

BactProNET: A Structural-Mechanistic Platform for Deciphering Target-Mediated Antimicrobial Resistance

Author information

Ke Wang^{1, a}, Yiru Liu^{1, a}, Xue Wang¹, Haitao Zhou¹, Longfei Chen^{2, *}, Jingyun Xu^{2, *}

¹School of Clinical Medicine, Wannan Medical University, Wuhu 241002, China

²School of Basic Medical Sciences, Wannan Medical University, Wuhu 241002, China

^aThese authors contributed equally to this work

*Correspondence: Jingyun Xu (xujingyun@wnmc.edu.cn) and Longfei Chen (chenlf2019@zju.edu.cn)

Running title: BactProNET: Antibiotic-Target Interaction Database

Address: <http://www.wnmc-bioinfo.com/BactProNET/>

This file contains:

Supplementary Methods (Manual Curation Protocol, Steps 1-9)

Supplementary Figure S1. Alignment of AlphaFold2-predicted GyrA model with experimental PDB structures (4Z2C and 5CDQ)

Supplementary Figure S2. Mapping of known resistance mutations onto the optimal binding model of GyrA

Supplementary Table S1. Comparison of BactProNET with existing antimicrobial resistance and structural databases

Supplementary Table S2. Summary of experimentally validated resistance mutations (Excel, provided separately)

Supplementary Table S3. Alignment quality metrics for mutation mapping to *E. coli* homologous sites

Supplementary Table S4. AlphaFold2 model confidence metrics (pLDDT and PAE) for predicted target proteins (Excel, provided separately)

Supplementary Table S5. Molecular docking parameters and AlphaFold2 model validation

Supplementary Table S6. Software version information

Supplementary Table S7. Summary of $\Delta\Delta G$ values for resistance mutations across different antibiotic classes

Supplementary Table S8. Spatial distances between known resistance mutations and contact residues

Supplementary Data S1. GyrA entropy scores (Excel, provided separately)

Supplementary Data S2. GyrA conservation scores (Excel, provided separately)

Supplementary Data S3. GyrB entropy scores (Excel, provided separately)

Supplementary Data S4. GyrB conservation scores (Excel, provided separately)

Supplementary Data S5. ParC entropy scores (Excel, provided separately)

Supplementary Data S6. ParC conservation scores (Excel, provided separately)

Supplementary Methods

Manual Curation Protocol

Step 1. Identification of protein families from antibiotics

We first collected 323 antibiotics across 12 major classes. For each antibiotic, we queried the ChEMBL database for its "Drug Mechanism" and "Target Type" fields to identify the protein family targeted by that antibiotic. Examples include:

- Quinolones → DNA gyrase (GyrA/GyrB) and topoisomerase IV (ParC/ParE)
- β -lactams → penicillin binding proteins (PBPs) and β -lactamases
- Sulfonamides → dihydropteroate synthase (DHPS)
- Trimethoprim → dihydrofolate reductase (DHFR)
- Rifamycins → DNA-directed RNA polymerase
- Glycopeptides → peptidoglycan synthesis enzymes
- Macrolides → 50S ribosomal proteins
- Aminoglycosides → 30S ribosomal proteins
- Tetracyclines → 30S ribosomal proteins

Through this process, we obtained a set of protein families associated with antibiotic resistance.

Step 2. Literature search and identification of bacterial protein targets

For each protein family identified in Step 1, we searched PubMed using the following template:

`([Protein Name[Title/Abstract]]) AND (resistance[Title/Abstract]) AND (mutation[Title/Abstract])`

No language or publication date restrictions were applied. The search was performed iteratively for each protein family name and its synonyms. Initial retrieval yielded 3,247 articles (before deduplication).

Step 3. Deduplication

All retrieved records were pooled and deduplicated using EndNote X9 (automatic

duplicate detection based on title, author, year, and journal), followed by manual verification of ambiguous cases. After deduplication, 1,856 articles remained.

Step 4. Inclusion and exclusion criteria

Inclusion criteria:

- Original research article (excluding reviews, editorials, conference abstracts, commentaries)
- Reports one or more specific amino acid mutations in a bacterial protein target
- Provides experimental evidence linking the mutation(s) to resistance
- Involves bacterial species (any genus)

Exclusion criteria:

- Reports only nucleotide mutations without confirmed amino acid changes
- Describes resistance solely via gene expression changes, efflux pumps, or horizontal gene transfer without specific mutations
- Lacks sufficient experimental detail to extract mutation information
- In vitro evolution studies without clinical or reference strain validation (retained only if the mutation was reported as naturally occurring in clinical isolates)

Step 5. Screening workflow

A team of three independent curators performed screening in two phases:

Phase 1 (title/abstract screening): Each article was independently reviewed by two curators against the inclusion/exclusion criteria. Articles flagged by either curator proceeded to full-text review. After title/abstract screening, 687 articles remained.

Phase 2 (full-text screening): The same two curators independently reviewed the full text. Disagreements were resolved by consensus discussion involving the third curator. After full-text screening, 330 articles remained.

Step 6. Data extraction

For each included article, curators manually extracted the following fields:

- Bacterial species

- Protein target (standardised name)
- Amino acid mutation (wild type → variant)
- Associated antibiotic(s)

Through this extraction process, we compiled 107 unique bacterium–protein combinations (e.g., *Escherichia coli* GyrA, *Staphylococcus aureus* GyrA, *Klebsiella pneumoniae* ParC).

Step 7. Definition of "key resistance mutation"

A mutation was classified as a "key resistance mutation" if the original authors explicitly reported that the mutation confers resistance. Mutations reported only as "polymorphisms" or "silent mutations" without evidence of resistance were excluded.

Step 8. Handling of conflicting reports

When multiple publications reported conflicting resistance conclusions for the same mutation, we applied the following hierarchy:

Priority 1: Studies using isogenic strain backgrounds (same species, same genetic background)

Priority 2: Most recent publication (by year)

Priority 3: If the conflict could not be resolved, the mutation was flagged as "ambiguous" in the database, excluded from the primary analysis, but retained in a separate supplementary table.

Step 9. Final dataset

Through this curation process, starting from 3,247 initial articles and after deduplication and two rounds of screening, we obtained 330 articles. From these, we extracted 243 unique amino acid substitutions across 196 unique mutation sites on 107 unique bacterium–protein combinations. All curated data are accessible via the BactProNET platform.

Supplementary Figure S1. Alignment of AlphaFold2-predicted GyrA model with experimental PDB structures (4Z2C and 5CDQ).

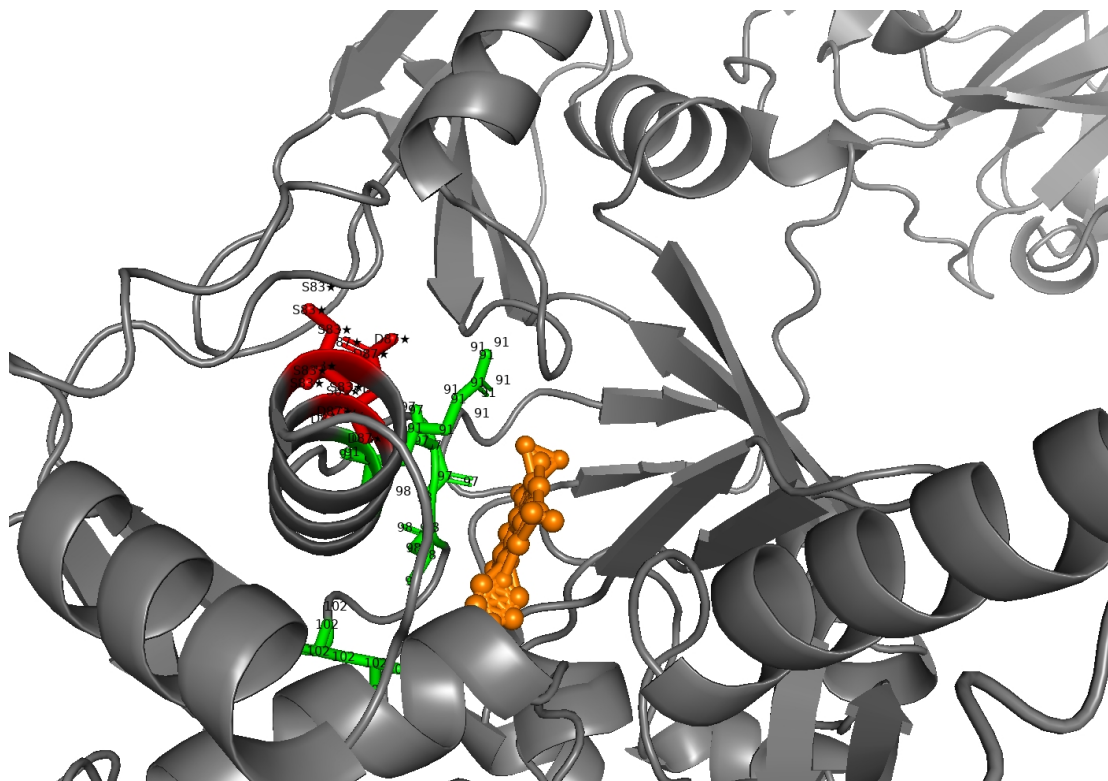


Color coding: Red: Moxifloxacin bound in PDB 4Z2C (experimental ternary complex); Green: Moxifloxacin bound in PDB 5CDQ (experimental ternary complex); Blue: Moxifloxacin docked in AlphaFold2-predicted GyrA model (DNA-free).

The predicted protein backbone aligned well with the protein components of both experimental structures (not colored for clarity). RMSD values are 0.377 Å (2,812 atoms aligned) for 5CDQ and 0.770 Å (2,655 atoms aligned) for 4Z2C, confirming that the AlphaFold2 model accurately captures the overall structure of DNA gyrase.

The binding positions of moxifloxacin differ substantially between the experimental structures and our model, with centroid distances of 78.0 Å (4Z2C) and 88.9 Å (5CDQ). This discrepancy reflects the conformational change in the binding pocket induced by DNA binding, which is a known feature of the fluoroquinolone mechanism of action, rather than an error in our docking methodology.

Supplementary Figure S2. Mapping of known resistance mutations onto the optimal binding model of GyrA.



The optimal binding model obtained from molecular docking shows the ligand (orange sticks) bound within the fluoroquinolone resistance-determining region (QRDR) of GyrA. Green stick residues (ARG91, SER97, LEU98, LEU102) represent contact residues located within 4.5 Å of the ligand. Red stick residues (S83, D87) indicate known resistance mutation sites reported in *Glaesserella parasuis* GyrA. Spatial distance analysis showed that D87 is located 6.1 Å from the contact residue ARG91, while S83 is located 12.1–17.6 Å from contact residues. These observations support the reliability of the model in identifying key resistance-associated residues.

Supplementary Table S1. Comparison of BactProNET with existing antimicrobial resistance and structural databases

Feature	BactProNET	CARD	ResFinder	MEGARes	NDARO	PDB	UniProt
Primary focus	Target-protein amino acid substitutions	All resistance mechanisms	Acquired resistance genes	AMR genes (clinical + environmental)	Pathogen + AMR data	Protein 3D structures	Protein sequences and annotations
Antibiotic information (structure, MoA, indication)	Yes (12 classes, 323 agents)	Yes (18 classes; drug count not specified)	Yes (drug count not specified)	No	No	No	No
Bacterial species information	Yes (44 species)	Yes (414 pathogens)	No	No	Yes (753 species, 97 taxa)	No	No
Target protein 3D structures	Yes (AlphaFold2, 107 targets)	No	No	No	No	Yes (X-ray, NMR, cryo-EM; 251,806 structures)	No
Resistance mutation sites / unique substitutions	Yes (196 sites / 243 mutations)	Yes (4,480 SNPs, all mechanisms)	No	No	Yes (AMRFinderPlus, 15,000 records)	No	Not systematically tracked
Wild-type vs. mutant docking (ΔG , $\Delta\Delta G$, hydrogen bond values)	Yes (640 WT + 1,697 mutant models)	No	No	No	No	No	No
Conservation score (Shannon entropy)	Yes	No	No	No	No	No	No
Phylogenetic tree + MSA	Yes	No	No	No	No	No	No
Built-in BLAST + MSA tools for user sequences	Yes (both)	Yes (BLAST via RGI)	Yes (BLAST only)	No	Yes (BLAST only)	Yes (BLAST only)	Yes (BLAST only)
Data source	Literature (330 publications) + NCBI, UniProt, PubChem, ChEMBL, CARD, DrugBank	Literature + NCBI (3,354 publications)	Literature + GenBank	CARD, ResFinder, PointFinder, AMRFinderPlus, BacMet	NCBI RefSeq	Experimental structures	Literature + computational (572,970 entries)

Note: MoA, mechanism of action; WT, wild-type; MSA, multiple sequence alignment; BLAST, Basic Local Alignment Search Tool. CARD, ResFinder, MEGARes, and NDAO cover diverse resistance mechanisms (efflux pumps, enzymatic inactivation, regulatory mutations). BactProNET focuses specifically on target-protein amino acid substitutions. Direct quantitative comparison is not meaningful due to scope differences.

Supplementary Table S3. Alignment quality metrics for mutation mapping to *E. coli* homologous sites

Protein	Retained Regions (amino acid positions, <i>E. coli</i> numbering)	Gap-free
GyrA (23 regions)	76 – 192, 613 – 689, 691 – 699, 704 – 746, 750 – 793, 795 – 910, 949 – 1026, 1028 – 1125, 1127 – 1129, 1132 – 1143, 1174 – 1177, 1191 – 1224, 1226 – 1242, 1279 – 1294, 1333 – 1337, 1339 – 1351, 1384 – 1392, 1148 – 1484, 1486 – 1489, 1492 – 1504, 1506 – 1541, 1545 – 1556, 1560 – 1562	Yes
GyrB (10 regions)	14 – 42, 44 – 186, 190 – 221, 225 – 493, 495 – 555, 557 – 561, 563 – 571, 732 – 735, 745 – 814, 816 – 819	Yes
ParC (25 regions)	14 – 125, 127 – 230, 232 – 260, 266 – 269, 273 – 314, 317 – 344, 347 – 354, 356 – 418, 423 – 548, 552 – 593, 596 – 604, 606 – 620, 627 – 648, 650 – 662, 666 – 679, 682 – 684, 686 – 703, 705 – 713, 715 – 720, 727 – 730, 733 – 741, 743 – 762, 765 – 768, 770 – 775, 778 – 790	Yes

*Note: All mutation positions fall within well-aligned, gap-free columns, supporting accurate mapping of resistance mutations to *E. coli* homologous sites.*

Supplementary Table S5. Molecular docking parameters and AlphaFold2 model validation.

Category	Parameter	Setting / Value
Docking	Software	AutoDock 4.2.6
	Grid size	320 × 320 × 320 grid points
	Grid spacing	0.375 Å
	Actual box size	120 Å × 120 Å × 120 Å
	Grid center	Geometric center of each protein
	GA runs	10
	Population size	150
	Maximum energy evaluations	2,500,000
	Maximum generations	27,000
	Clustering RMSD tolerance	2.0 Å
	Ligand flexibility	Fully flexible (up to 16 rotatable bonds)
	Receptor flexibility	Rigid
Validation	Protein RMSD (vs PDB 4Z2C)	0.770 Å
	Protein RMSD (vs PDB 5CDQ)	0.377 Å

Note: The grid box (120 Å × 120 Å × 120 Å) is centered at the geometric center of each protein, covering the entire protein volume. Docking was therefore performed as global protein surface docking without restriction to a predefined binding pocket. Predicted binding poses should be interpreted as hypothesis-generating; for proteins with experimentally validated active sites, poses overlapping with those regions are most likely to be biologically relevant.

Supplementary Table S6. Software version information.

Category	Software/Tool	Version
Web Server	Apache	v2.4.58
Database	MySQL	v8.0.42
Backend Runtime	Node.js	v22.17.0
Front-end Framework	Bootstrap	v4.0.0-beta2
Front-end Library	jQuery	v3.2.1
3D Visualization	NGL Viewer	v0.10.4

Supplementary Table S7. Summary of $\Delta\Delta G$ values for resistance mutations across different antibiotic classes.

Antibiotic Class	Protein/Target	Total	Destabilizing (n, %)	Stabilizing (n, %)	Median (kcal/mol)	Min (kcal/mol)	Max (kcal/mol)
Quinolones	GyrA	97	92 (94.8%)	5 (5.2%)	-0.422	-1.950	+1.086
Quinolones	GyrB	23	22 (95.7%)	1 (4.3%)	-0.716	-1.977	+0.381
Quinolones	ParC	30	26 (86.7%)	4 (13.3%)	-0.288	-2.535	+0.395
Quinolones	ParE	8	6 (75.0%)	2 (25.0%)	-0.405	-1.740	+0.369
Glycopeptides	Multiple	13	12 (92.3%)	1 (7.7%)	-0.814	-1.326	+0.084
Oxazolidinones	Multiple	2	1 (50%)	1 (50.0%)	-0.149	-0.547	+0.250
Sulfonamides	Multiple	9	8 (88.9%)	1 (11.1%)	-0.790	-2.024	+0.030
Lipopeptides	Multiple	6	6 (100%)	0 (0%)	-0.636	-0.995	-0.335
Beta-lactam	PBP	28	26 (92.9%)	2 (7.1%)	-0.842	-2.043	+0.349
Beta-lactam	β -lactamase	27	24 (88.9%)	3 (11.1%)	-0.769	-2.483	+1.471
Total	All above	243	223 (91.8%)	20 (8.2%)	—	-2.535	+1.471

Note: Global median is not reported due to heterogeneous distributions across different target proteins.

Supplementary Table S8. Spatial distances between known resistance mutations and contact residues.

Known resistance mutation	Contact residue	Distance (Å)
D87	ARG91	6.1
S83	ARG91	12.1
S83	SER97	17.6

Note: QRDR, quinolone resistance-determining region. Distances were measured between the alpha-carbon atoms of the mutation site and the contact residue in the AlphaFold2-predicted GyrA structure from Glaesserella parasuis.