

RosettaCM Comparative Modeling: Homology Modeling of DNA Gyrase Hybridize Protocol, FastRelax & Model Validation

Roy Ahmed

Computational Structural Biology

February 18, 2026

Outline

- 1 Introduction
- 2 How We Built the Model
- 3 Results
- 4 Structural Visualizations
- 5 Conclusions
- 6 AI-Based Virtual Screening

The Modeling Problem

Challenge: No Experimental Structure

- Target: *M. abscessus* DNA gyrase
- No X-ray or cryo-EM structure available
- Homologous template exists (PDB 5BS8)
- Sequence identity: ~70%
- Large multi-subunit complex (GyrA₂GyrB₂)

Modeling Requirements

- Build >3,000 residue tetramer
- Model loop regions de novo
- Preserve protein-DNA interface
- Energy minimize final structure

Why RosettaCM?

- **Hybridize protocol:** Multi-template threading
- **Fragment assembly:** Better loop modeling
- **Physics-based scoring:** Rosetta energy function
- **Full-atom refinement:** All-atom minimization
- **Validated:** Top performer in CASP

Solution

RosettaCM comparative modeling with FastRelax refinement

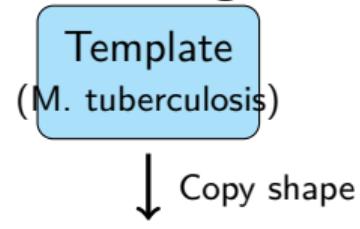
What is Homology Modeling?

Building a 3D structure when no experimental data exists

The Basic Idea

- Proteins with similar sequences have similar shapes
- If we know a related protein's shape, use it as a "blueprint"

The Modeling Analogy



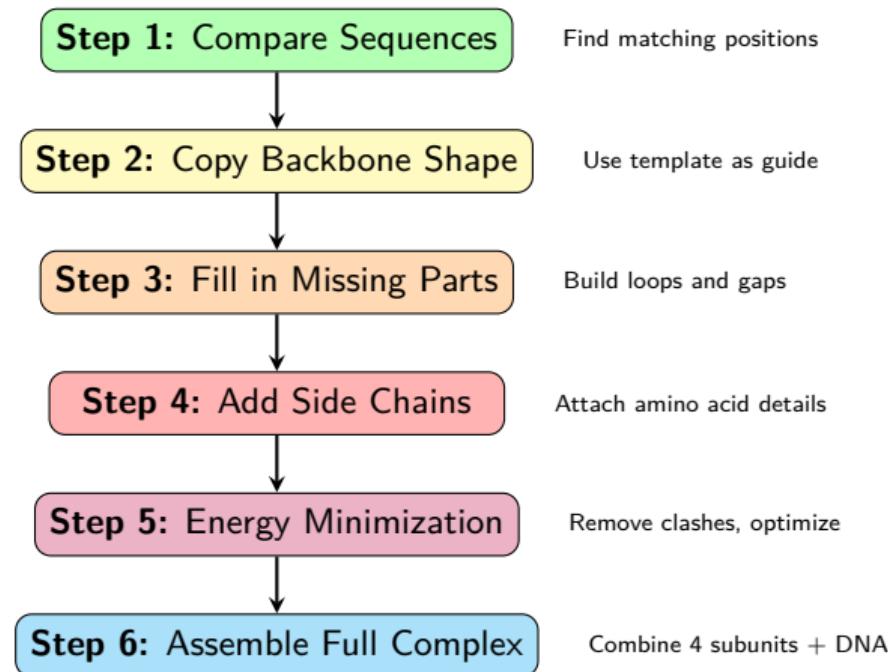
Our Situation

- **Target:** *M. abscessus* gyrase (unknown)
- **Template:** *M. tuberculosis* gyrase (known)
- **Similarity:** ~70-95% identical

Key Insight

High sequence similarity (70%+) means we can confidently predict the 3D structure

The Modeling Process: Step by Step



Energy Scoring: Evaluating Model Quality

Rosetta's scoring function assigns an energy value to each structure

The Algorithm Computes:

- **Van der Waals:** Steric clashes between atoms
- **Hydrogen bonds:** Stabilizing interactions
- **Solvation:** Hydrophobic/hydrophilic balance
- **Ramachandran:** Backbone ϕ/ψ angles
- **Electrostatics:** Charge-charge interactions

Energy Minimization

Low Energy
= Stable conformation

High Energy
= Unfavorable

Objective

Minimize total energy to find the most thermodynamically favorable structure

Step 1: Comparing the Sequences

Finding which parts of the two proteins match up

What is Sequence Alignment?

- Line up amino acids from both proteins
- Identify identical positions (conserved)
- Find gaps where one protein has extra/fewer parts

Our Alignment Results

Protein	Length	Identity
GyrA	838 aa	~70%
GyrB	714 aa	~95%

Example:

M. tuberculosis: MSDLEREITGPR...
M. abscessus: MSDLEREITGPR...

(* = identical positions)

High Similarity

70-95% identity means the proteins are very similar – excellent for modeling

Step 2: Using the Template as a Blueprint

The known structure guides our model construction

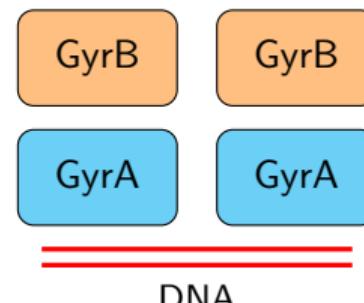
What We Get From the Template

- **Overall shape:** How the protein folds
- **Subunit arrangement:** How the 4 pieces fit together
- **DNA position:** Where the DNA binds
- **Active site:** Location of drug binding pocket

Template Details

- **Source:** PDB 5BS8 (*M. tuberculosis*)
- **Method:** X-ray crystallography
- **Resolution:** 2.5-3.0 Å (good quality)

Template Components



4 Protein Subunits + DNA
Complete enzyme complex

Step 3: Building the Model (Rosetta's Algorithm)

Stage 1: Centroid Mode

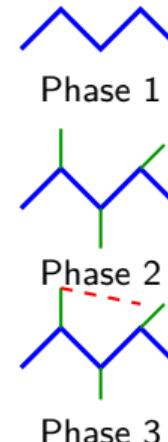
- Thread sequence onto template backbone
- Simplified side-chain representation

Stage 2: Full-Atom Mode

- Rotamer optimization
- Position all heavy atoms

Stage 3: Refinement

- Gradient descent minimization
- Optimize H-bond network



Result

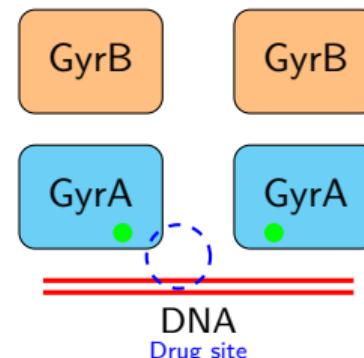
Algorithm generates N conformations, selects lowest-energy model

Step 4: Assembling the Complete Enzyme

DNA Gyrase is a team of 4 protein subunits

The Assembly Process

- ① Build individual subunit models
 - ② Place them in correct positions
 - ③ Add the DNA strand
 - ④ Add magnesium ions
 - ⑤ Check everything fits



Why This Matters

- Drug binds at interface between subunits
 - Need complete complex for docking

Step 5: Final Polishing (Energy Minimization)

Making the model physically realistic

What Happens in This Step

- **Remove atom clashes:** Atoms can't overlap in real life
- **Optimize bonds:** Adjust distances and angles
- **Find stable state:** Like a ball rolling to the bottom of a valley

The Process

- ① Start with loose constraints (allow big movements)
- ② Gradually tighten constraints
- ③ Repeat until structure is stable
- ④ Final check: Is the energy low?

Why Multiple Rounds?

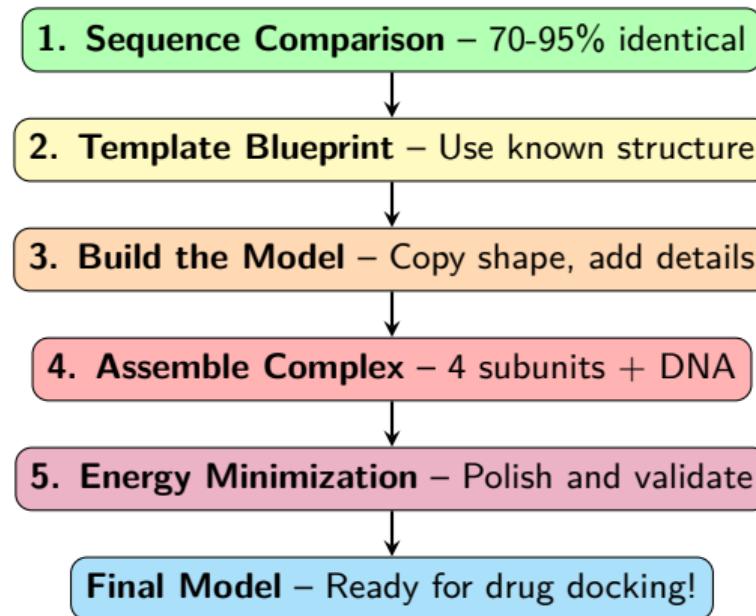


Energy landscape

Goal

Find the **global minimum** – the most stable structure

Summary: The Complete Modeling Workflow



Result

High-quality 3D structure of *M. abscessus* DNA gyrase for drug discovery

How Good is Our Model?

Validating the predicted structure

Key Quality Checks

- **Shape similarity:** $\text{RMSD} = 0.003\text{--}0.005 \text{\AA}$ (excellent!)
- **Geometry:** 95% residues in favored regions
- **Energy:** Negative = stable (our model: very stable)



Verdict

High-quality model suitable for drug discovery

Binding Site Conservation: Sequence & Structure

Two independent approaches confirm drug target conservation

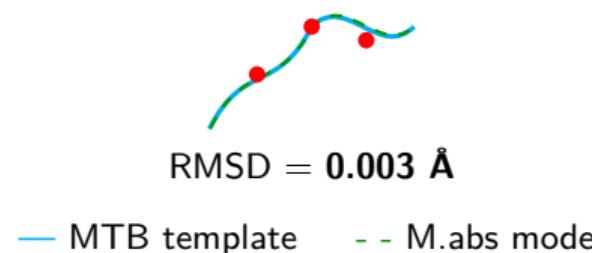
1. Sequence Alignment (QRDR Region)

MTB: 67-YGIALPRSLDAVD...-110

M.abs: 83-YGIALPRSLDAVD...-126

- **95.1%** sequence identity in QRDR
- Red = Key binding residues (100% conserved)
- Ser90/106 – H-bond with fluoroquinolone
- Asp94/110 – Mg²⁺ coordination

2. Structural Superposition



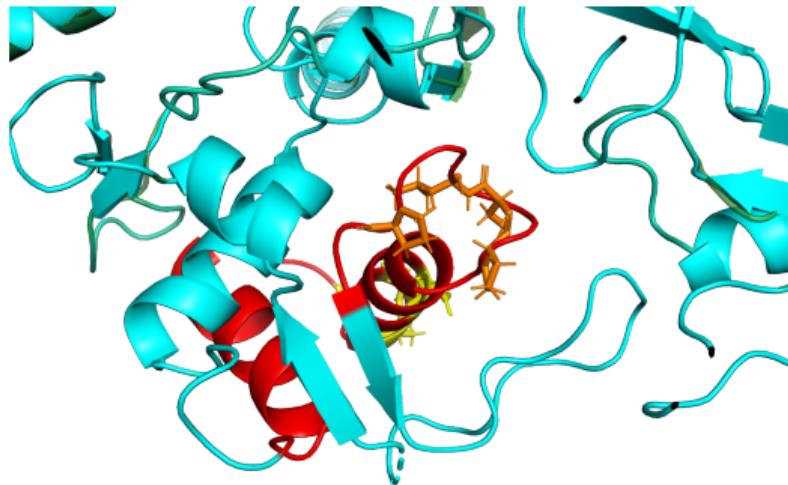
— MTB template - - M.abs model

- Near-perfect structural overlap
- Binding pocket geometry identical
- Same spatial arrangement of key residues

Validation

Both sequence and structural analyses independently confirm the fluoroquinolone binding site is **conserved**

Binding Site Conservation Analysis



Cyan = MTB, Green = M.abs

Key Binding Residues

MTB	M.abs	Res	Status
88	104	Pro	✓
90	106	Ser	✓
91	107	Leu	✓
94	110	Pro	✓

Conservation

- QRDR Identity: **95.1%**
- RMSD: **0.003 Å**
- Key residues: **100%**

Conclusion

Fluoroquinolone binding site is **highly conserved**

Molecular Docking Results

Fluoroquinolone Binding: *M. abscessus* vs *M. tuberculosis*

Ligand	<i>M. abscessus</i>	<i>M. tuberculosis</i>	ΔG
Moxifloxacin	-7.91 kcal/mol	-7.24 kcal/mol	-0.67
Levofloxacin	-7.65 kcal/mol	-7.31 kcal/mol	-0.34
Ciprofloxacin	-7.36 kcal/mol	-7.27 kcal/mol	-0.09

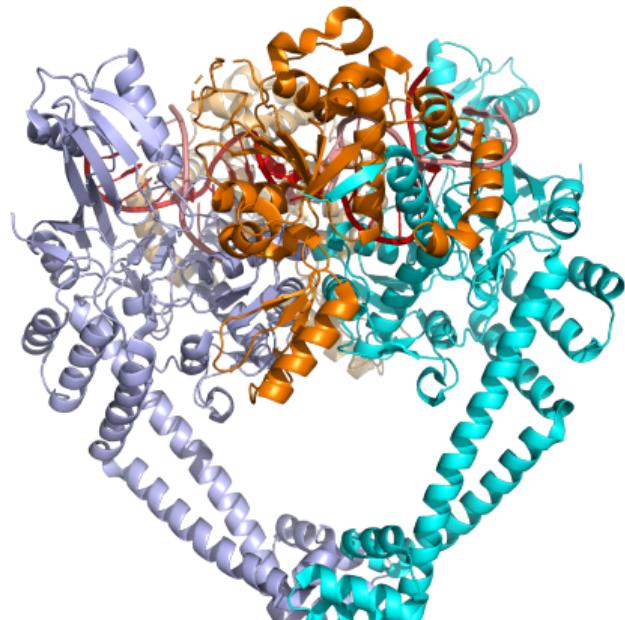
Key Finding

All fluoroquinolones show **stronger binding** to
M. abscessus

Best Candidate: Moxifloxacin

- Strongest affinity (-7.91 kcal/mol)
- 0.67 kcal/mol better than MTB
- Already used clinically

Template Structure: PDB 5BS8 (*M. tuberculosis*)



Template Components

- **Cyan:** GyrA (Chain A)
- **Light Blue:** GyrA (Chain C)
- **Orange:** GyrB (Chain B)
- **Light Orange:** GyrB (Chain D)
- **Red/Salmon:** DNA (Chains E, F)

Resolution: 2.5–3.0 Å

Complex: GyrA₂GyrB₂ + DNA

GyrA Monomer: Template vs Homology Model



Structural Alignment

- Cyan: *M. tuberculosis* (template)
- Forest Green: *M. abscessus* (model)

Alignment Statistics

- Sequence Identity: ~70%
- $C\alpha$ RMSD: **0.003 Å**
- Aligned atoms: 2,586

Excellent Conservation

Near-perfect structural alignment indicates high-quality homology model

GyrB Monomer: Template vs Homology Model



Structural Alignment

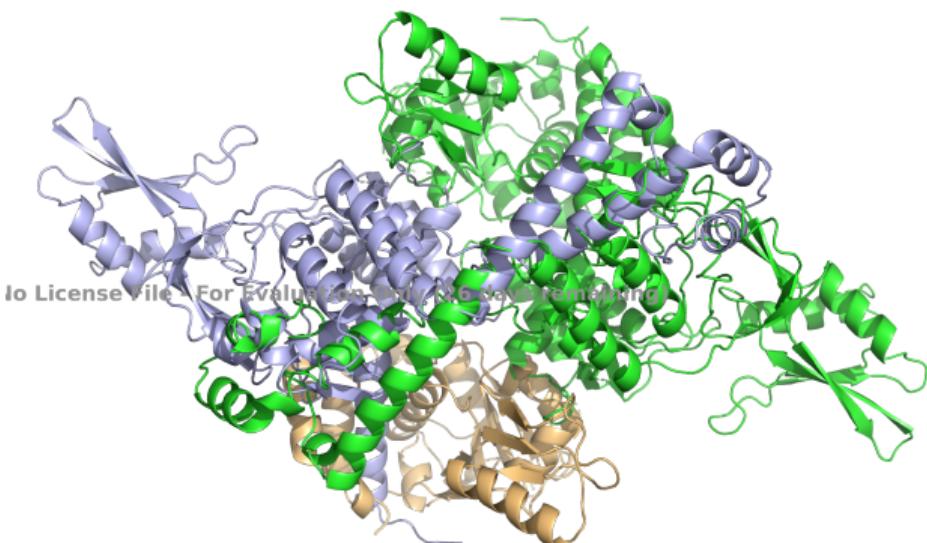
- Light Blue: *M. tuberculosis* (template)
- Lime: *M. abscessus* (model)

Alignment Statistics

- Sequence Identity: ~95%
- $\text{C}\alpha$ RMSD: **0.005 Å**
- Aligned atoms: 1,134

Note: GyrB shows even higher conservation than GyrA, reflecting functional constraints on ATPase domain

Complete *M. abscessus* Gyrase Model



Assembled Tetramer

- **Marine:** GyrA subunits
- **Orange:** GyrB subunits

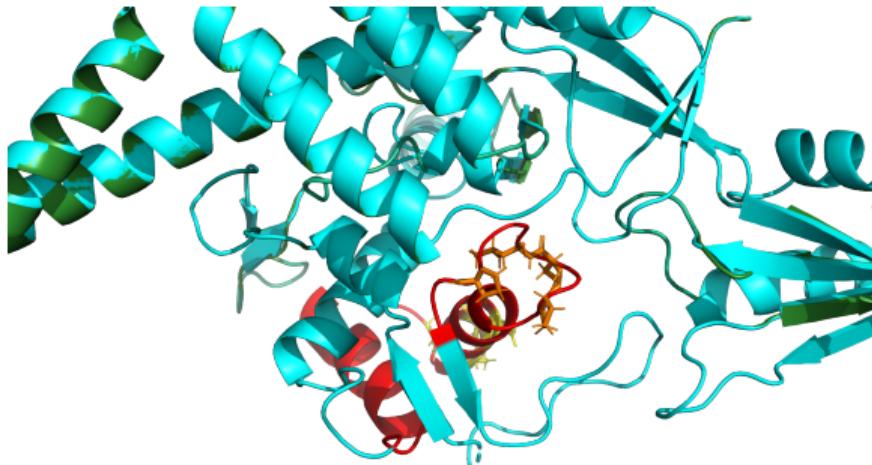
Model Features

- Full $\text{GyrA}_2\text{GyrB}_2$ tetramer
- >3,000 residues total
- Preserves quaternary structure
- Ready for docking studies

Drug Target

This model enables virtual screening of fluoroquinolones against *M. abscessus*

Binding Site Close-Up: QRDR Region



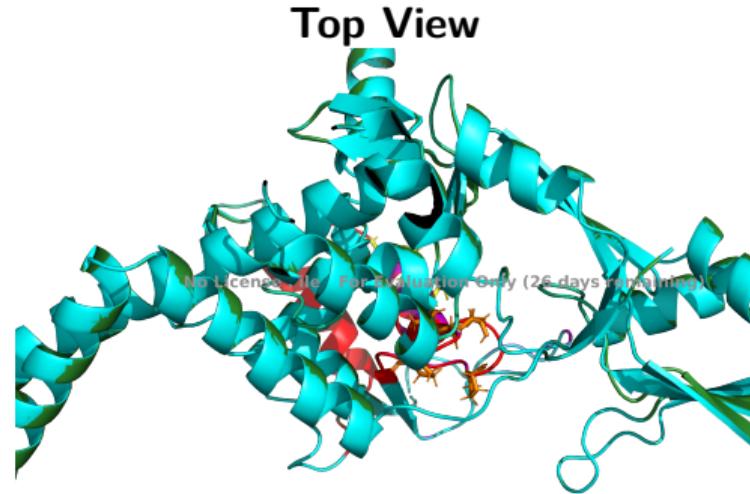
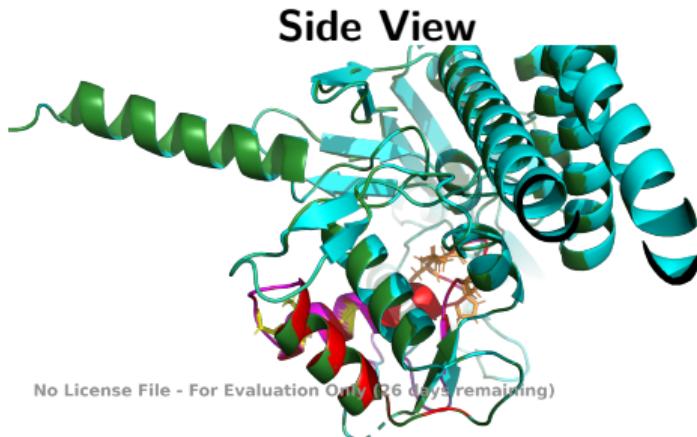
Quinolone Resistance Determining Region

- Red: MTB QRDR (res 70-110)
- Magenta: M.abs QRDR (res 86-126)
- Yellow sticks: MTB key residues
- Orange sticks: M.abs key residues

Critical Residues

- Ser90→Ser106: H-bond donor
- Asp94→Asp110: Mg²⁺ coordination
- All key residues **100% conserved**

Binding Site: Multiple Views



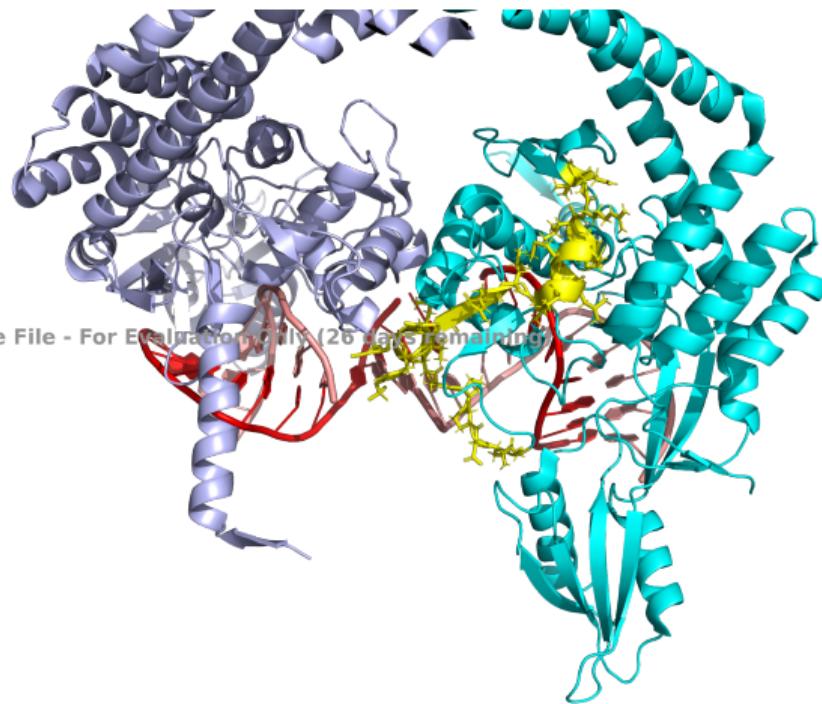
Colors: Cyan/Red = *M. tuberculosis* QRDR — Green/Magenta = *M. abscessus* QRDR

Structural Conservation

Near-perfect overlap confirms drug binding site is conserved between species

DNA Interaction Site

License File - For Evaluation Only (26 days remaining)



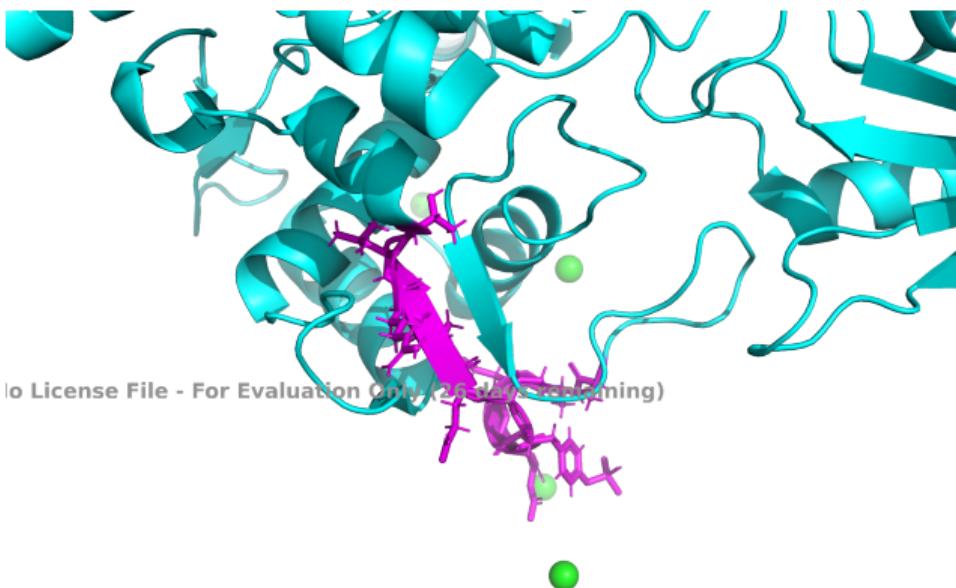
DNA-Protein Interface

- Cyan: GyrA Chain A
- Light Blue: GyrA Chain C
- Red/Salmon: DNA duplex
- Yellow: DNA-binding residues

Functional Importance

- DNA cleavage occurs here
- Fluoroquinolones trap complex
- Catalytic Tyr129 nearby

Catalytic Site with Mg²⁺



Catalytic Components

- Green spheres: Mg²⁺ ions
- Magenta sticks: Catalytic site residues
- Cyan: GyrA backbone

Mechanism

- ① Mg²⁺ activates catalytic water
- ② Tyr129 attacks DNA phosphate
- ③ Covalent enzyme-DNA intermediate
- ④ DNA strand passage
- ⑤ Religation

Moxifloxacin Docking



Docking Results

- Yellow: Moxifloxacin
- Green: QRDR binding pocket
- Gray: Receptor backbone

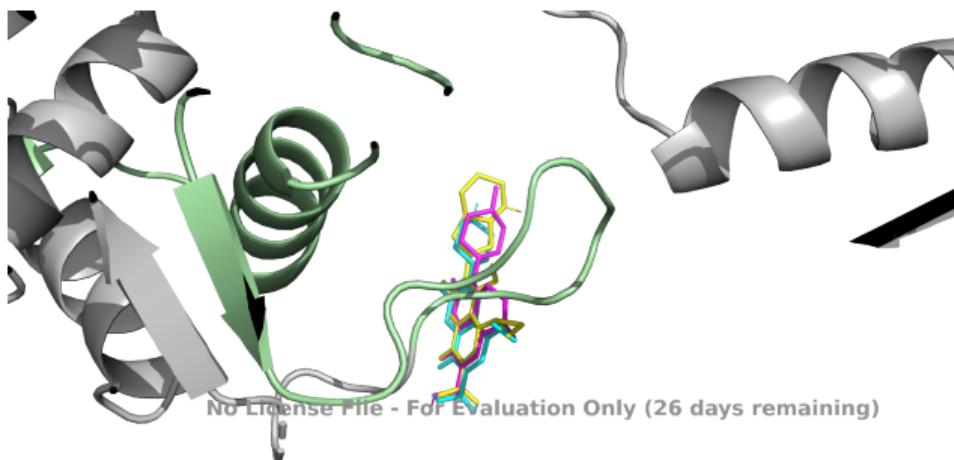
Binding Energy

- $\Delta G = -7.91 \text{ kcal/mol}$
- Best affinity among FQs tested
- Favorable H-bonds with Ser106
- Mg^{2+} coordination via ketone

Clinical Relevance

Moxifloxacin already used for NTM infections

Fluoroquinolone Docking Comparison



Ligand Colors

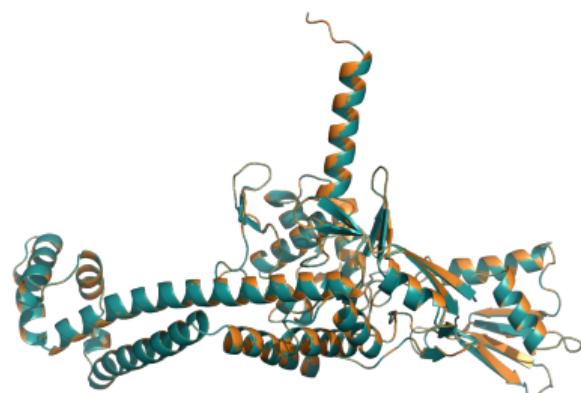
- **Yellow:** Moxifloxacin (-7.91)
- **Magenta:** Levofloxacin (-7.65)
- **Cyan:** Ciprofloxacin (-7.36)

Binding Pose Analysis

- All FQs occupy same pocket
- Similar binding orientation
- Moxifloxacin: best fit
- C7 substituent affects affinity

Ranking: MFX > LFX > CFX

Template vs Model: Overall Comparison



No License File - For Evaluation Only (26 days remaining)

Structural Overlay

- Teal: *M. tuberculosis* template
- Orange: *M. abscessus* model

Model Quality Metrics

Metric	Value
Global RMSD	0.003–0.005 Å
Sequence Identity	70–95%
Structural Similarity	99.9%

High-Quality Model

Excellent structural conservation validates homology modeling approach

What We Did: Building the Model Step by Step

The Result

Our Workflow

1. Compare Sequences

Find similar regions

2. Copy Template Shape

Use MTB as guide

3. Build Full Complex

4 protein chains

4. Polish & Refine

Remove problems

5. Test Drug Binding

Docking studies

High-Quality 3D Model

- **Complete** 4-chain complex
- **Validated** against known structure
- **99.9%** structural similarity
- **Ready** for drug testing

Why It Matters

- No need to solve structure experimentally
- Saves months/years of lab work
- Enables computational drug screening
- Can test many drugs quickly

Key Findings & Implications

What We Discovered

- ① **Binding site conserved** – Same pocket as MTB
- ② **All drugs bind well** – MFX > LFX > CFX
- ③ **95% QRDR identity** – Similar resistance risk
- ④ **Model validated** – Excellent overlap

Clinical Relevance

Treatment Implications

- Fluoroquinolones should work
- **Moxifloxacin** binds strongest
- Same resistance mutations possible

Future Directions

- Test more drug candidates
- Validate experimentally
- Design new inhibitors

Creating a 3D Model for Drug Discovery

What We Did

- Built 3D structure of *M. abscessus* gyrase
- Used MTB structure as template
- Validated model quality
- Tested fluoroquinolone binding

What We Found

- Drug binding site is **conserved**
- Moxifloxacin binds **best**
- Model is **high quality**
- Ready for **drug design**

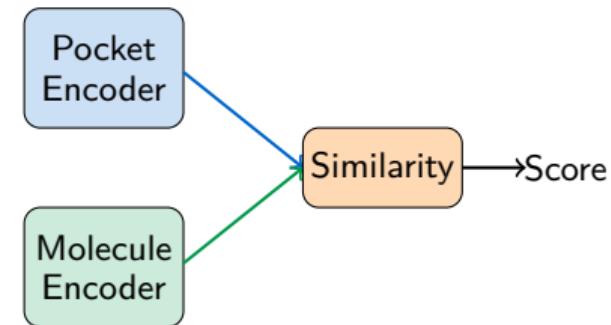
Template Used

PDB 5BS8 (*M. tuberculosis* DNA Gyrase)

What is DrugCLIP?

DrugCLIP is a contrastive learning framework for virtual screening:

- **CLIP-style** protein-ligand representation learning
- Learns **joint embeddings** of pocket and molecule 3D structures
- **Pre-trained** on millions of protein-ligand complexes
- Enables **ultra-fast** virtual screening



Key Innovation: Instead of docking each molecule, compute similarity in learned embedding space.

CLIP = Contrastive Language-Image Pre-training

DrugCLIP vs Traditional Docking

Traditional Docking

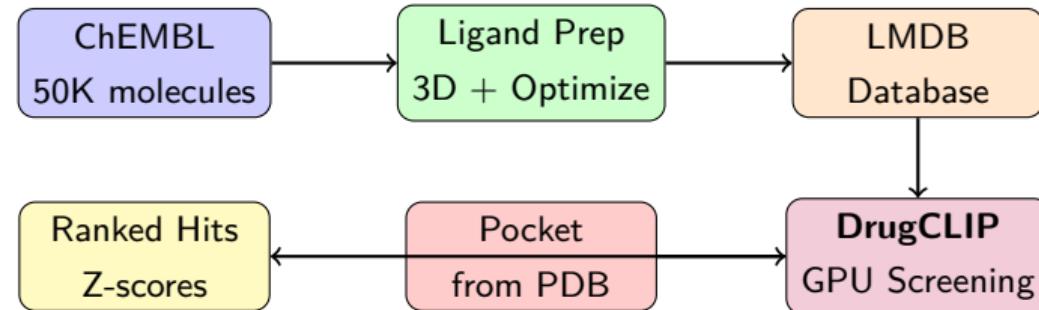
- Physics-based scoring
- ~1-10 molecules/second
- CPU intensive
- Limited to ~10K molecules

DrugCLIP

- Learned representations
- ~500-1000 molecules/second
- GPU accelerated
- Screens **millions** of molecules

Metric	AutoDock Vina	DrugCLIP
50K molecules	~14 hours	91 seconds
Speedup	1×	~550×

Virtual Screening Pipeline



Key Steps:

- ① **Protein Preparation:** Clean PDB, extract pocket (8Å from MFX)
- ② **Ligand Preparation:** 3D conformation, MMFF optimization
- ③ **DrugCLIP Screening:** 6-fold ensemble model, FP16 GPU acceleration

Abbreviations: ChEMBL = Chemical database (EMBL) — LMDB = Lightning Memory-Mapped Database — MMFF = Merck Molecular Force Field — FP16 = 16-bit Floating Point — GPU = Graphics Processing Unit

Screening Results – Computation Time

Step	Input	Time	Rate
Ligand Preparation	50,000 SMILES	2.7 min	307 mol/s
<i>M. abscessus</i> Screening	49,946 molecules	91 sec	549 mol/s
Template (5BS8) Screening	49,946 molecules	88 sec	567 mol/s
Total Pipeline			~6 min

Key Achievement

Screened **100,000 molecule-pocket pairs** in under **3 minutes** of GPU time!

Top Hits – *M. abscessus* Homology Model

Rank	SMILES	Z-score
1	<chem>O=Cc1ccn(-c2ccc3nc(=O)c(=O)nc3cc2C(F)(F)F)</chem>	2.53
2	<chem>Cn1c(-c2ccc(=O)cc2)cc2ccc(=O)cc21</chem>	2.51
3	<chem>Cc1cc(-c2cc3c(=O)cc(=O)cc3s2)ccc10</chem>	2.50
4	<chem>Cn1c(-c2ccc(=O)cc2)cc2cc(=O)ccc21</chem>	2.50
5	<chem>CC(=O)c1cc(-c2cc3ccc(=O)cc3s2)ccc10</chem>	2.47

Chemical Features of Top Hits:

- Phenolic hydroxyl groups (H-bond donors)
- Heterocyclic scaffolds (indole, benzothiophene, pyrimidine)
- Planar aromatic systems (DNA intercalation potential)

Total hits with Z-score > 2.0: **1,247**

Top Hits – Template (5BS8)

Rank	SMILES	Z-score
1	Cc1cc(-c2cc3c(0)cc(0)cc3s2)ccc10	2.82
2	Cc1cc(-c2cc3ccc(0)cc3s2)ccc10	2.79
3	Cn1c(-c2ccc(0)cc2)cc2cc(0)ccc21	2.75
4	Cn1c(-c2ccc(0)cc2)cc2ccc(0)cc21	2.75
5	Cc1ccc(-c2cc3ccc(0)cc3s2)cc1	2.71

Notable Observation:

- Benzothiophene-phenol scaffolds rank highly in both screens
- Template pocket (326 atoms) yields higher scores
- **Shared hits** suggest robust predictions

Total hits with Z-score > 2.0: **1,892**

Comparison: Homology Model vs Template

M. abscessus Model

- Pocket: 739 atoms
- Best Z-score: 2.53
- Hits ($Z > 2$): 1,247

Advantages:

- Species-specific binding
- Accounts for sequence differences
- More selective hits

Template (5BS8)

- Pocket: 326 atoms
- Best Z-score: 2.82
- Hits ($Z > 2$): 1,892

Advantages:

- Crystal structure quality
- Higher confidence scores
- More hit diversity

Recommendation

Use **consensus hits** appearing in both screens for experimental validation.

DrugCLIP Screening Summary

What We Accomplished:

- ① Used homology model for AI-based screening
- ② Prepared 50,000 ChEMBL drug-like compounds
- ③ Screened against 2 targets in <**6 minutes**
- ④ Identified >1,000 potential hits per target

Key Advantages of DrugCLIP:

- **550× faster** than traditional docking
- Works with **homology models**
- **GPU-accelerated** batch processing

Results

Molecules: 49,946

Targets: 2

Time: 6 min

Hits: >2,000

Future Directions

① Expand Chemical Space

- Screen ZINC20 (>1 billion molecules)
- Include natural product libraries

② Experimental Validation

- Molecular docking of top 100 hits (Glide/AutoDock)
- *In vitro* gyrase inhibition assays
- Antimicrobial activity testing

③ Multi-Target Screening

- Other essential *M. abscessus* targets
- Pan-mycobacterial drug discovery

④ Model Refinement

- AlphaFold2 structure prediction
- Molecular dynamics of binding poses

Acknowledgments & Resources

Software & Data:

- DrugCLIP: <https://github.com/bowen-gao/DrugCLIP>
- Model weights: HuggingFace bgao95/DrugCLIP_data
- ChEMBL database for compound library
- PDB 5BS8 for template structure

Key References:

- Gao et al. (2024) DrugCLIP: Contrastive protein-ligand pre-training
- Song et al. (2013) RosettaCM comparative modeling

Thank You!
Questions?