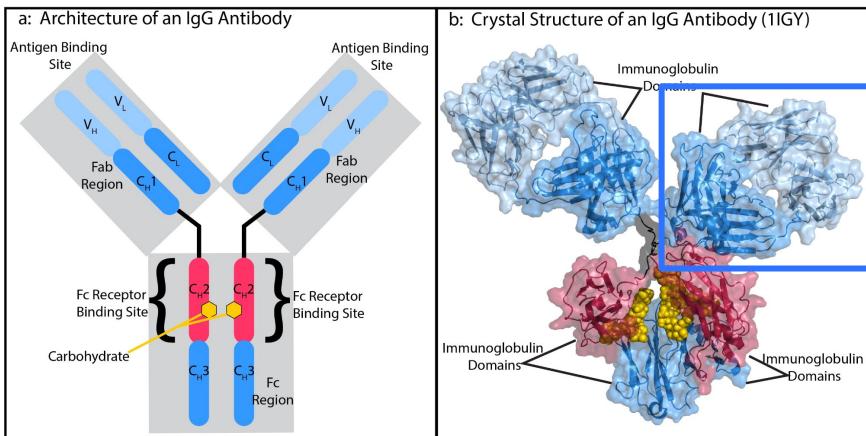
Antibody structure from a textbook

Real experimental antibody structure

When you add antigens the system becomes even larger and more complex

Antibody structure

- Antibody structure = β -strands + loops
- Antibody binding = loops
- **Paratope** (part of the **antibody** that binds to the **antigen**)
- **Epitope** (part of the **antigen** that binds to the **antibody**)

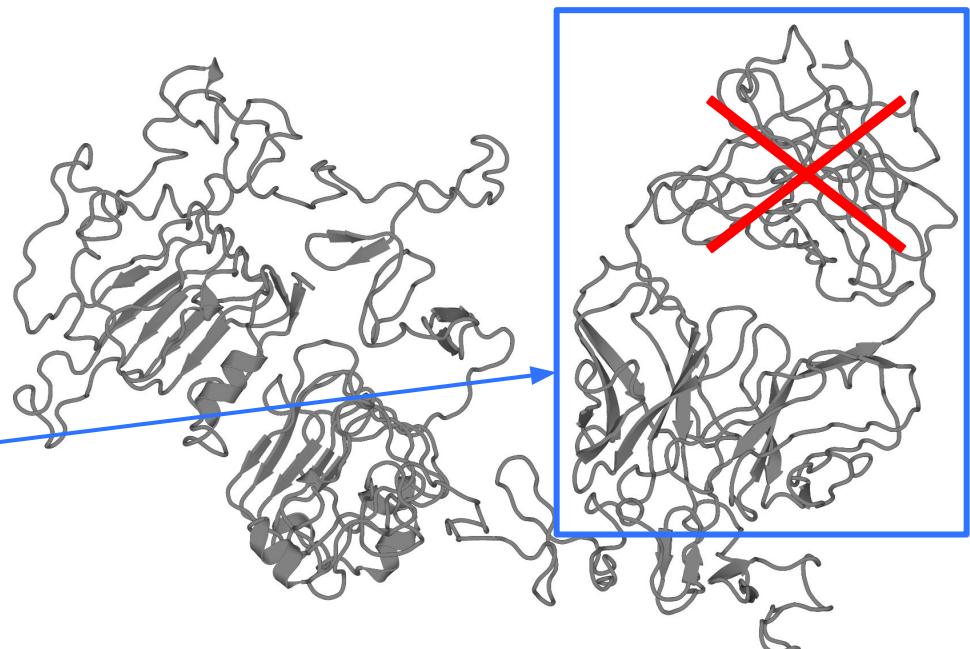
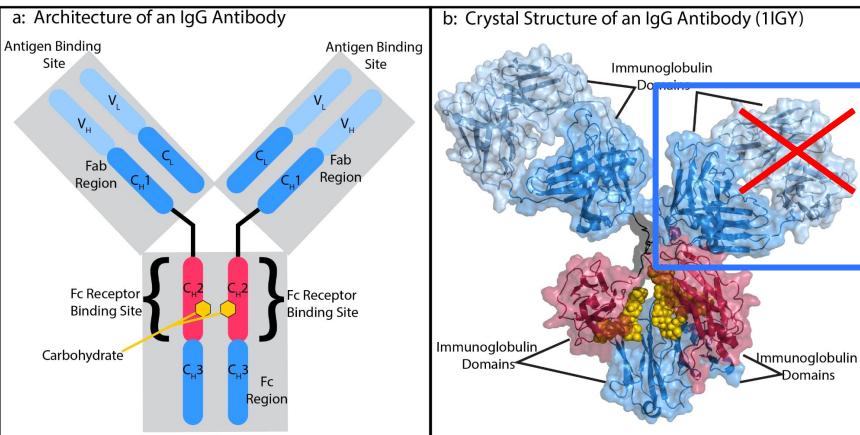


PDB ID: 6OGE

Current computational tools struggle to model the structure of the full antibody or antibody-antigen complex

Antibody structure

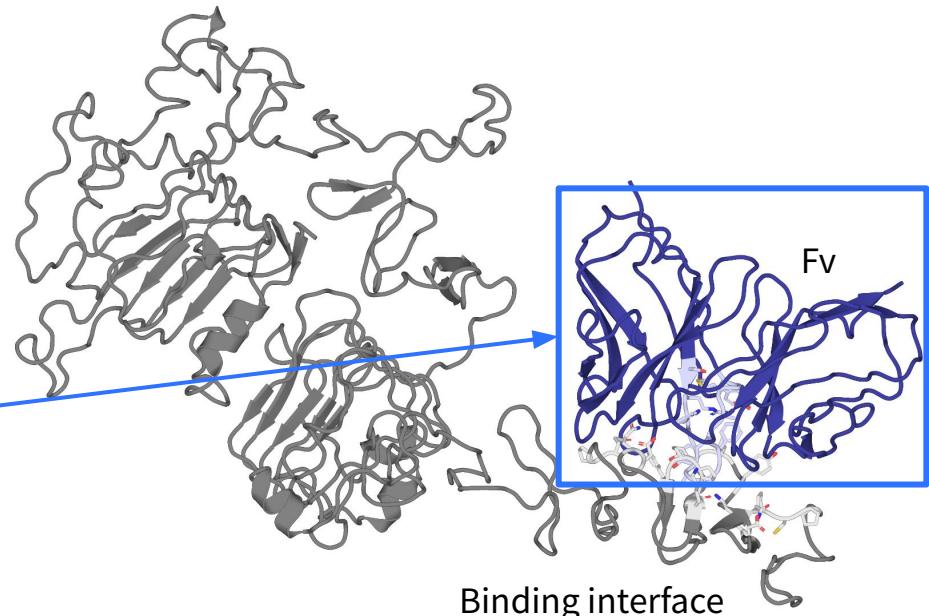
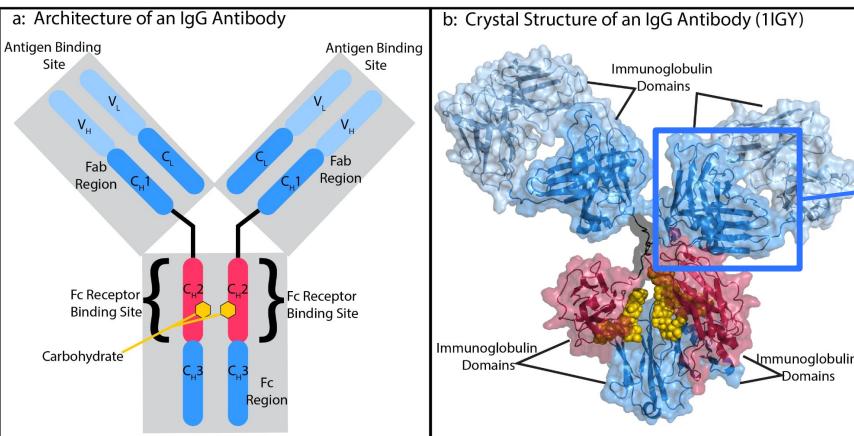
- Antibody structure = β -strands + loops
- Antibody binding = loops
- **Paratope** (part of the **antibody** that binds to the **antigen**)
- **Epitope** (part of the **antigen** that binds to the **antibody**)



Current computational tools struggle to model the structure of the full antibody or antibody-antigen complex

Antibody structure

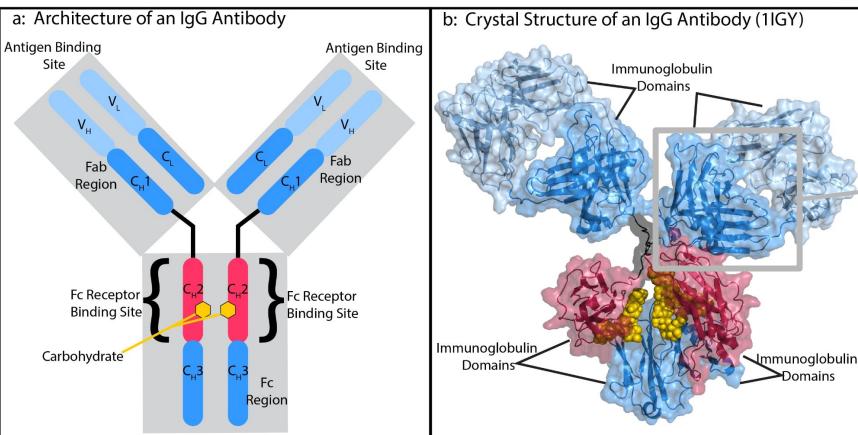
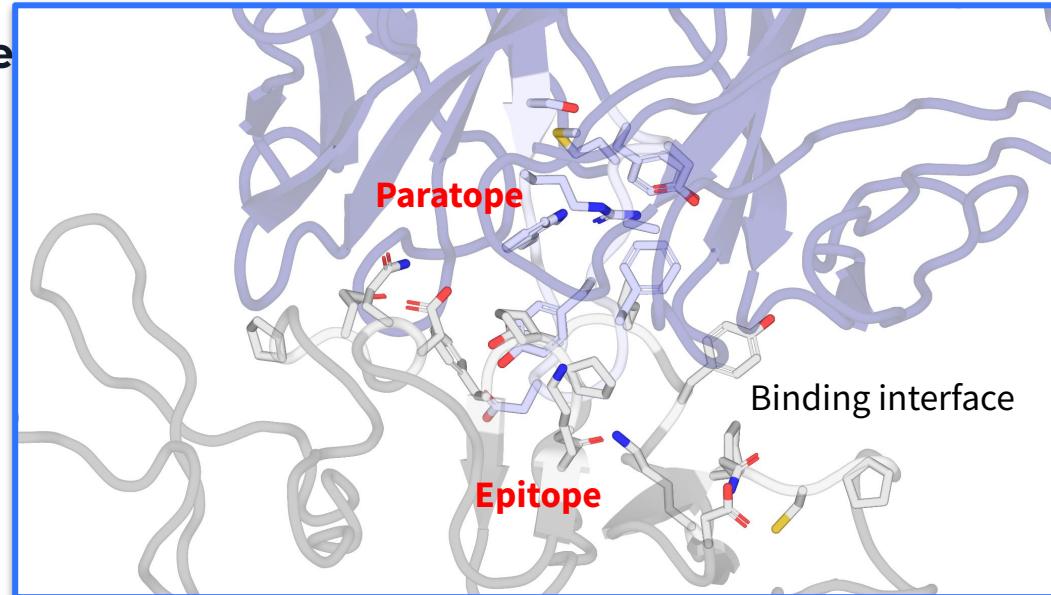
- Antibody structure = β -strands + loops
- Antibody binding = loops
- **Paratope** (part of the **antibody** that binds to the **antigen**)
- **Epitope** (part of the **antigen** that binds to the **antibody**)



Current computational tools struggle or antibody-antigen complex

Antibody structure

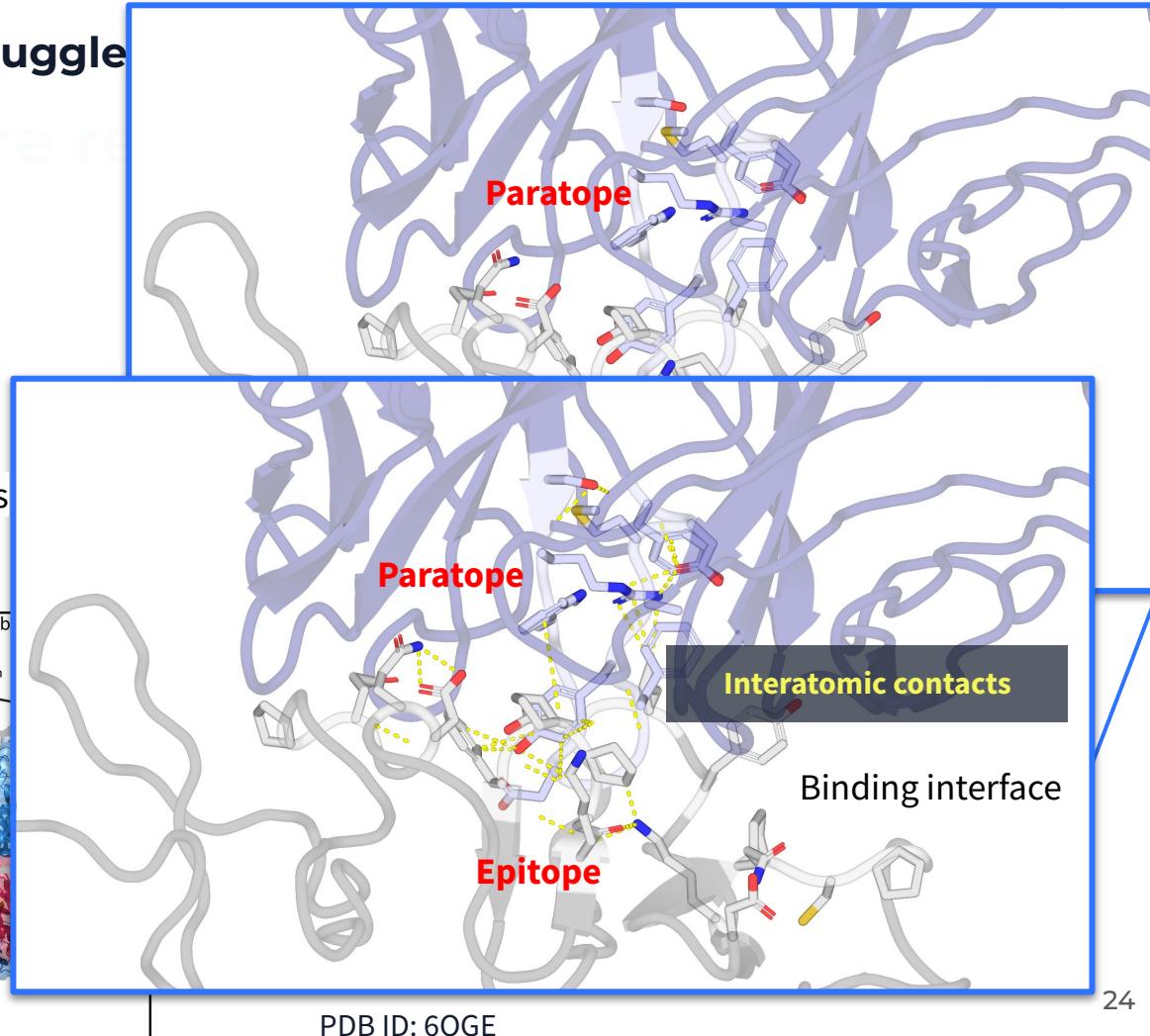
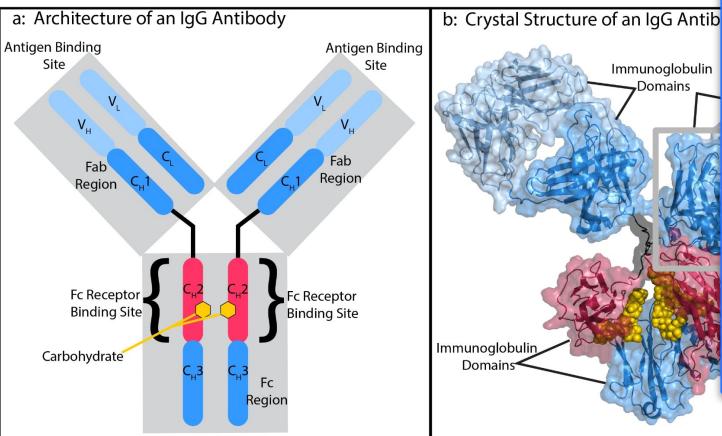
- Antibody structure = β -strands + loops
- Antibody binding = loops
- **Paratope** (part of the **antibody** that binds to the **antigen**)
- **Epitope** (part of the **antigen** that binds to the **antibody**)



Current computational tools struggle with antibody-antigen complex

Antibody structure

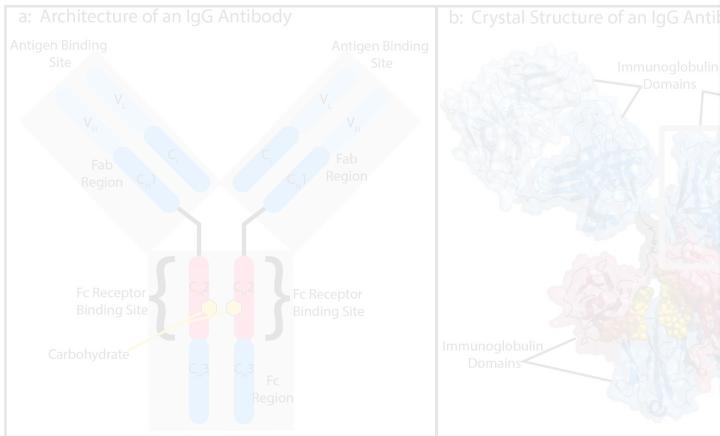
- Antibody structure = β -strands + loops
- Antibody binding = loops
- **Paratope** (part of the **antibody** that binds to the **antigen**)
- **Epitope** (part of the **antigen** that binds to the **antibody**)



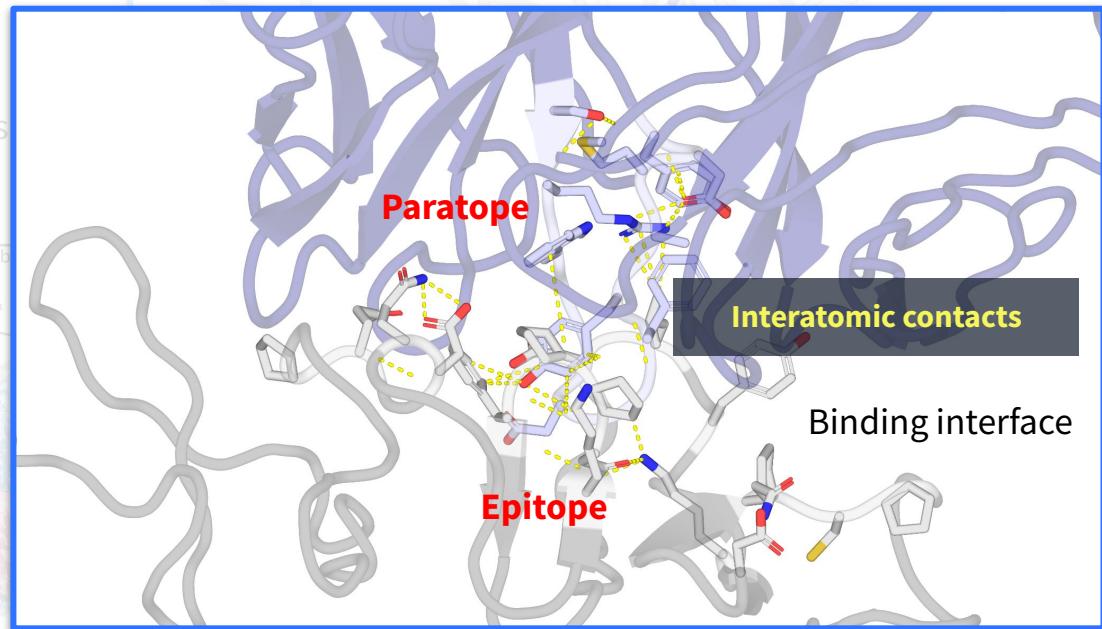
Current computational tools struggle with antibody-antigen complex

Antibody structure

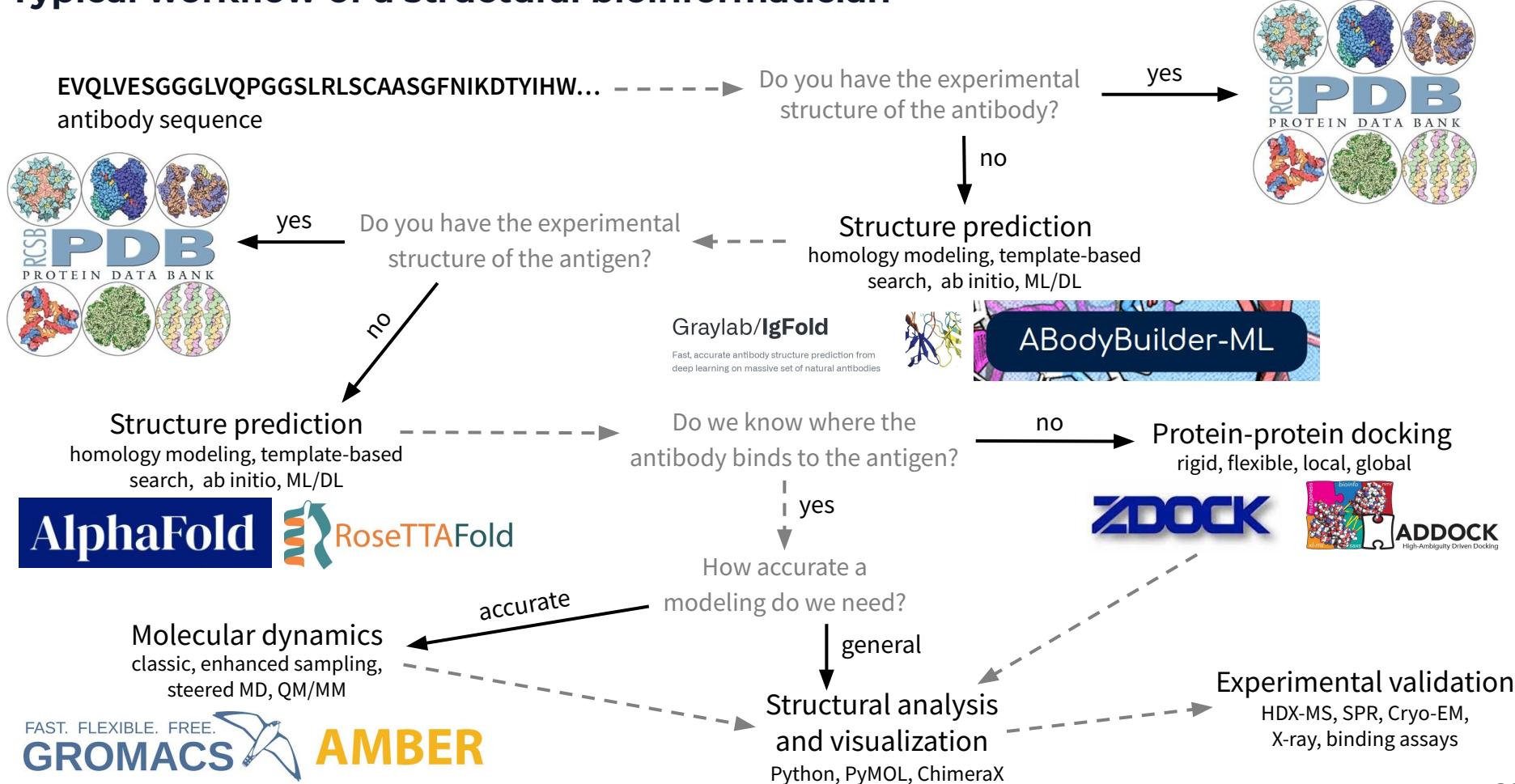
- Antibody structure = β -strands + loops
- Antibody binding = loops
- **Paratope** (part of the **antibody** that binds to the **antigen**)
- **Epitope** (part of the **antigen** that binds to the **antibody**)



Structural data can help in real world
applications like developability studies
or antibody-antigen binding prediction



Typical workflow of a structural bioinformatician



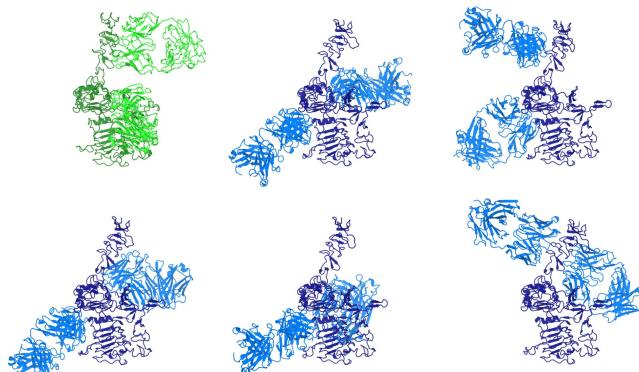
AlphaFold3 predictions antibody and antibody-antigen structures

Failure:

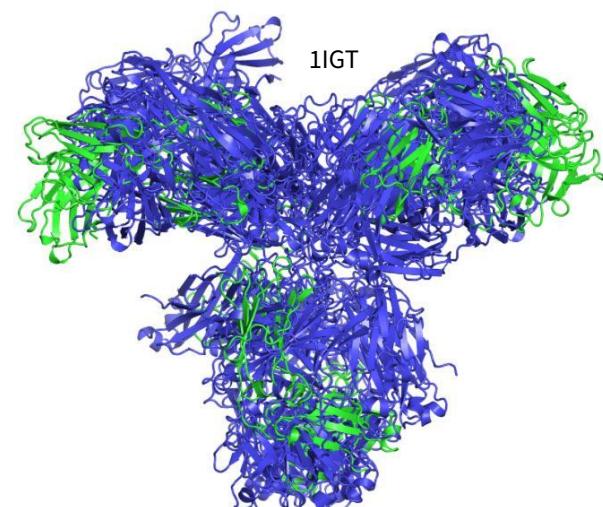
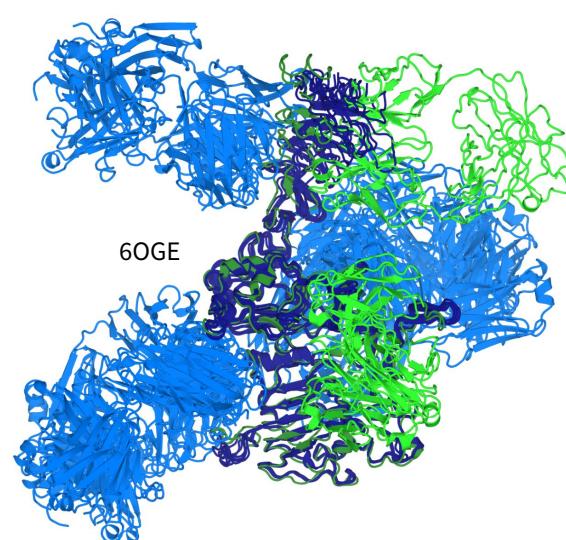
biomolecules with changing conformations or flexible and disordered regions whose structure depends a lot on the environment and external factors

- intrinsically disordered proteins
- antibodies and their complexes
- protein-switches
- membrane proteins

experiments → green (protein), gray (nucleic acid), AlphaFold3 → blue (protein), red (nucleic acid)

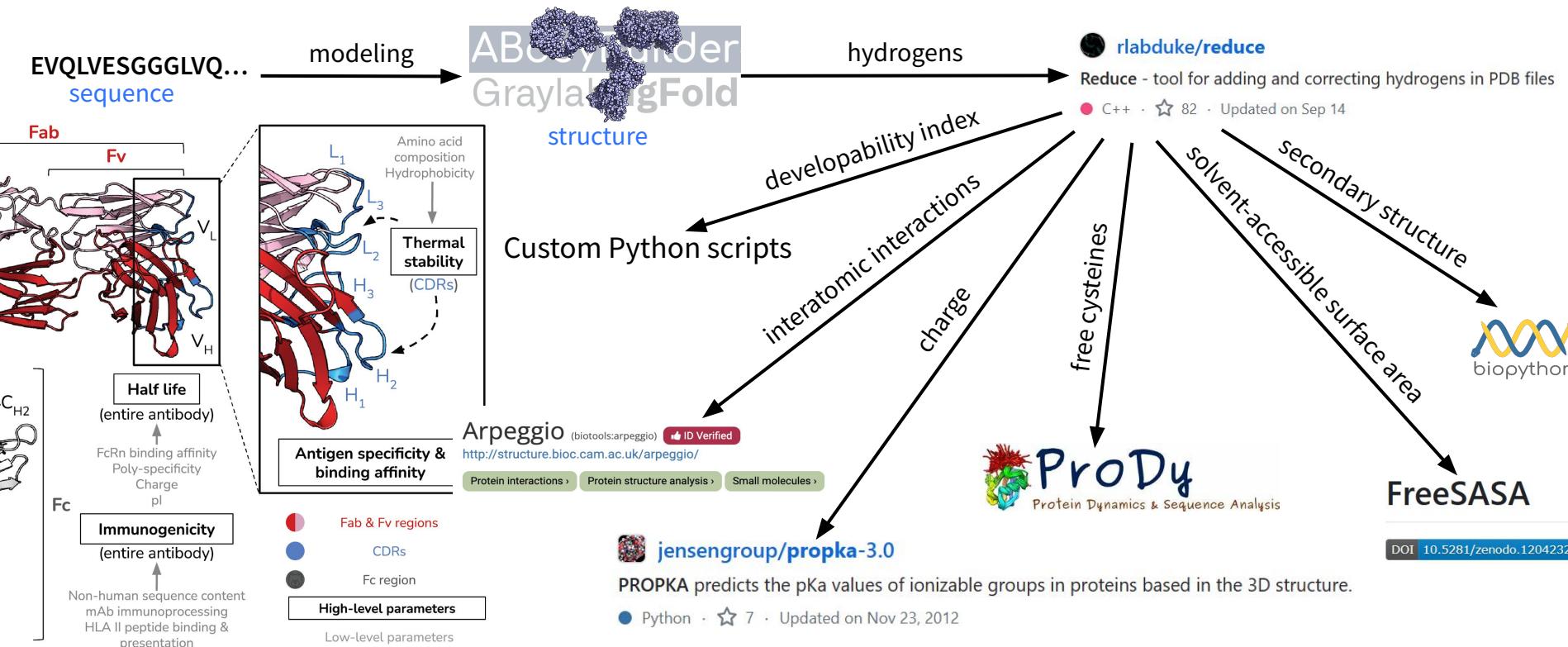


Antibody-antigen complex

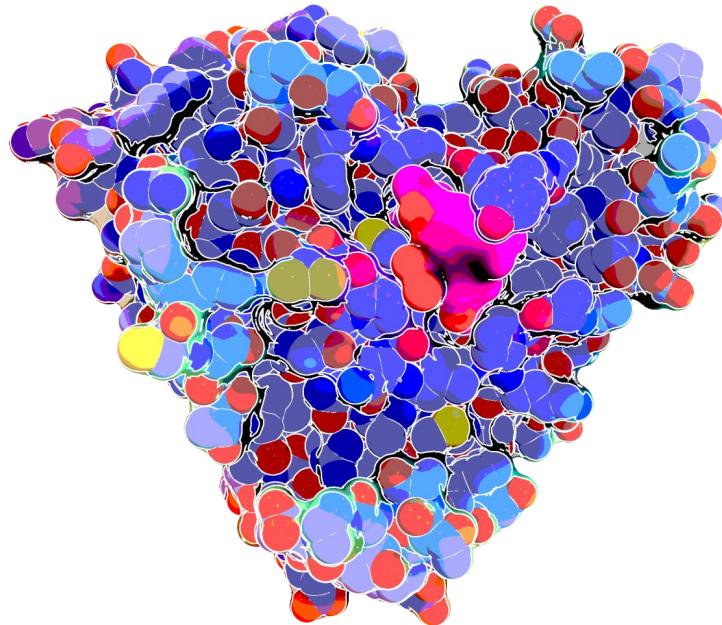


Antibody

Structure is not the end...



Thank you for your attention!



ribes.ev@gmail.com



eva-smorodina



eavsmorodina.com

Basic practice: antibody (1IGT), antibody-antigen (6OGE), VHH-antigen (7OLZ)

Trastuzumab (Herceptin) – monoclonal antibody against breast cancer (binds to HER2)

RCSB PDB Deposit Search Visualize Analyze Download Learn About Documentation Careers COVID-19 MyPDB

PROTEIN DATA BANK 208,702 Structures from the PDB 1,068,577 Computed Structure Models (CSM)

3D Structures Enter search term(s), Entry ID(s), or sequence Include CSM

Advanced Search | Browse Annotations

PDB-101 PDB EMDResource NAKB wwwPDB PDB-Dev

Structure Summary 3D View Annotations Experiment Sequence Genome Ligands Versions

Biological Assembly 1 1N8Z

3D View: Structure | 1D-3D View | Electron Density | Validation Report | Ligand Interaction

Global Symmetry: Asymmetric - C1
Global Stoichiometry: Hetero 3-mer - A1B1C1

Find Similar Assemblies

Biological assembly 1 assigned by authors.

Macromolecule Content

- Total Structure Weight: 114.42 kDa

This is version 1.5 of the entry. See complete history.

