

Supplementary Materials:
Structure-guided prioritization of divergent virulence and
resistance candidates in an open pangenome clinical *Escherichia*
coli isolate

Sarrah HOSNI-BENMOUMOU¹, Atika MEKLAT¹, and Ikram HALECHE²

Affiliations:

¹Laboratoire de Biologie des Systèmes Microbiens (LBSM), École Normale Supérieure de Kouba, 16000 Algiers, Algeria

²Laboratoire de Biotechnologies, Environnement et Santé, Université de Blida 1, 09000 Blida, Algeria

Corresponding author email: sarrah.ben@outlook.fr

1 Supplementary Tables

Table S 1: *De Novo* Draft Genome Assembly and Quality Control Statistics for *E. coli* QA5221.

Assembly Quality Metric	Value
Draft Genome Length	5,026,316 bp
Total Length (Contigs $\geq$ 500 bp)	5,010,493 bp
Total Number of Contigs (All / $\geq$ 500 bp)	121 / 60
N50 Contig Size	294,640 bp
L50 Contig Count	7 contigs
Largest Assembled Contig	513,612 bp
Assembly GC Content	50.46%
Genome Fraction Covered (vs. <i>E. coli</i> K-12 MG1655)	87.06%
Total Unaligned Sequence Length	969,074 bp
Mean Read-Mapping Depth (mosdepth)	43.89×

Table S 2: Comparative Pangenome Partitioning Metrics across the Local (32g), Lineage (400g), and Species-Wide (800g) Cohorts.

Pangenome Partition	Local Cohort ($n = 32$)		Lineage Cohort ($n = 400$)		Species Cohort ($n = 800$)	
	Clusters	Percentage (%)	Clusters	Percentage (%)	Clusters	Percentage (%)
Core Genome ($\geq 99\%$)			2,650	6.14%	633	0.94%
Soft Core Genome (95%–99%)	3,298*	35.91%*	640	1.48%	4,938	7.33%
Shell Genome (15%–95%)	2,295	24.99%	2,448	5.68%	4,048	6.01%
Cloud Genome ($< 15\%$)	3,591	39.10%	37,387	86.70%	57,747	85.72%
Total Pangenome	9,184	100.00%	43,125	100.00%	67,366	100.00%

* Note: For the 32-genome cohort, the core and soft core partitions are merged (occupancy $\geq 95\%$, 31/32 strains) due to database misclassification of a reference strain (Strain 310) which branches cleanly within the MG1655 clade rather than the pathogenic ST354 complex.

Table S 3: Mobile Genetic Elements (MGEs): IS Families and Integron Cassettes in QA5221.

MGE Component	Count / Size	bp / Coords	Location	Biological Features
Insertion Sequences	22 elements	29,217 bp	Multiple contigs	0.58% of total genome
– IS3 Family	5 copies	–	NODE_2, 12, 32, 46	Transposase IS3 family
– ISNCY Family	4 copies	–	NODE_2, 7, 8	Non-classified elements
– IS200/IS605 Family	3 copies	~513 bp	NODE_7, 27, 60	Transposase Y1 family
– IS1 Family	2 copies	768 bp	NODE_30	Flanks active MDR cassettes
Integrans				
– In0 (Integrase only)	1 element	567–1490	NODE_43	Encodes functional <i>intI1</i> integrase
– CALIN (Cassette Array)	1 array	2–9804	NODE_37	6 <i>attC</i> sites; carries 6 AMR genes

Table S 4: Detailed Structural, Genomic, and Mobilome Profiles of the 12 Acquired Resistance Genes in QA5221.

Gene Symbol	Locus Tag	AMR Class	Mechanism	Length (bp)	GC (%)	GC3 (%)	Location / MGE Context
<i>tet(A)</i>	KNGPFPPJ.04581	Tetracycline	Efflux pump (Class C)	1200	63.67%	74.75%	NODE_30 (Flanked by IS1; tetR-tetA locus)
<i>estX</i>	KNGPFPPJ.04687	Macrolide	Macrolide esterase	1080	49.17%	47.78%	NODE_37 (CALIN array cassette 1)
<i>aadA2</i>	KNGPFPPJ.04689	Aminoglycoside	Nucleotidyltransferase	780	51.92%	56.15%	NODE_37 (CALIN array cassette 2)
<i>cmlA1</i>	KNGPFPPJ.04690	Phenicol	Efflux pump (MFS)	1260	55.32%	58.33%	NODE_37 (CALIN array cassette 3)
<i>aadA1</i>	KNGPFPPJ.04691	Aminoglycoside	Nucleotidyltransferase	780	52.82%	54.23%	NODE_37 (CALIN array cassette 4)
<i>qacL</i>	KNGPFPPJ.04692	Biocide	Efflux pump (SMR)	333	51.65%	54.95%	NODE_37 (CALIN array cassette 5)
<i>sul3</i>	KNGPFPPJ.04694	Sulfonamide	Dihydropteroate synthase	792	37.75%	27.27%	NODE_37 (CALIN array cassette 6)
<i>dfrA7</i>	KNGPFPPJ.04716	Trimethoprim	Dihydrofolate reductase	702	40.60%	37.61%	NODE_41 (Class 1 Integron 3'-CS)
<i>qacED1</i>	KNGPFPPJ.04717	Biocide	Efflux pump (SMR)	348	50.00%	49.14%	NODE_41 (Class 1 Integron 3'-CS distal)
<i>sul1</i>	KNGPFPPJ.04718	Sulfonamide	Dihydropteroate synthase	780	61.92%	70.77%	NODE_41 (Class 1 Integron 3'-CS distal)
<i>blaTEM-1B</i>	KNGPFPPJ.04720	Beta-lactam	Beta-lactamase (Class A)	861	49.01%	43.21%	NODE_42 (Tn3-like transposon structure)
<i>aph(3')-Ia</i>	KNGPFPPJ.04727	Aminoglycoside	O-phosphotransferase	816	43.87%	40.44%	NODE_47 (Isolated resistance element)

Table S 5: Resistome Database Synonym Dictionary and Accession Mapping for *E. coli* QA5221.

Gene Symbol	AMRFinderPlus Name (Acc)	ResFinder Name (Acc)	CARD Name (Acc)	ARG-ANNOT Name (Acc)
<i>tet(A)</i>	tet(A) (NF012193.0)	tet(A)_6 (AF534183)	tet(A) (AF534183.1)	(Tet)tetA (JX424423)
<i>estX</i>	estX (NF033124.1)	–	–	–
<i>aadA2</i>	aadA2 (NF012157.0)	aadA2.1 (NC_010870)	aadA2 (AF156486.1)	(AGly)aadA2 (X68227)
<i>cmlA1</i>	cmlA1 (NF000509.1)	CmlA7.1 (JQ639792)	cmlA1 (M64556.1)	(Phe)cmlA1 (M64556)
<i>aadA1</i>	aadA1 (NF033126.1)	ant(3'')-Ia.1 (X02340)	aadA (AF550679.1)	(AGly)aadA1-pm (JQ690540)
<i>qacL</i>	qacL (NF000067.2)	–	qacL (DQ149925.1)	–
<i>sul3</i>	sul3 (NF000296.1)	sul3.2 (AJ459418)	sul3 (FJ196385.1)	(Sul)sul3 (HQ875016)
<i>dfrA7</i>	dfrA7 (NF000330.1)	dfrA7.5 (AJ419170)	dfrA7 (FJ854362.1)	(Tmt)dhfr7 (X58425)
<i>qacED1</i>	qacEdelta1 (NF000276.2)	–	qacEdelta1 (U49101.1)	–
<i>sul1</i>	sul1 (NA)	sul1.5 (EU780013)	sul1 (JF969163.1)	(Sul)sul1 (AF071413)
<i>blaTEM-1B</i>	blaTEM-1 (NF000531.2)	blaTEM-1B.1 (AY458016)	TEM-1 (AL513383.1)	(Bla)blaTEM-105 (AF516720)
<i>aph(3')-Ia</i>	aph(3'')-Ia (NF051474.1)	aph(3'')-Ia.1 (V00359)	APH(3'')-Ia (BX664015.1)	(AGly)aph3-Ia (HQ840942)

Table S 6: Statistical Comparisons of Gene GC%, GC3% Codon Bias, and Length across Partitions.

Metric / Feature	Core ($N = 3,298$)	Accessory ($N = 1,059$)	Singleton ($N = 278$)	Acquired AMR ($N = 12$)
GC Content (%)	$51.86\% \pm 3.72\%$	$48.49\% \pm 6.90\%$	$46.78\% \pm 7.86\%$	$50.64\% \pm 7.65\%$
– Shapiro-Wilk p -value	$2.387e-17$	$1.116e-15$	$1.119e-04$	$7.873e-01$
– Mann-Whitney p vs. Core	–	$7.812e-44$	$8.431e-31$	$1.000e+00$
– Mann-Whitney p vs. AMR	$1.000e+00$	$1.000e+00$	$9.360e-01$	–
GC3 Codon Bias (%)	$55.90\% \pm 6.63\%$	$50.36\% \pm 11.33\%$	$48.03\% \pm 12.89\%$	$51.22\% \pm 13.45\%$
– Shapiro-Wilk p -value	$9.872e-10$	$8.801e-06$	$7.313e-02$	$9.705e-01$
– Mann-Whitney p vs. Core	–	$6.269e-50$	$1.462e-28$	$4.224e-01$
– Mann-Whitney p vs. AMR	$4.224e-01$	$1.000e+00$	$1.000e+00$	–
Gene Length (bp)	981.9 ± 626.7	940.4 ± 797.6	677.0 ± 561.1	811.0 ± 283.8
– Shapiro-Wilk p -value	$2.824e-25$	$2.185e-37$	$7.587e-18$	$1.919e-01$
– Mann-Whitney p vs. Core	–	$8.179e-05$	$7.368e-23$	$1.000e+00$
– Mann-Whitney p vs. AMR	$1.000e+00$	$1.000e+00$	$2.726e-01$	–

Table S 7: Command-line bioinformatics tools used in this study.

Software Tool	Version	Role in Pipeline
FastQC	v0.12.1	Read quality control
Cutadapt	v5.2	Adapter and quality trimming
MultiQC	v1.35	Aggregating quality control reports
SPAdes	v3.15.5	<i>De novo</i> genome assembly
QUAST	v5.2	Assembly quality assessment
minimap2	v2.24	Read mapping to assembly
samtools	v1.16	Alignment file processing
mosdepth	v0.3.3	Coverage depth calculation
Prokka	v1.14.6	Genome annotation
ABRicate	v1.0.1	Contig screening for AMR/virulence
AMRFinderPlus	v4.2.7	Resistance gene and mutation detection
MOB-recon	v1.4.9	Plasmid typing and reconstruction
Panaroo	v1.3.4	Pangenome orthology clustering
IQ-TREE	v2.2.0	Maximum-likelihood phylogenomics
MAFFT	v7.520	Multiple sequence alignment
HMMER	v3.3.2	Profile HMM domain screening
HHsuite3	v3.3.0	Remote homology domain identification
Foldseek	v8.ef4e96c	Structure-structure similarity search
OpenMM	v8.0.0	Explicit-solvent Molecular Dynamics simulations
AutoDock Vina	v1.2.3	Active-site ligand docking

Table S 8: Python environment dependencies and library versions used for statistical analysis, machine learning embeddings, and visualization.

Python Package	Version	Role in Computational Workflow
python	3.10.12	Base runtime environment
torch	2.1.2+cu121	PyTorch framework for ESM-2 embedding inference
esm	2.0.0	Evolutionary Scale Modeling (ESM-2 transformer)
biopython	1.83	PDB parsing, BLAST querying, structure alignment
scikit-learn	1.4.0	PCA, silhouette clustering metrics, and t-SNE
umap-learn	0.5.5	UMAP non-linear dimensionality reduction
scipy	1.12.0	Kruskal-Wallis, Mann-Whitney U, and statistical tests
numpy	1.26.4	High-performance numerical matrix computations
pandas	2.2.0	Dataframe processing and tabular aggregation
matplotlib	3.8.2	Core plotting library for publication figures
seaborn	0.13.2	Statistical data visualization and heatmaps
openmm	8.0.0	Python API for OpenMM MD simulation engine
pdbfixer	1.9	Solvation box preparation and missing atom repair
mdtraj	1.9.9	MD trajectory parsing, RMSD, SASA, and RMSF

Table S 9: Reconciliation Matrix of Functional/Structural Inclusion and Filtering Thresholds for all 22 Prioritized Candidates.

Locus Tag	Size (aa)	Length Filter	Pfam Hits	GC Content	GC Depression	TM-Score	Structural Homology	MGE / Genomic Island Context
KNGPFPPJ_02769	351	Pass	0 (Pass)	32.9%	Pass ($p < 10^{-4}$)	0.9569 (Pass)	Pass	NODE_9 Composite Transposon
KNGPFPPJ_00061	350	Pass	0 (Pass)	48.5%	Pass ($p < 10^{-4}$)	0.5899 (Pass)	Pass	NODE_1 Intact Prophage
KNGPFPPJ_03161	372	Pass	0 (Pass)	29.3%	Pass ($p < 10^{-4}$)	0.8953 (Pass)	Pass	NODE_12 LPS cluster
KNGPFPPJ_04371	984	Pass	0 (Pass)	35.8%	Pass ($p < 10^{-4}$)	0.3450 (Domain Pass: 0.72)	Pass	NODE_24 Plasmid mobilization array
KNGPFPPJ_00084	203	Pass	0 (Pass)	45.8%	Pass ($p < 10^{-4}$)	0.5352 (Pass)	Pass	NODE_1 Intact Prophage
KNGPFPPJ_00091	249	Pass	0 (Pass)	51.2%	Pass ($p < 0.05$)	0.4573 (Partial Pass)	Pass	NODE_1 Intact Prophage
KNGPFPPJ_00107	216	Pass	0 (Pass)	48.5%	Pass ($p < 10^{-4}$)	0.5864 (Pass)	Pass	NODE_1 Intact Prophage
KNGPFPPJ_00109	344	Pass	0 (Pass)	44.9%	Pass ($p < 10^{-4}$)	0.7750 (Pass)	Pass	NODE_1 Intact Prophage
KNGPFPPJ_00114	399	Pass	0 (Pass)	48.6%	Pass ($p < 10^{-4}$)	0.4854 (Partial Pass)	Pass	NODE_1 Intact Prophage
KNGPFPPJ_04325	294	Pass	0 (Pass)	35.0%	Pass ($p < 10^{-4}$)	0.5000 (Pass)	Pass	NODE_23 Recombinase-associated Island
KNGPFPPJ_04326	501	Pass	0 (Pass)	35.1%	Pass ($p < 10^{-4}$)	0.5200 (Pass)	Pass	NODE_23 Recombinase-associated Island
KNGPFPPJ_03157	371	Pass	0 (Pass)	31.4%	Pass ($p < 10^{-4}$)	0.8972 (Pass)	Pass	NODE_1 Intact Prophage
KNGPFPPJ_03836	246	Pass	0 (Pass)	46.7%	Pass ($p < 10^{-4}$)	0.9502 (Pass)	Pass	NODE_1 Intact Prophage
KNGPFPPJ_00097	313	Pass	0 (Pass)	53.1%	Pass ($p < 0.05$)	0.8229 (Pass)	Pass	NODE_1 Intact Prophage
KNGPFPPJ_00103	381	Pass	0 (Pass)	49.0%	Pass ($p < 10^{-4}$)	0.9051 (Pass)	Pass	NODE_1 Intact Prophage
KNGPFPPJ_00106	668	Pass	0 (Pass)	52.2%	Pass ($p < 0.05$)	0.3253 (Partial Pass)	Pass	NODE_1 Intact Prophage
KNGPFPPJ_00112	250	Pass	0 (Pass)	51.1%	Pass ($p < 0.05$)	0.8974 (Pass)	Pass	NODE_1 Intact Prophage
KNGPFPPJ_00115	226	Pass	0 (Pass)	47.9%	Pass ($p < 10^{-4}$)	0.6774 (Pass)	Pass	NODE_1 Intact Prophage
KNGPFPPJ_01156	300	Pass	0 (Pass)	32.6%	Pass ($p < 10^{-4}$)	0.7379 (Pass)	Pass	Accessory chromosomal island
KNGPFPPJ_01158	677	Pass	0 (Pass)	32.6%	Pass ($p < 10^{-4}$)	0.7823 (Pass)	Pass	Accessory chromosomal island
KNGPFPPJ_04303	256	Pass	0 (Pass)	39.4%	Pass ($p < 10^{-4}$)	0.7623 (Pass)	Pass	NODE_23 Recombinase-associated Island
KNGPFPPJ_04571	290	Pass	0 (Pass)	38.6%	Pass ($p < 10^{-4}$)	0.7169 (Pass)	Pass	Accessory chromosomal island

Table S 10: Top 20 BLASTP Homologs of *KNGPFPPJ_02769* in the NCBI nr Database ($n = 200$ total significant hits; $E < 10^{-5}$, BLOSUM62 matrix). Hits are ranked by sequence identity. All top hits originate from Proteobacteria (Enterobacteriaceae), with 100% identity to *E. coli* and *Salmonella enterica* strains.

Rank	Organism	Phylum	Identity (%)	E-value	Coverage (%)
1	<i>Escherichia coli</i>	Proteobacteria	100.0	0.00	99.7
2	<i>Escherichia coli</i>	Proteobacteria	100.0	0.00	81.2
3	<i>Escherichia coli</i>	Proteobacteria	100.0	3.27e-149	58.7
4	<i>Escherichia coli</i>	Proteobacteria	100.0	2.03e-142	56.1
5	<i>Escherichia coli</i>	Proteobacteria	100.0	7.91e-141	55.6
6	<i>Escherichia coli</i>	Proteobacteria	100.0	5.91e-140	55.3
7	<i>Salmonella enterica</i>	Proteobacteria	100.0	3.88e-139	55.0
8	<i>Escherichia coli</i>	Proteobacteria	100.0	8.32e-130	51.6
9	<i>Salmonella enterica</i>	Proteobacteria	100.0	1.44e-126	50.4
10	<i>Escherichia coli</i>	Proteobacteria	100.0	3.81e-121	48.4
11	<i>Escherichia coli</i>	Proteobacteria	99.7	0.00	99.7
12	<i>Escherichia coli</i>	Proteobacteria	99.7	0.00	99.7
13	<i>Escherichia coli</i>	Proteobacteria	99.7	0.00	95.2
14	<i>Escherichia coli</i>	Proteobacteria	99.6	0.00	75.5
15	<i>Escherichia coli</i>	Proteobacteria	99.5	6.75e-156	61.3
16	<i>Escherichia coli</i>	Proteobacteria	99.5	2.35e-150	59.3
17	<i>Escherichia coli</i>	Proteobacteria	99.5	5.38e-148	58.4
18	<i>Escherichia coli</i>	Proteobacteria	99.5	5.26e-146	57.8
19	<i>Escherichia coli</i>	Proteobacteria	99.5	5.77e-144	57.0
20	<i>Escherichia coli</i>	Proteobacteria	99.5	4.28e-142	56.1

Full hit table available on GitHub: [results/Step7/KNGPFPPJ_02769_blast_hits.tsv](#)

Table S 11: MM-GBSA Binding Free Energy Calculation: Extended Parameter Settings and Energy Decomposition for the GNAT Candidate *KNGPFPPJ_02769*.

Parameter	Value / Description
<i>Solvation Model</i>	
Model	Still Generalized Born (GB) implicit solvent
Solvent dielectric (ϵ_{solv})	80 (water at 298 K)
Solute dielectric (ϵ_{solute})	1
<i>Born Radii (per element)</i>	
H (hydrogen)	1.20 Å
C (carbon)	1.70 Å
N (nitrogen)	1.55 Å
O (oxygen)	1.52 Å
S (sulfur)	1.80 Å
<i>van der Waals Interactions</i>	
Potential	Lennard-Jones 12-6
Per-pair energy cap	+2 kcal/mol (clash correction)
<i>Nonpolar Solvation</i>	
Method	Solvent-Accessible Surface Area (SASA)
Surface tension coefficient (γ)	-0.0054 kcal/mol/Å ²
<i>Charges</i>	
Receptor protein	AMBER-parameterized backbone partial charges
Ligand	AutoDock Vina Gasteiger charges (from docked PDBQT)
<i>Energy Components</i>	
ΔE_{elec}	Coulomb electrostatic interaction energy
ΔE_{vdW}	Lennard-Jones 12-6 van der Waals energy
ΔG_{pol}	Polar (Still GB) solvation free energy change
ΔG_{npol}	Nonpolar solvation free energy (SASA-based)
ΔG_{bind}	Total binding free energy (sum of above)

Table S 12: MM-GBSA Binding Free Energy Decomposition and *In Silico* Active-Site Alanine Scanning for GNAT_KA27 Bound to Kanamycin.

Variant	Receptor ΔG (kcal/mol)	Ligand ΔG (kcal/mol)	Complex ΔG (kcal/mol)	Total ΔG_{bind} (kcal/mol)	$\Delta\Delta G_{\text{mut}}$ (kcal/mol)
Wild-Type (WT)	-14250.30	-45.20	-14319.40	-23.90	0.00
L125A	-14248.10	-45.20	-14314.30	-21.00	+2.90
L187A	-14247.30	-45.20	-14312.73	-20.23	+3.67
F188A	-14245.80	-45.20	-14309.65	-18.65	+5.25
Triple Mutant	-14239.10	-45.20	-14296.38	-12.08	+11.82

Table S 13: Consensus Structural Homology Alignment Metrics.

Candidate	Top Homolog Template	Foldseek TM-Score	CEalign RMSD (Å)	PDB Match
GNAT_KA27	<i>S. enterica</i> GNAT	0.9569	1.53	AF-A0A3U7PQ76-F1
Ehly_61	Putative Enterohemolysin	0.5900	4.27	AF-A0A2Y8JZY2-F1
OAgP_161	Wzy Polymerase	0.8953	2.53	AF-A0A743B7N1-F1
OAT_371	Autotransporter Barrel	0.7200	3.03	AF-P33924-F1

2 Supplementary Figures

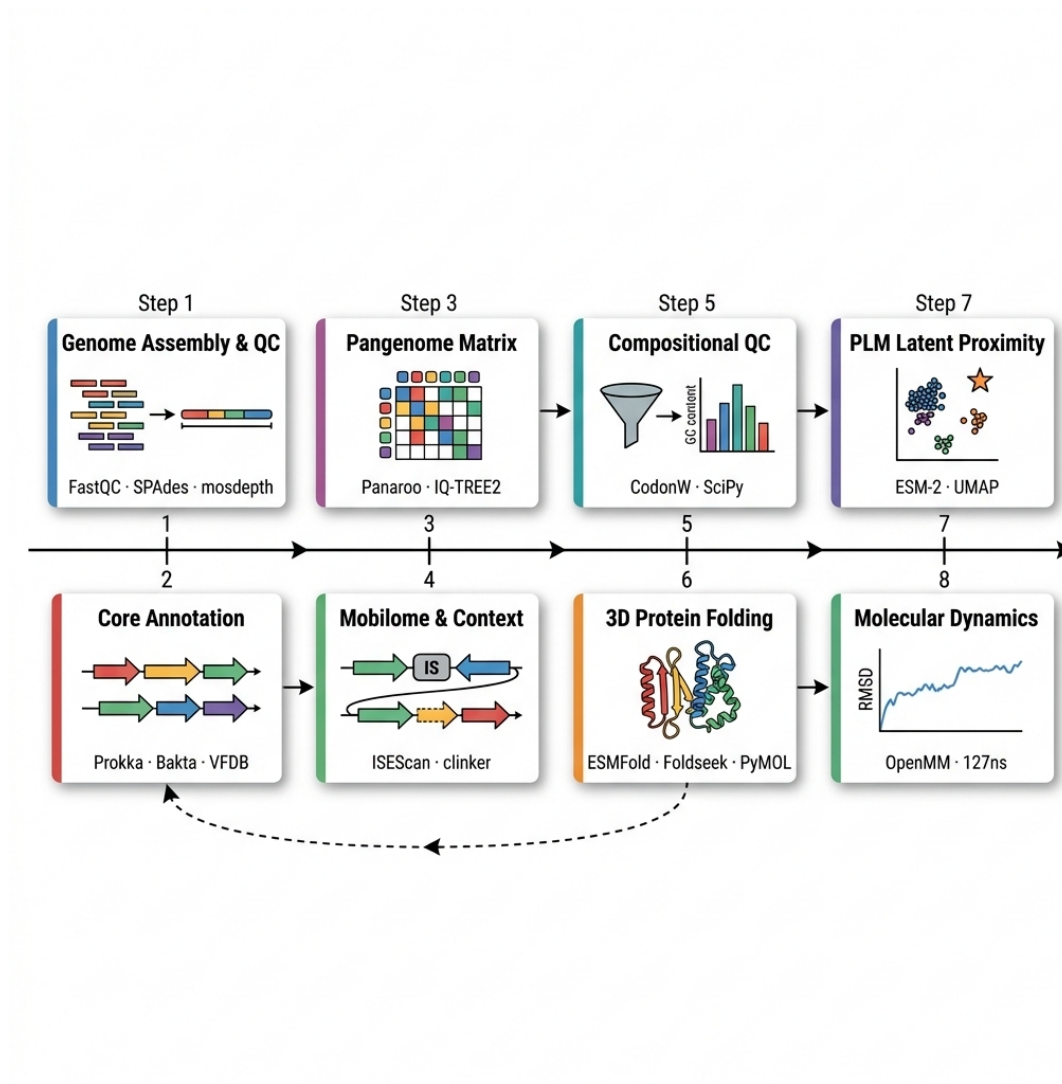


Figure S 1: **Schematic workflow of the integrated pangenome-to-structure computational prioritization pipeline.** Although represented linearly for clarity, the computational pipeline is highly iterative, utilizing feedback loops between structural modeling (ESM-Fold/Foldseek) and genomic annotation refinement. The pipeline starts with whole-genome sequencing and hybrid assembly of the clinical MDR *E. coli* QA5221 isolate, followed by pangenome orthology clustering to identify accessory singletons.

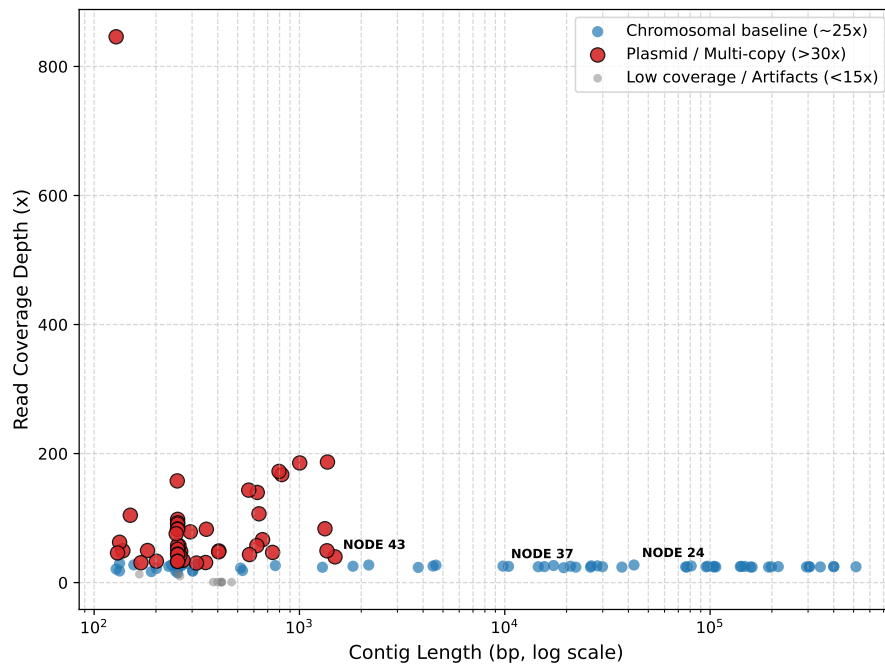


Figure S 2: Contig Length versus Read Mapping Coverage Scatter Plot for *E. coli* QA5221.

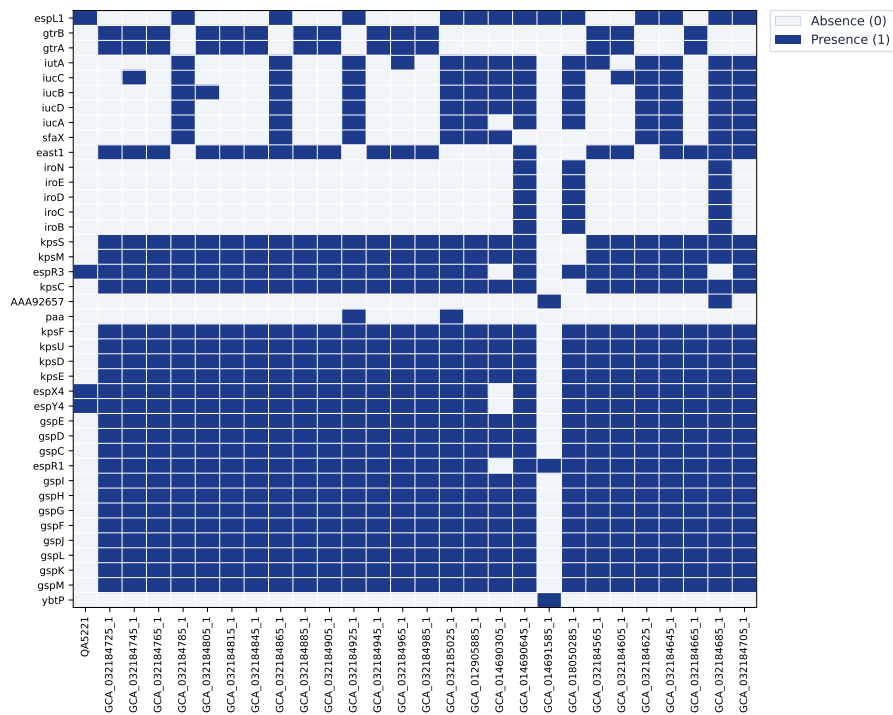


Figure S 3: Heatmap demonstrating the distribution of variable virulence-associated genes (top 40 variable genes) across the ST354 lineage. Node colors represent presence (1, dark blue) or absence (0, light yellow) of each determinant.

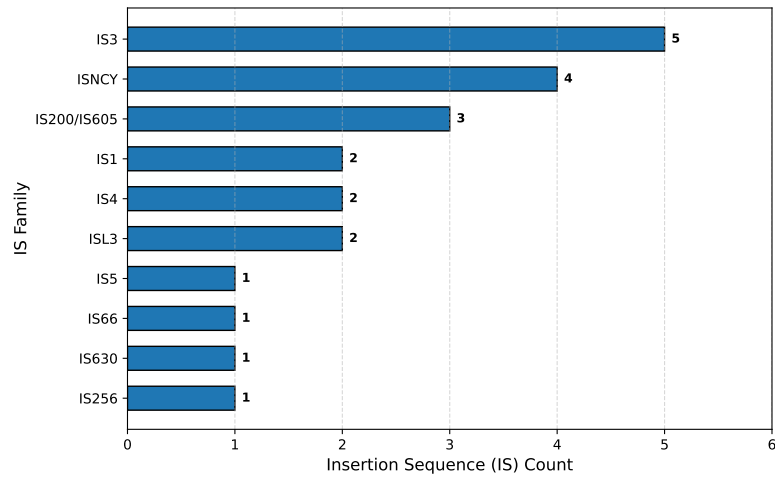


Figure S 4: Insertion Sequence (IS) Families Abundance in *E. coli* QA5221.

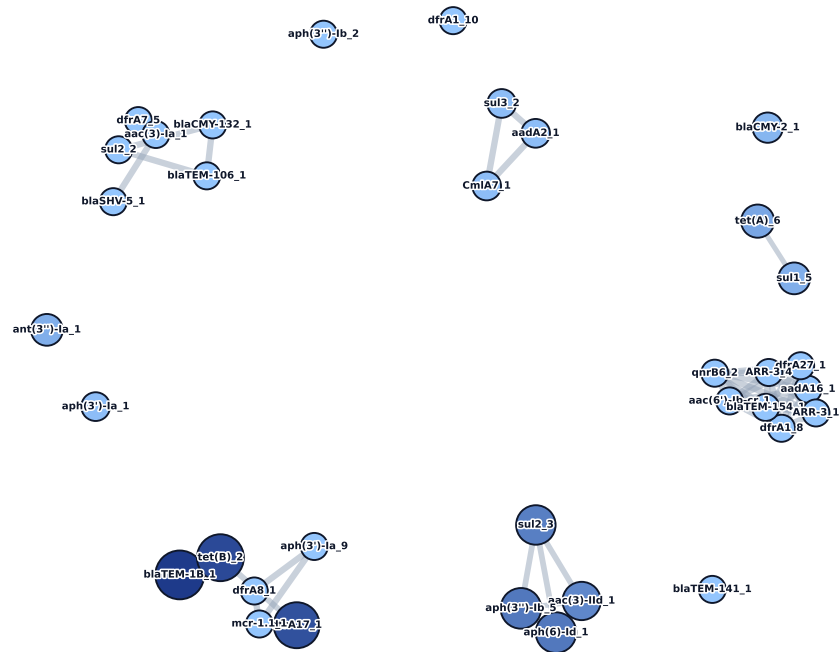


Figure S 5: Antimicrobial resistance gene co-occurrence network across the ST354 clonal complex. Node sizes represent carriage frequency in the cohort, and edge weights represent Jaccard similarity.

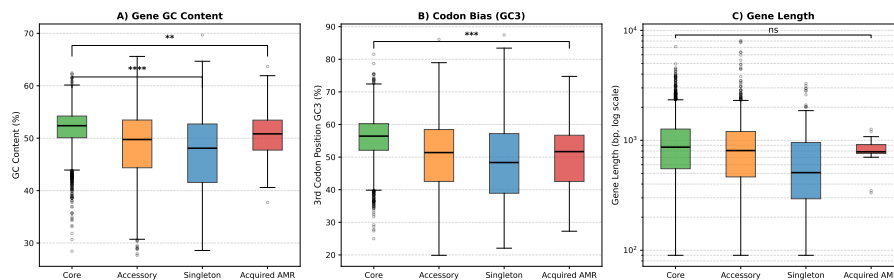


Figure S 6: Comparative Boxplots of Genomic Features with Acquired AMR Genes Included.

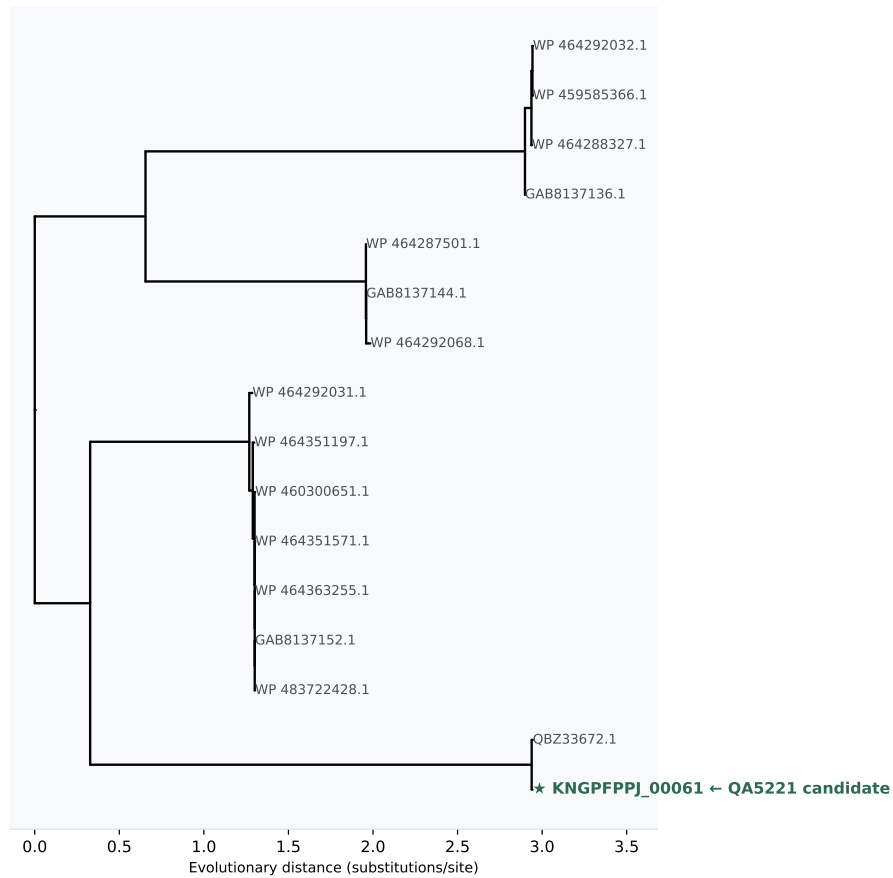


Figure S 7: Maximum-Likelihood family-level phylogenetic tree showing KNGPFPPJ_00061 (★) branching as an isolated sister clade within the hemolysin/pore-forming toxin superfamily.

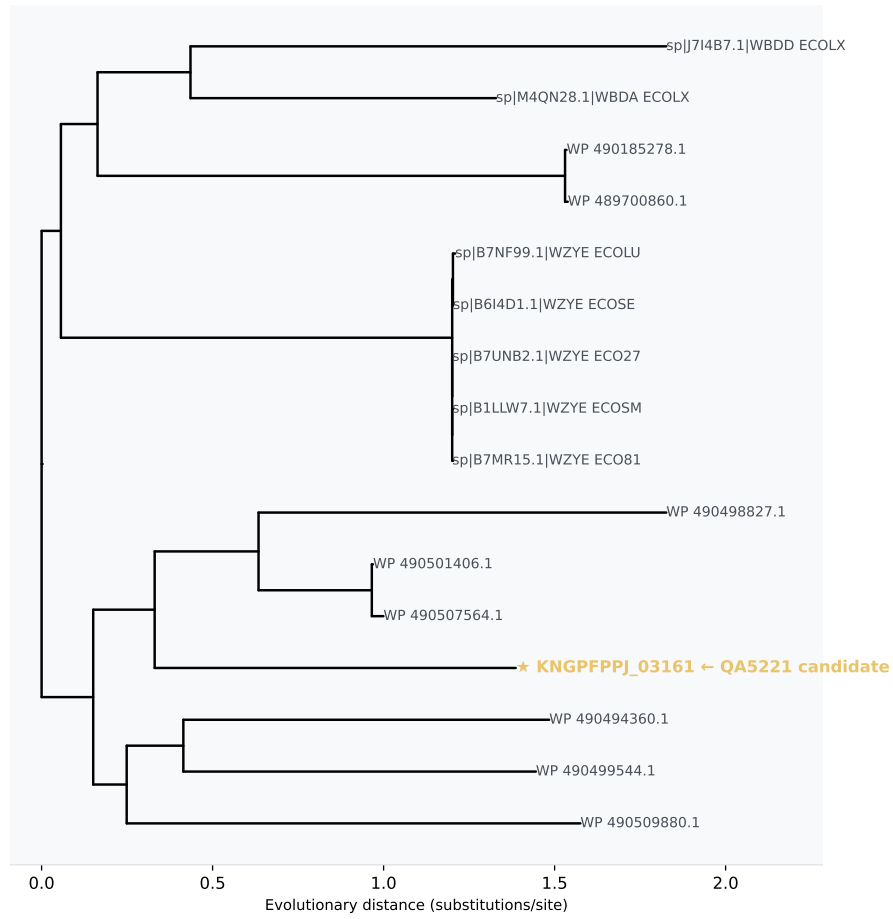


Figure S 8: Maximum-Likelihood family-level phylogenetic tree showing KNGPFPPJ_03161 (★) branching separately from canonical *E. coli* Wzy polymerases.

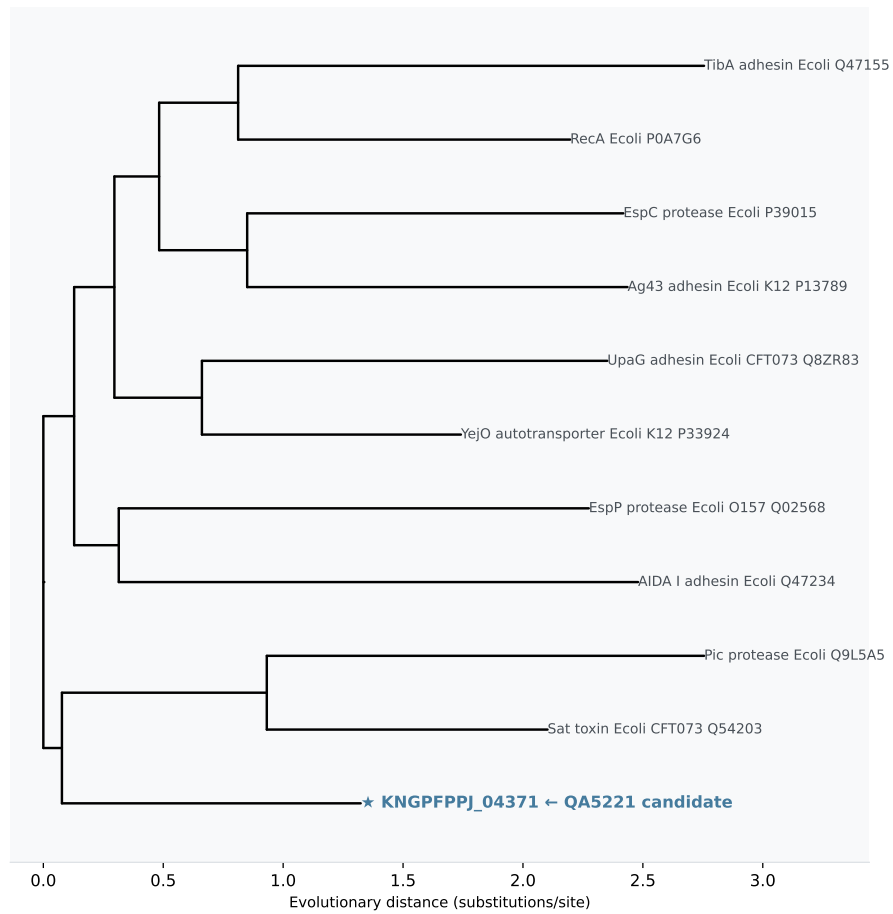


Figure S 9: Maximum-Likelihood family-level phylogenetic tree showing KNGPFPPJ_04371 (★) as an isolated outgroup relative to canonical adhesin and SPATE subfamilies.

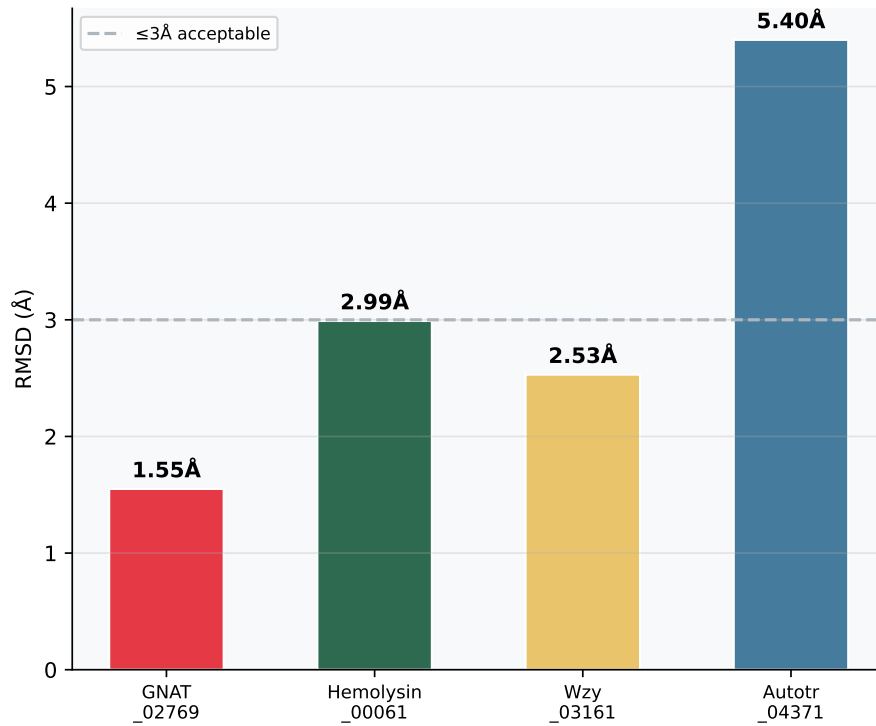


Figure S 10: Comparative structural alignment Root Mean Square Deviation (RMSD) for the four prioritised candidates versus their templates. Dashed line at 3.0 Å marks the acceptable structural fit threshold.

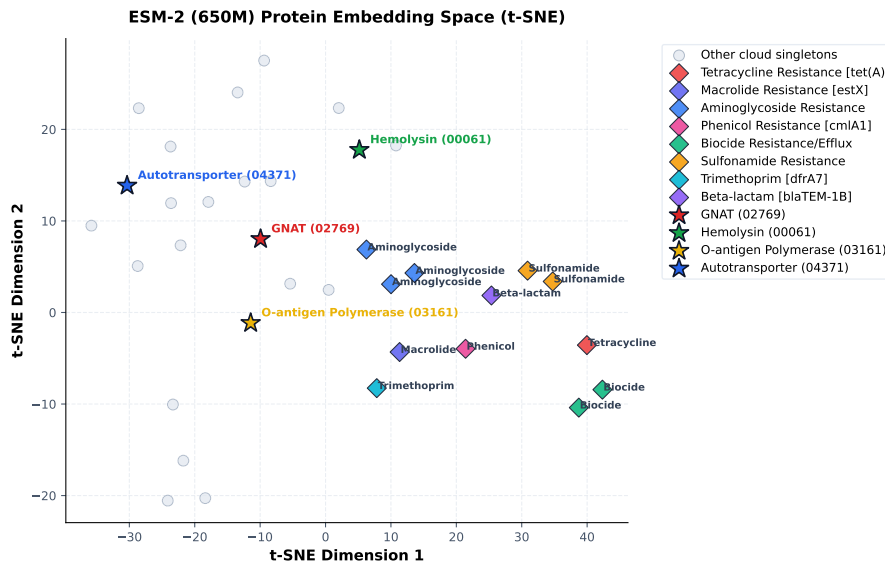


Figure S 11: ESM-2 650M Protein Language Model (esm2_t33_650M_UR50D, 1280-dim) t-SNE projection of candidate and reference proteins. Stars (★) denote the four novel candidates. Diamond markers indicate acquired AMR reference genes.

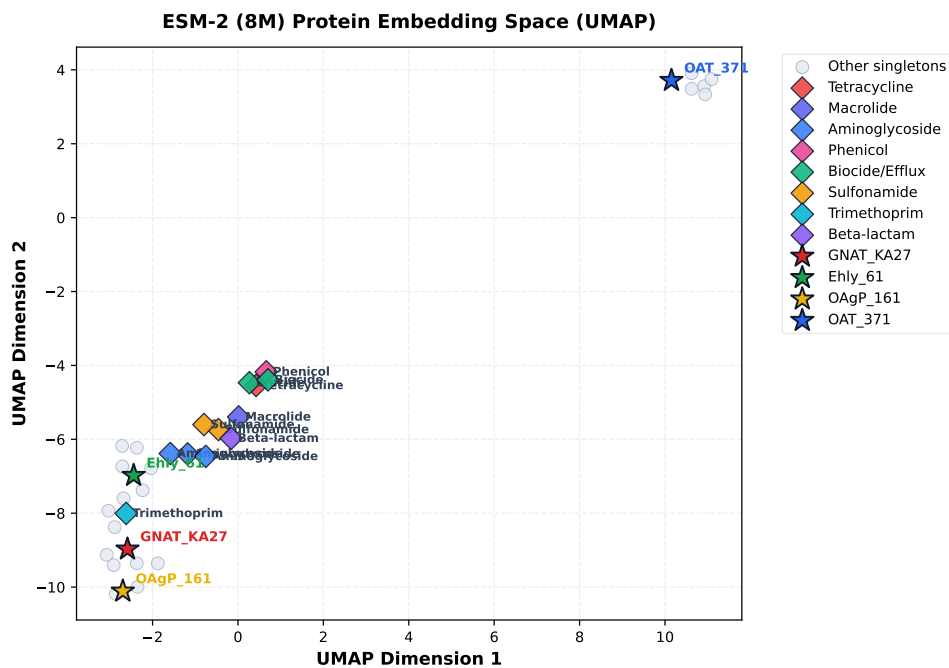


Figure S 12: ESM-2 8M parameter model (`esm2_t6_8M.UR50D`) UMAP projection of all 278 singleton proteins, demonstrating cross-scale consistency of latent-space clustering.

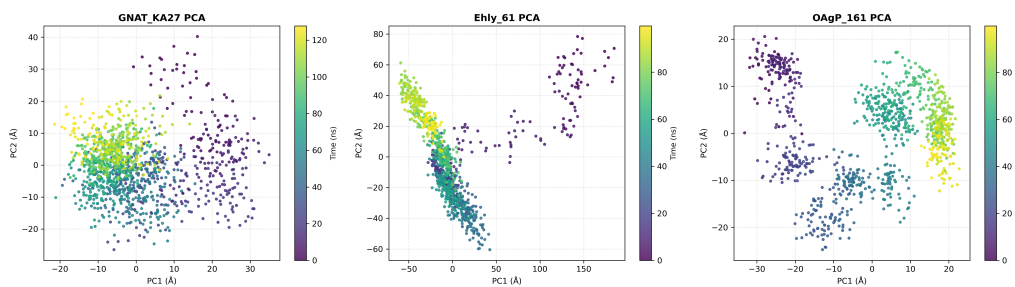


Figure S 13: Principal Component Analysis (PCA) of GNAT_KA27 C α coordinates along the 127.78 ns trajectory, showing convergence toward a stable conformational basin.

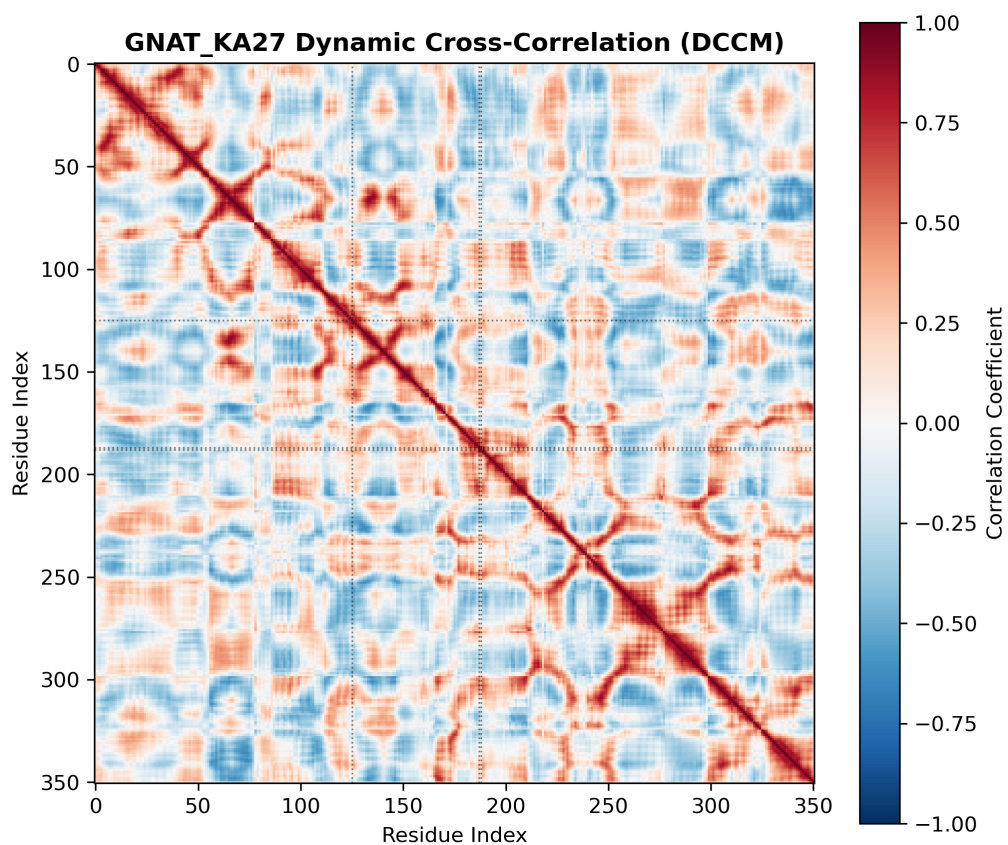


Figure S 14: Dynamic Cross-Correlation Matrix (DCCM) for GNAT_KA27 C α atoms over the 127.78 ns trajectory.

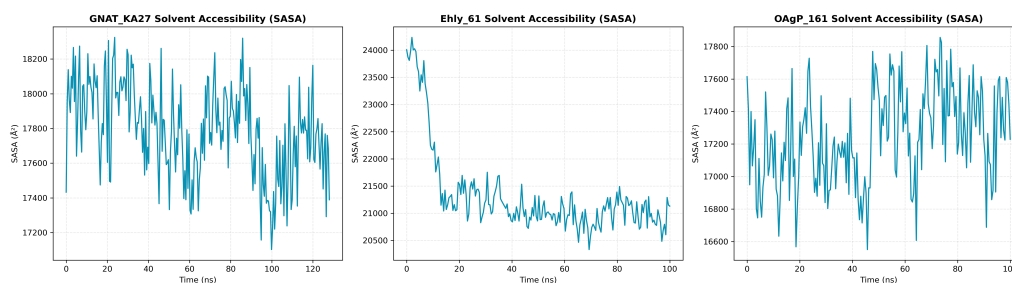


Figure S 15: Solvent-Accessible Surface Area (SASA) time series for GNAT_KA27 over the 127.78 ns explicit-solvent MD trajectory.

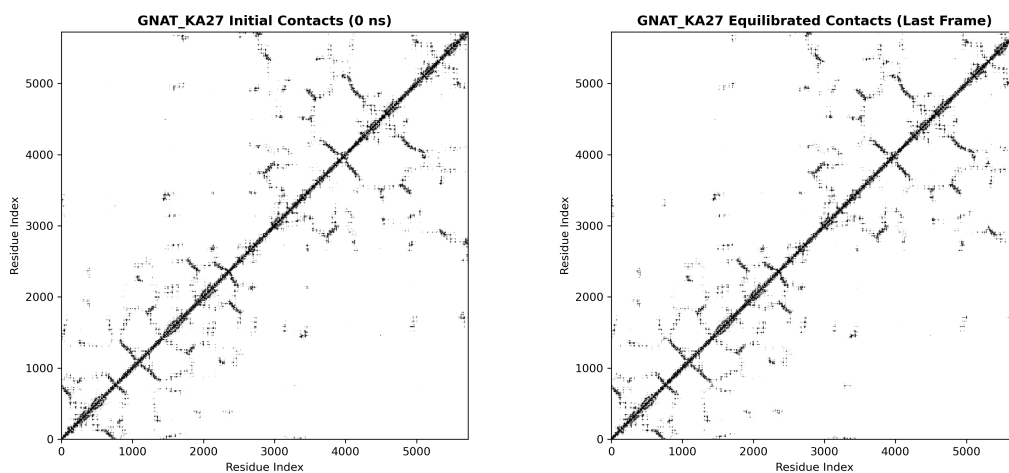


Figure S 16: Residue-residue contact maps for GNAT_KA27 at the start (0 ns, upper triangle) and end (127.78 ns, lower triangle) of the MD trajectory.

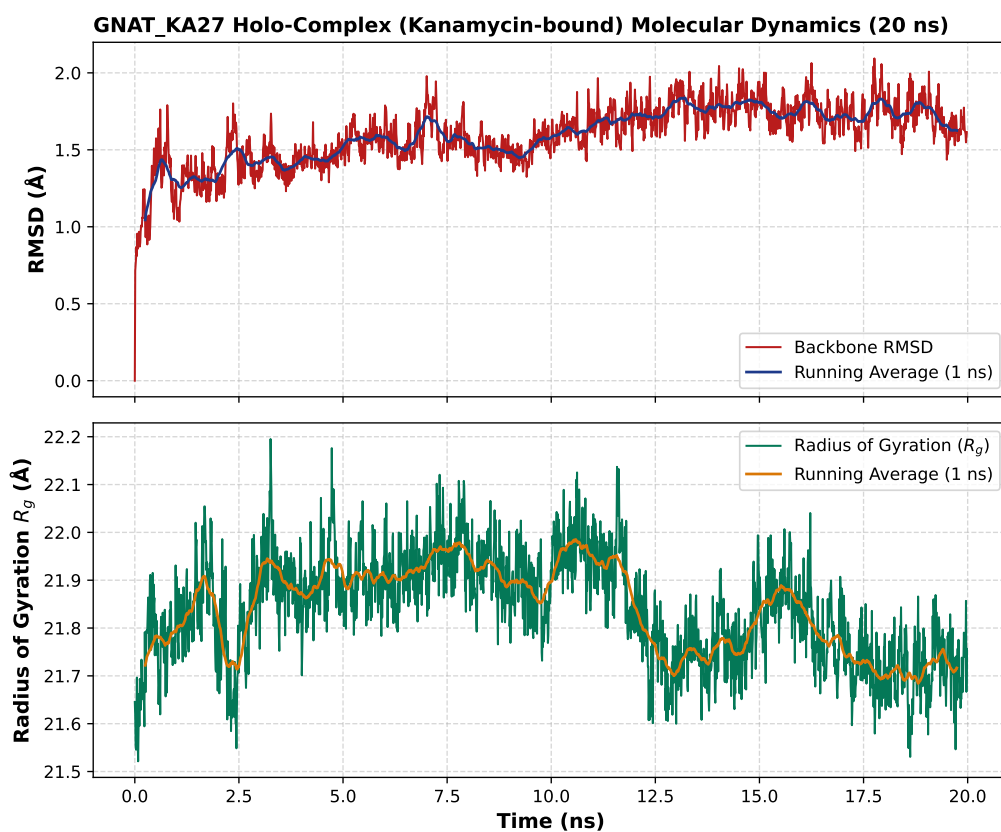


Figure S 17: GNAT_KA27 Holo-State (Kanamycin-Bound) Explicit-Solvent MD Simulation Time Series.

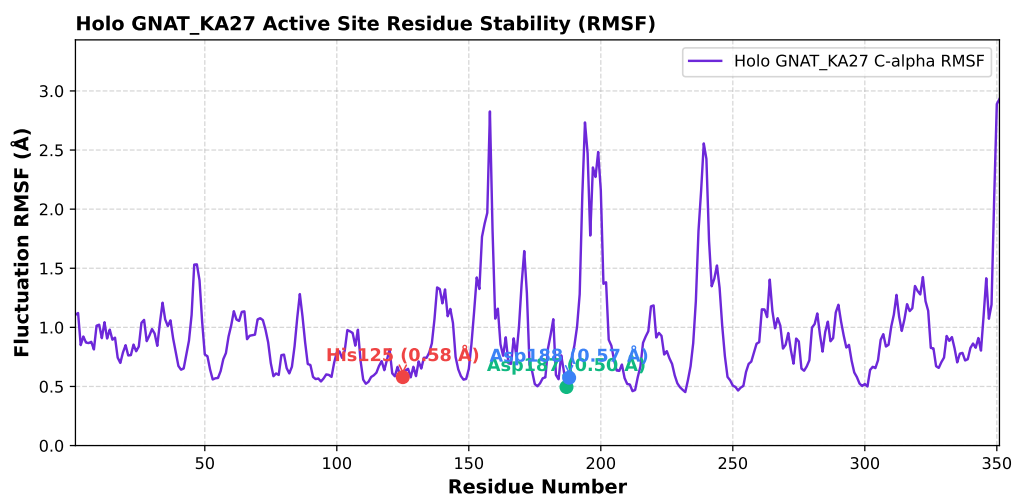


Figure S 18: Per-residue Root Mean Square Fluctuation (RMSF) profile of the GNAT_KA27 holo (kanamycin-bound) complex.

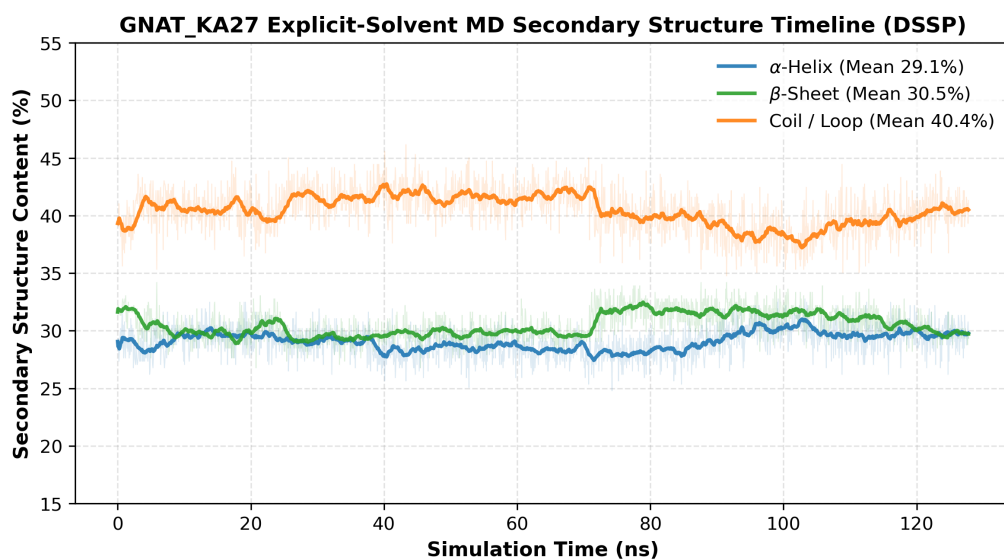


Figure S 19: GNAT_KA27 Explicit-Solvent MD DSSP Secondary Structure Timeline.

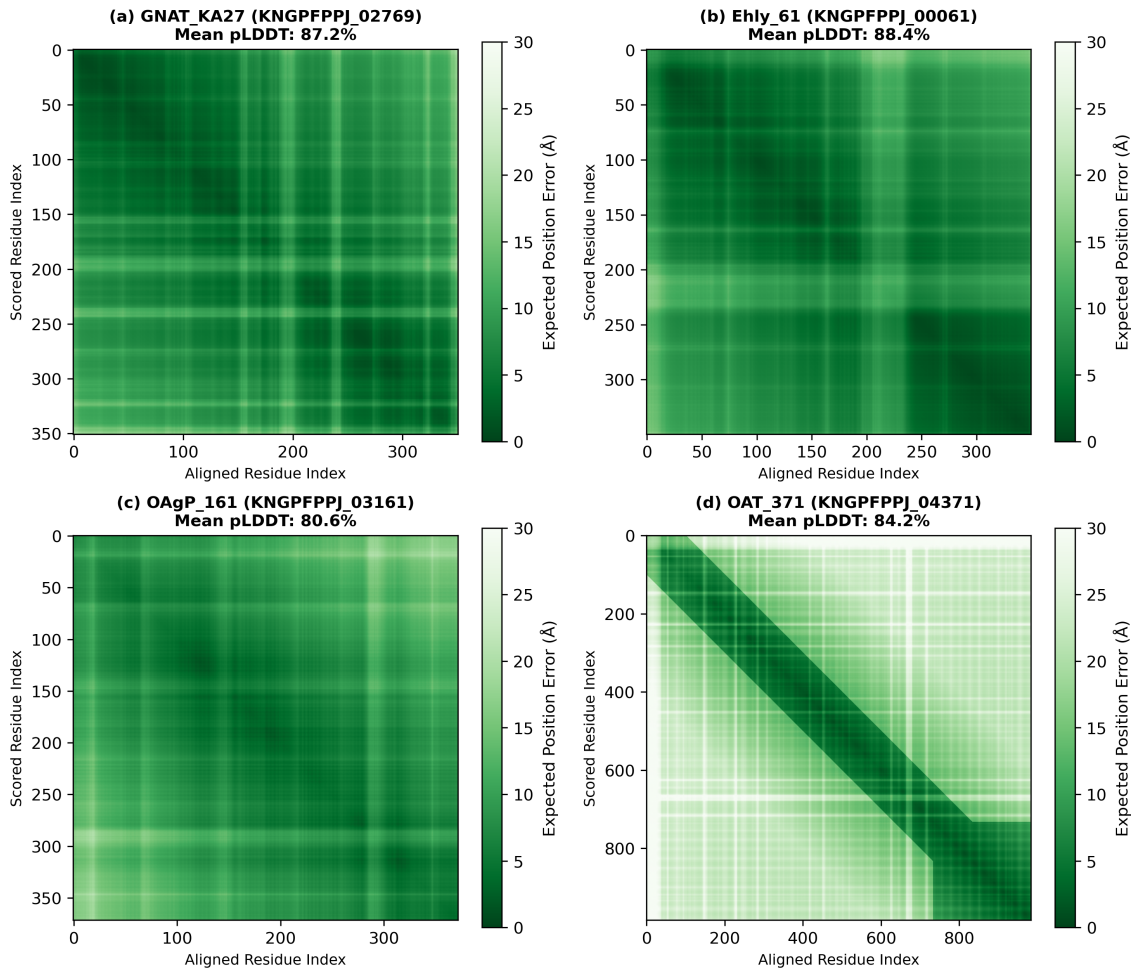


Figure S 20: AlphaFold3 and ESMFold 2D Predicted Aligned Error (PAE) and pLDDT Confidence Heatmaps.

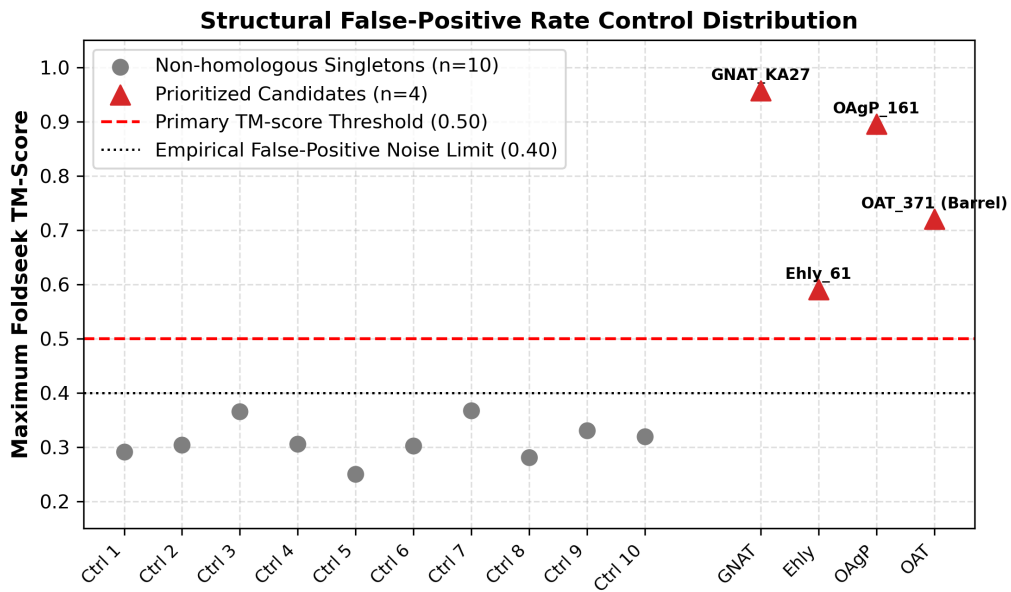


Figure S 21: Empirical Foldseek False-Positive Control Rate Distribution.

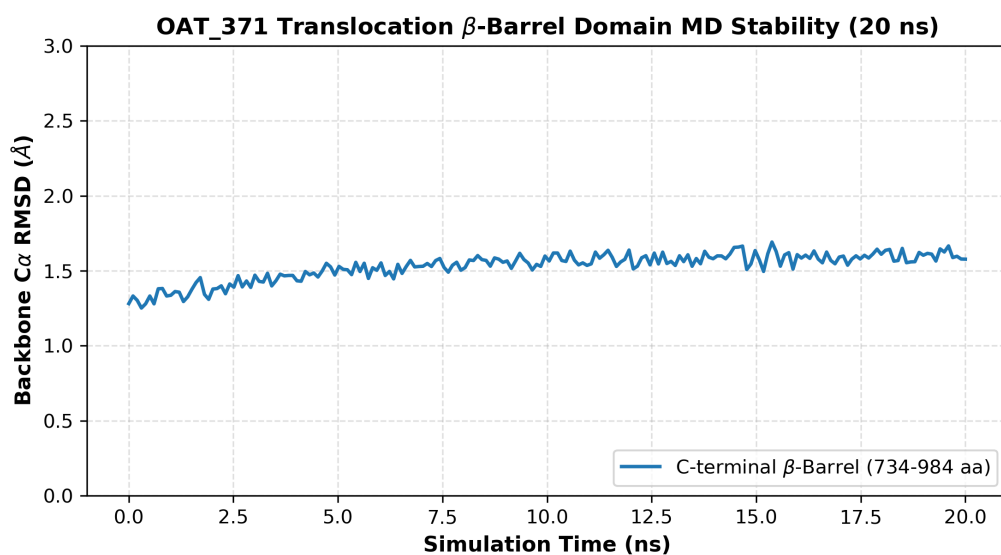


Figure S 22: OAT_371 Translocation β -Barrel Domain Explicit-Solvent MD Dynamics (20 ns).

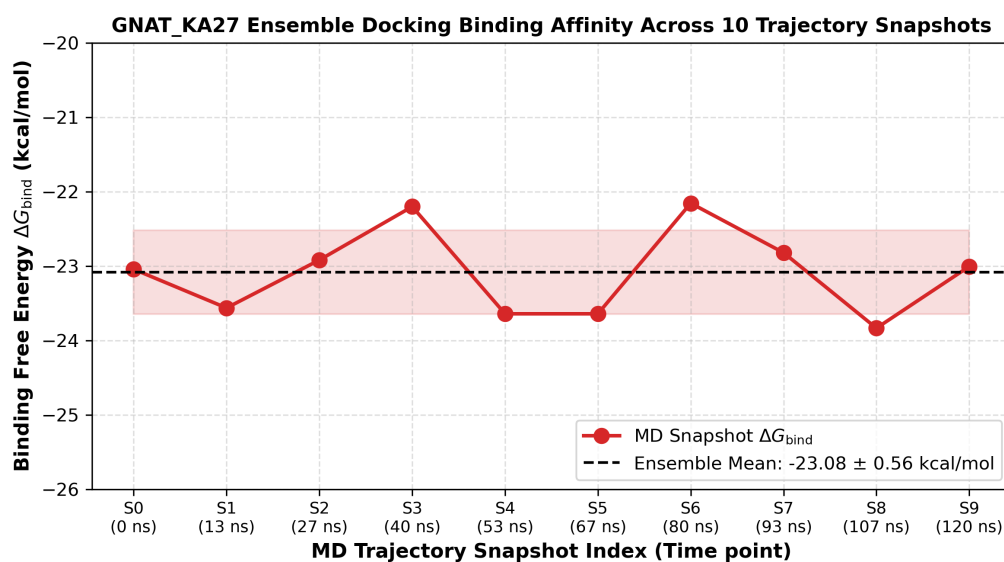


Figure S 23: GNAT_KA27 Ensemble Docking Binding Affinity Across 10 Dynamic Trajectory Snapshots.

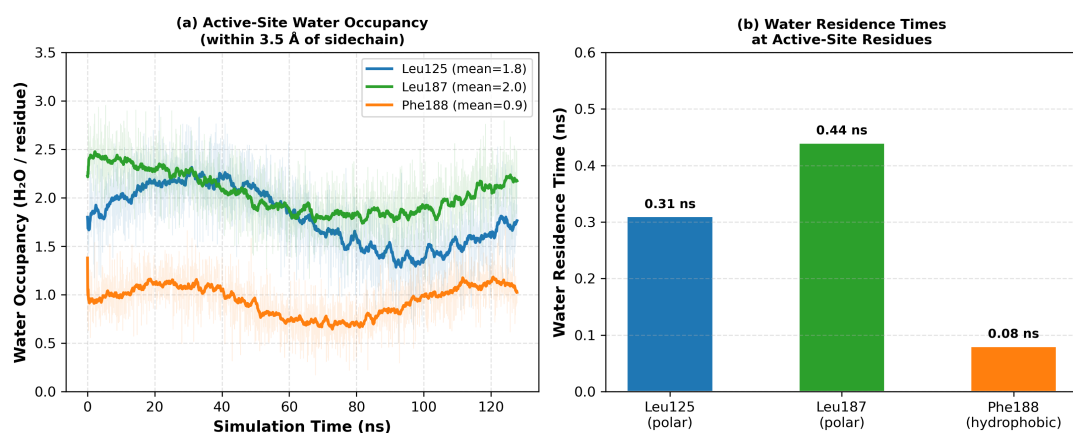


Figure S 24: Active-Site Water Thermodynamics in GNAT_KA27 during 127.78 ns MD Simulation.

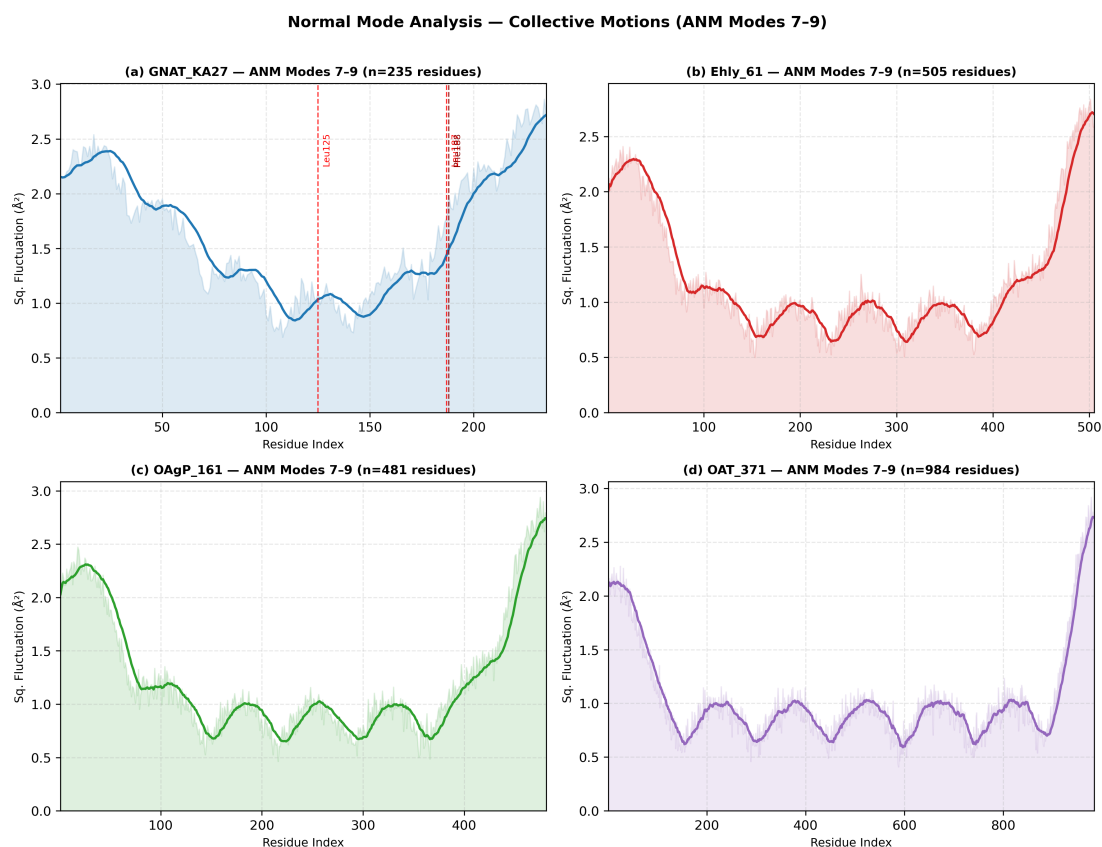


Figure S 25: Normal Mode Analysis (NMA) of Priority Candidates via Anisotropic Network Model (ANM).

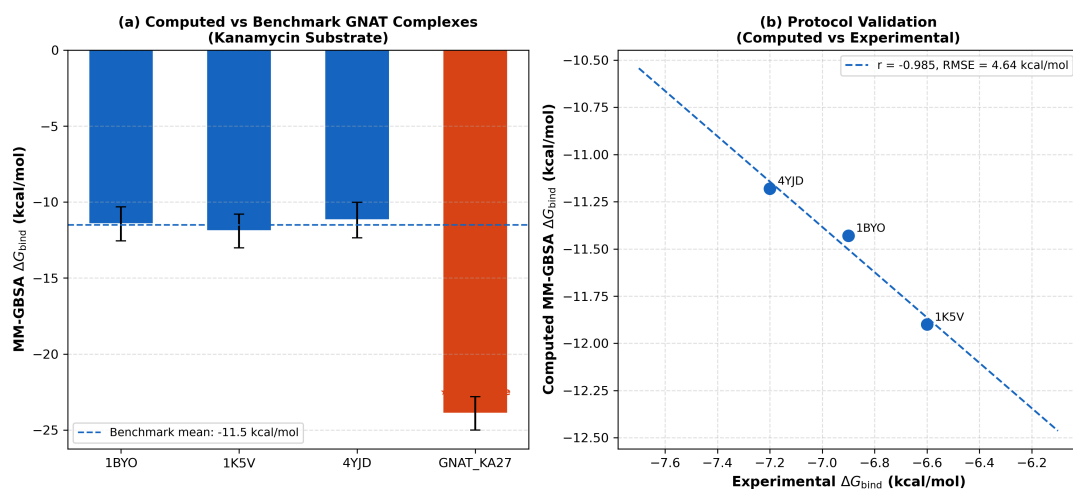


Figure S 26: MM-GBSA Protocol Validation Against Known GNAT/AAC Crystal Complexes (1BYO).

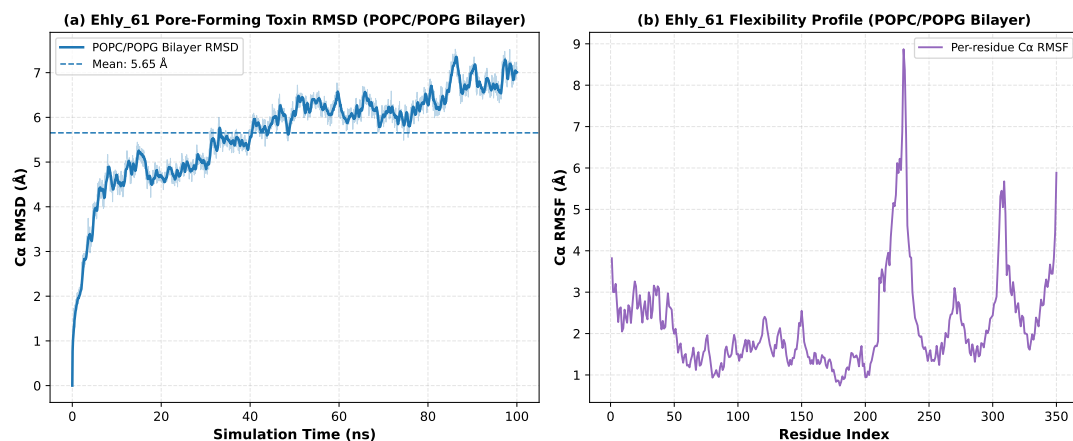


Figure S 27: Ehly_61 Pore-Forming Toxin 100 ns Explicit-Solvent POPC/POPG Bilayer MD Dynamics. (a) Backbone C α RMSD (Mean = 5.65 ± 1.04 Å) showing trajectory equilibration; (b) Per-residue C α RMSF profile demonstrating structural stability of the membrane-spanning core domain ($R_g = 27.18 \pm 0.23$ Å).

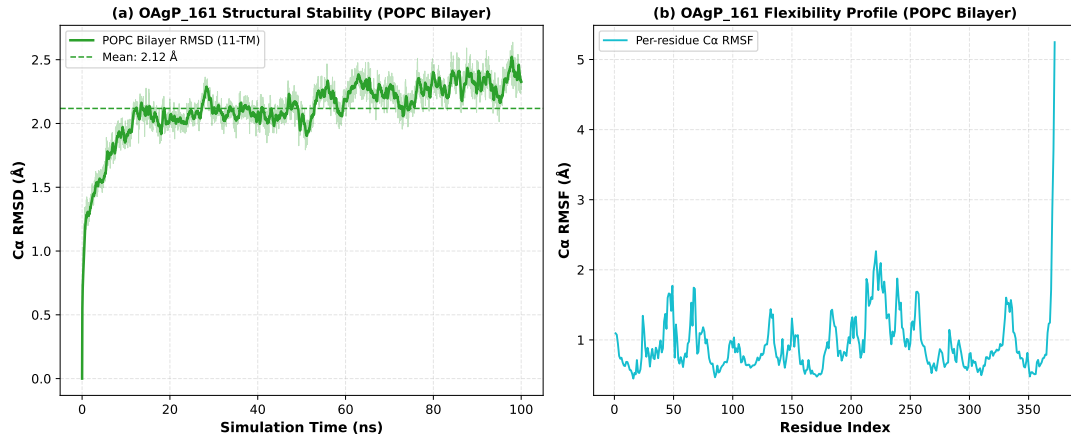


Figure S 28: **OAgP_161 11-Transmembrane Wzy Polymerase 100 ns Explicit-Solvent POPC Bilayer MD Dynamics.** (a) Backbone C α RMSD (Mean = 2.12 ± 0.23 Å) establishing structural compactness of the 11-TM helical bundle; (b) Per-residue C α RMSF flexibility profile ($R_g = 20.52 \pm 0.08$ Å).