Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Nebulized Covalent Killing of Macrolide-Resistant Mycoplasma pneumoniae: Computational Target Prediction and Design of a Proximity-Triggered Phosphorus(V) Conjugate

Chen Shen

shenchen1110@126.com

Beijing Children's Hospital

Deze Li

Beijing Children's Hospital

Research Article

Keywords: Mycoplasma pneumoniae, P116, proximity-triggered covalent conjugate, architecture inversion, tyrosine trigger, structural bioinformatics, macrolide resistance

Posted Date: November 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11654766/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Nebulized Covalent Killing of Macrolide-Resistant Mycoplasma pneumoniae: Computational Target Prediction and Design of a Proximity-Triggered Phosphorus(V) Conjugate

Chen Shen[1]*, Deze Li[1]

[1] Beijing Children's Hospital, Capital Medical University, Beijing, China

*Correspondence: Chen Shen, shenchen1110@126.com

Abstract

Macrolide-resistant *Mycoplasma pneumoniae* (MRMP) has reached crisis prevalence—97.1% of 421 isolates in a 2023–2024 multicenter survey in China and more than 90% across twenty provinces—while the pediatric formulary remains confined to three antibiotic classes that all act on intracellular housekeeping machinery. No mechanism-diverse drug exists for this pathogen, in part because covalent anchoring strategies have lacked trigger-site knowledge for bacterial surface proteins. Here we address that gap computationally through a proteome-scale tyrosine-trigger screen of thirteen *M. pneumoniae* surface proteins (82 qualifying hits), corroboration against experimental cryo-electron-microscopy structures (PDB 8A9A and 9GVG), antibody-epitope and pocket mapping, and conservation analysis across four complete genomes spanning both dominant P1 genotypes (P1-1 and P1-2). The screen identified the essential lipid-acquisition protein P116 as the sole Tier-1 target. Four trigger tyrosines (Tyr671, Tyr566, Tyr188, Tyr206) display phenol-hydroxyl solvent-accessible surface areas of 35.2–43.5 Å² in the physiological dimer, and Tyr398 resides inside a validated monoclonal-antibody epitope (residue SASA 52.4 Å² on the isolated P116 chain of the 9GVG complex; 31.9 Å² in the antibody-bound state). A 533 Å³ subsidiary pocket lies only 2.7 Å from Tyr671, but its entrance aperture measures only 4.0–4.5 Å—too narrow for any aromatic ring (in-plane widths: benzene 7.02 Å, naphthalene 7.36 Å)—which forced an architecture inversion: the recognition group was moved from the anchor end of the conjugate to the leaving-group side of the electrophilic warhead, so that binding occurs in a solvent-side groove without crossing any aperture. The resulting lead, K2 ($C_{29}H_{27}FN_2O_8PS^-$, monoisotopic mass 613.1215 Da; two-carbon tether), achieves a qualifying docked-pose fraction of 0.85 (range 0.80–0.90 across four random seeds; 0.78 on an independent seed set), with a median warhead-P···Tyr671-OH distance of 5.60 Å, whereas the warhead-inactivated control K2-N scores 0.00 and the two top-scoring

compounds by docking score (−10.62 and −10.45 kcal/mol) produce zero qualifying poses—motivating ranking by pose fraction rather than docking score. Successful exchange of K2 is predicted to generate a +375.0421 Da covalent adduct, a proposed intact-protein mass-spectrometry readout. All eighteen trigger sites are invariant across the four genomes. Because the organism colonizes the airway lumen extracellularly, the design is compatible with nebulized local delivery at the site of infection. This work proposes the first proximity-triggered covalent antibacterial strategy and delivers a conservation-anchored target map together with synthetically tractable, architecture-inverted conjugate designs, providing a defined experimental path toward mechanism-diverse agents against MRMP.

Keywords: *Mycoplasma pneumoniae*; P116; proximity-triggered covalent conjugate; architecture inversion; tyrosine trigger; structural bioinformatics; macrolide resistance

1. Introduction

1.1 The macrolide-resistance crisis and the mechanism ceiling

Mycoplasma pneumoniae is a leading cause of community-acquired pneumonia, accounting for 10–40% of cases in children, and its clinical management now faces a structural antibiotic dilemma [1]. A multicenter retrospective analysis of 421 *M. pneumoniae* isolates collected from four regions of China between November 2023 and January 2024 reported macrolide resistance in 97.1% of samples, driven predominantly by the A2063G substitution in the 23S ribosomal RNA gene [2]. A nationwide cross-sectional study of more than 1.6 million cases showed that the proportion of macrolide-resistant *M. pneumoniae* (MRMP) rose from 80.00% in July 2023 to 93.02% in May 2024, with 20 provinces exceeding 90%, a trajectory that contrasts sharply with the 5.3% increase over the preceding nine years [1].

The therapeutic consequences are amplified by the demographic profile of the disease. Children constitute the principal patient group, yet the antibiotic formulary active against this wall-less organism is restricted to three classes: macrolides, tetracyclines, and fluoroquinolones [1]. Tetracyclines are contraindicated in children younger than eight years because of dental and skeletal toxicity, and fluoroquinolones are used with caution in pediatric populations because of cartilage concerns, leaving macrolides as the de facto first-line and frequently the only well-tolerated option for the most affected age group [2, 3]. When the dominant genotype carries high-level macrolide resistance, the effective formulary therefore collapses toward a single, age-restricted margin. Because all three remaining classes act on intracellular housekeeping machinery and resistance mutations are already fixed in the circulating P1-1 lineage,

incremental medicinal chemistry within these scaffolds addresses neither the mechanism ceiling nor the resistance already present [1, 2]. New antibacterial mechanisms that engage targets orthogonal to the ribosome are accordingly needed.

1.2 P116 as an essential surface target

The P116 protein (gene *mpn_213*) offers a rare combination of essentiality, surface accessibility, and structural definition among *M. pneumoniae* proteins. Single-particle cryo-electron microscopy (cryo-EM) revealed that P116 is a homodimer with a previously unobserved fold, spanning approximately 240 Å in an arch-like architecture and enclosing an exceptionally large hydrophobic cavity—approximately 18,000 Å³ (1.8×10^4 Å³) in the lipid-filled state—which is fully accessible to solvent and extracts cholesterol, phosphatidylcholine, and sphingomyelin from host lipoproteins such as high-density lipoprotein (HDL) [4]. P116 is essential for viability, is encoded by a single genomic copy, and is present at only approximately 34 copies per cell (as compiled in [4]), yet it elicits the second strongest antigenic response among *M. pneumoniae* proteins after the P1 adhesin [4]. The authors of the structure determination themselves proposed that interference with lipid uptake could control *Mycoplasma* infection and identified P116 as a candidate vaccine target [2].

Despite this explicit invitation, drug-like progress against P116 has been limited. The principal advance is monoclonal antibody (mAb) 3G9, which binds the C-terminal domain of P116, impairs cholesterol acquisition, and produced the strongest growth inhibition in a panel of anti-P116 antibodies, though suppression remained incomplete; the same study showed that the W681A substitution reduces low-density lipoprotein (LDL) cholesterol uptake by 60% [4]. The action of mAb 3G9 is, however, reversible and its growth inhibition incomplete, and no small molecule, antibody–drug conjugate, or covalent conjugate directed at P116 has been reported [4]. P116 therefore represents a structurally mature but mechanistically underexploited target: its lipid-uptake function is essential, its surface display is uniform across the cell, and its only pharmacological validation to date derives from a noncovalent biologic [4, 5].

1.3 Proximity-triggered covalent chemistry as a new mechanism space

Covalency is a historically validated antibacterial principle: the β-lactams, which remain the most successful antibiotic class, act through acylation of the active-site serine of penicillin-binding proteins, and the broader resurgence of covalent drugs has demonstrated that irreversible engagement can confer durable target occupancy with acceptable selectivity [6]. Two recent developments extend covalent pharmacology beyond enzyme active sites. First, Wen and colleagues introduced a binding-to-release (BTR) strategy in which an electrophile positioned by noncovalent target binding is cleaved directly by a

proximal nucleophilic residue within the binding pocket, thereby decoupling payload release from receptor-mediated endocytosis; the strategy was realized through phosphorus(V)–phenol exchange (PhoPEx) chemistry and validated on the tumor-associated targets fibroblast activation protein (FAP) and programmed cell death ligand 1 (PD-L1) [7]. Second, the proximity-enabled reactive therapeutic (PERx) platform established that genetically encoded latent bioreactive amino acids, including fluorosulfate-l-tyrosine (FSY) and its accelerated successor FFY, convert noncovalent protein binders into covalent drugs [8, 9]. Applied to nanobodies against severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), FFY incorporation enabled cross-linking of the non-catalytic receptor-binding-domain residue Tyr369 within approximately 10 minutes at physiological pH, demonstrating that surface tyrosines lacking catalytic activation can serve as covalent triggers when a suitably positioned warhead is presented [9].

Neither chemistry has been applied to bacteria. BTR has been demonstrated on small-molecule and peptide conjugates against mammalian tumor targets, and PERx requires genetic code expansion confined to engineered biologics against viral antigens [7, 9]. To our knowledge, covalent targeting of a bacterial surface protein through proximity-triggered exchange chemistry remains unexplored, despite its conceptual fit with extracellular airway colonizers such as *M. pneumoniae*, whose essential surface machinery is continuously exposed to the airway lumen.

1.4 The present work

This paper proposes that the unexploited space between BTR-type covalent chemistry and bacterial surface biology can be opened computationally by answering three questions: which *M. pneumoniae* surface proteins carry exploitable tyrosine triggers, where are those triggers located relative to antibody epitopes and ligandable pockets, and can they be reached by synthetically tractable conjugates. To this end, we performed a proteome-scale trigger screen across 13 *M. pneumoniae* surface proteins using solvent-accessible surface area (SASA) analysis with a custom Shrake–Rupley implementation, hits corroborated against experimental cryo-EM structures of P116 under a dual-track structural-confidence policy, mapped the mAb 3G9 epitope and a tyrosine-adjacent subsidiary pocket, and—when pocket-aperture measurements invalidated direct pocket-wedge designs—developed an architecture-inverted phosphorus(V) warhead conjugate series whose lead, K2, carries its recognition group on the leaving-group side of the warhead. Strain-level conservation was assessed across four complete genomes. The contributions of this study are fourfold: (i) a reproducible screen identifying P116 as the only candidate supported by experimental structures, a validated antibody epitope, and a ligandable surface region; (ii) a computationally prioritized set of surface-exposed trigger tyrosines (Tyr671, Tyr566, Tyr188, Tyr206)

plus the epitope-resident Tyr398; (iii) a 533 Å³ subsidiary pocket adjacent to Tyr671 whose narrow entrance aperture (4.0–4.5 Å) motivated an architecture-inverted small-molecule design (lead K2, with parallel candidate K1, warhead-inactivated controls K2-N/K1-N, and fluorescent reporters K3/K4) whose predicted covalent adduct (+375.0421 Da for K2) is detectable in principle by intact-protein mass spectrometry; and (iv) evidence that all 18 trigger sites are fully conserved across four *M. pneumoniae* strains spanning the P1-1 and P1-2 genotypes. The remainder of this paper is organized as follows: Section 2 reports the screen, its structural corroboration against experimental structures, pocket-aperture analysis, architecture-inverted design, and conservation analysis; Section 3 discusses mechanism, selectivity, delivery, and limitations; and Section 4 details all computational methods.

2. Results

2.1 A Proteome-Scale Screen Localizes Trigger Tyrosines on the *M. pneumoniae* Surface

Proximity-triggered covalent chemistries, exemplified by the phosphorus(V)–phenol exchange (PhoPEx) warhead recently validated in the binding-to-release (BTR) drug-delivery paradigm, release or lock a payload only when a target-bound ligand places the phosphorus center within reach of a surface-exposed, non-catalytic tyrosine [7]. The feasibility of extending this mechanism to a bacterial surface protein therefore hinges on a single geometric question: which proteins on the *M. pneumoniae* cell surface present tyrosine phenol groups that are sufficiently solvent-exposed to serve as covalent triggers? To answer this question systematically, we performed a proteome-scale computational screen of thirteen *M. pneumoniae* surface and secreted proteins (Table 1; Figure 1). Solvent-accessible surface areas (SASA) were computed with an in-house implementation of the Shrake–Rupley rolling-ball algorithm (1.4 Å probe, 192 points per atom) [10] on AlphaFold v4 models retrieved through the AlphaFold Protein Structure Database [11], with accessions taken from UniProtKB [12]. Under a dual-track structural-confidence policy, a residue was scored as a trigger hit when its whole-residue SASA was ≥40 Å² and—because the screen rests on predicted models—its Cα pLDDT confidence score was ≥70; the N-terminal 25 residues were masked in all sequences to exclude signal peptides and lipobox motifs that are proteolytically removed or membrane-embedded in the mature proteins. A glycosylation-shielding filter was additionally applied at the sequence level: candidate triggers adjacent to potential N-linked glycosylation sequons (N-X-S/T, X ≠ P) would be excluded as potentially glycan-shielded. We note that *M. pneumoniae* lacks a classical (PglB-type) N-linked glycosylation pathway, and the mycoplasma protein modifications reported in the literature are motif-less, stochastic mono-hexosylations that do not form shielding glycan chains; accordingly, no candidate trigger tyrosine was excluded by this filter in the present panel, and the filter is retained for generalization of the screening method to other pathogens. The trigger-tyrosine

counts and tier assignments for all thirteen proteins are summarized in Table 1, and the distribution of hits across the panel is visualized in Figure 1.

Table 1. Trigger-tyrosine screen of thirteen *M. pneumoniae* surface proteins (hits: residue SASA ≥ 40 Å² and Cα pLDDT ≥ 70 on AlphaFold v4 models; N-terminal 25 residues masked).

Protein (gene)	Trigger hits	Structural basis	Tier
MPN372 (CARDS toxin)	12	AlphaFold model	2
MPN474 (lipoprotein)	11	AlphaFold model	—
MPN142 (P40/P90)	10	AlphaFold model; X-ray (6TLZ, 3.1 Å) for N-domain	2
MPN133 (nuclease)	8	AlphaFold model	—
MPN284	7	AlphaFold model	—
MPN288	7	AlphaFold model	—
MPN281	6	AlphaFold model	—
MPN491	5	AlphaFold model	—
MPN052	4	AlphaFold model	—
MPN213 (P116)	4	Cryo-EM (8A9A, 3.3 Å; 9GVG, 5.06 Å) + AlphaFold	1
MPN456	4	AlphaFold model	—
MPN388 (IbpM)	3	AlphaFold model	—
MPN309 (P65)	1	AlphaFold model	—

Total qualifying hits across the panel: 82. For MPN142 (P40/P90), the 10 hits derive from 17 surveyed tyrosines (10/17); the CARDS toxin shows the highest single-residue exposure of the panel (maximum residue SASA 187 Å²).

Table 1 shows that trigger-tyrosine density varies by more than an order of magnitude across the surface proteome, from twelve qualifying residues in the CARDS toxin to a single hit in P65. The low hit count of P65 is consistent with its predominantly disordered architecture: extended, low-confidence regions yield few residues that simultaneously satisfy the exposure and pLDDT filters. Integrating experimental structure availability, essentiality, and the existence of validated binders, P116 emerged as the sole Tier-1 target: it is essential and single-copy, its cryo-EM structures are available in lipid-bound and antibody-bound states (PDB 8A9A and 9GVG), and monoclonal antibodies against it are functionally validated [4, 5]. The P40/P90 adhesion-complex subunit (MPN142), whose N-terminal domain and sialyllactose complexes have been crystallographically characterized [13], and the CARDS toxin (MPN372), an ADP-ribosylating and vacuolating virulence factor of established structure and pathogenic importance [14], were assigned to Tier-2: both are attractive secondary targets whose screen hits currently rest on AlphaFold-only models across their full lengths. The remaining ten proteins, which combine model-only structures with lower hit counts, unestablished essentiality, or no known binders, were not advanced.

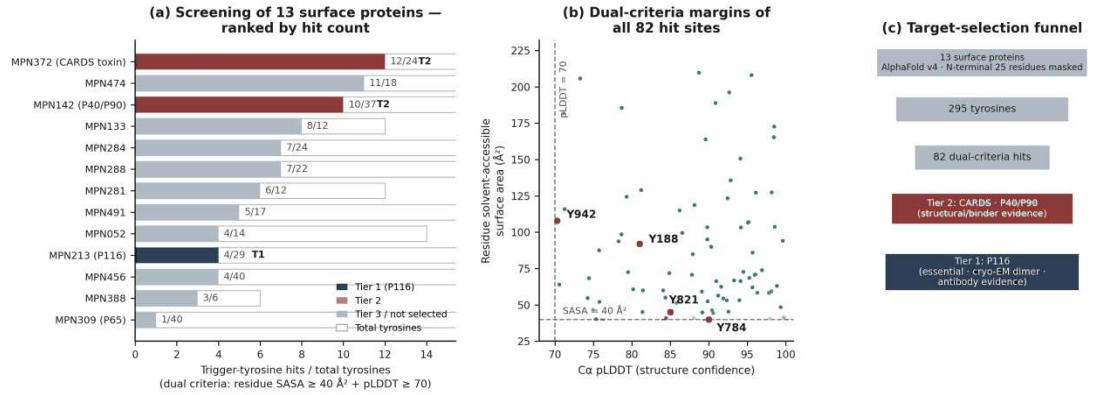


Figure 1. Proteome-scale trigger-tyrosine screen of thirteen M. pneumoniae surface proteins.

(a) Number of tyrosine residues per protein passing the dual filter of residue SASA ≥ 40 Å² and Cα pLDDT ≥ 70 (Shrake–Rupley algorithm, 1.4 Å probe, 192 points per atom, AlphaFold v4 models, N-terminal 25 residues masked), ranked by hit count: MPN372 (12), MPN474 (11), MPN142 (10), MPN133 (8), MPN284 and MPN288 (7), MPN281 (6), MPN491 (5), MPN052, MPN213 (P116), and MPN456 (4), MPN388 (3), and MPN309 (1). (b) Dual-filter margins of all 82 passing hits (residue SASA versus Cα pLDDT), with the four P116 hits annotated. (c) Target-selection funnel: tier assignment integrates experimental structure availability, essentiality, and the existence of validated binders, identifying P116 as the sole Tier-1 target despite its intermediate raw hit count.

2.2 Dual-Track Structural Corroboration of P116 Trigger Tyrosines in the Dimer Context

P116 is an essential, single-copy protein that extracts cholesterol, phosphatidylcholine, and sphingomyelin from host lipoproteins; cryo-EM has shown it to be a homodimer with an approximately 240 Å arch-like span enclosing a huge solvent-accessible hydrophobic cavity, present at only ~34 copies per cell (as compiled in [4]). Because the screen of Section 2.1 was performed on monomeric AlphaFold models, we next asked whether the candidate trigger tyrosines remain exposed in the physiological dimer. Structural confidence was tracked on two rails: AlphaFold models were used only where the local Cα pLDDT was ≥70, and experimental structures were qualified by resolution—no worse than 3.5 Å for SASA geometry measurements and no worse than 5.5 Å for epitope or binding-region localization, with antigen–antibody complex structures used only for epitope localization and exposure-context analysis. Phenol-hydroxyl (atom-level) SASA values were therefore computed directly on the lipid-bound dimer 8A9A (3.3 Å resolution; chains A/B, residues 60–868), which satisfies the ≤3.5 Å requirement for geometric measurement, while the Fab-bound complex 9GVG (5.06 Å resolution) was used to localize the 398–446 epitope and to analyze the exposure context of Tyr398 [4, 15]. Because the PhoPEx

mechanism engages the phenol oxygen specifically, we adopted an atom-level decision threshold of 35 Å² for the hydroxyl oxygen SASA, distinct from the 40 Å² residue-level threshold used in the initial screen. The complete geometry, together with the strain-conservation status established in Section 2.5, is compiled in Table 2; the spatial distribution of the prioritized triggers on the dimer is shown in Figure 2, and the quantitative accessibility ranking is presented in Figure 3.

Table 2. Geometry, assessment, and conservation of P116 trigger tyrosines (phenol-OH SASA and residue SASA computed on the 8A9A dimer; decision threshold 35 Å² at the atom level).

Trigger	Phenol-OH SASA, dimer (Å²)	Residue SASA, dimer (Å²)	≥35 Å² verdict	Pocket within 6 Å	Overall assignment	Conservation (4 genomes)
Tyr671	43.5	56.7	Pass	Yes (533 Å³ subsidiary pocket, 2.7 Å)	Small-molecule trigger (K-series design)	4/4; 0 mismatches in ±30 aa
Tyr566	41.9	74.3	Pass	Yes (~139 Å³ cavity)	Small-molecule trigger (passes dual criteria)	4/4; 0 mismatches in ±30 aa
Tyr188	36.8	99.5	Pass	No (no pocket within 6 Å)	Excluded from small-molecule route	4/4; 0 mismatches in ±30 aa
Tyr206	35.2	66.6	Pass	Yes (~272 Å³ pocket, 3.1 Å)	Small-molecule trigger (passes dual criteria)	4/4; 0 mismatches in ±30 aa
Tyr398	19.0	26.5	Fail	—	Epitope-resident; residue SASA 52.4 Å² on the isolated P chain (antibody removed) and 31.9 Å² in the antibody-bound state; antibody route	4/4; 0 mismatches in ±30 aa
Tyr821	17.3	73.3	Fail	—	Excluded	4/4; 0 mismatches in ±30 aa

Table 2 supports three conclusions. First, four tyrosines—Tyr671, Tyr566, Tyr188, and Tyr206—pass the 35 Å² atom-level accessibility threshold in the lipid-bound dimer (35.2–43.5 Å²); of these, Tyr671, Tyr566, and Tyr206 also pass the pocket criterion and constitute the primary small-molecule covalent-

trigger set. Second, exposure alone is not sufficient for small-molecule design: Tyr188, although well exposed, has no detectable pocket within 6 Å of its phenol oxygen in any parameter combination tested and is therefore excluded from the small-molecule route, remaining a valid anchor for antibody-format engagement; Tyr566 and Tyr206 also satisfy both criteria, with a ~139 Å³ cavity detected at Tyr566 and a ~272 Å³ pocket 3.1 Å from Tyr206, so that Tyr671, Tyr566, and Tyr206 all pass the dual requirement of phenol-OH SASA ≥ 35 Å² and a detectable pocket within 6 Å. Third, Tyr398 (19.0 Å²) and Tyr821 (17.3 Å²) fail the hydroxyl-accessibility threshold in the unliganded dimer; Tyr821 is excluded, whereas Tyr398 is analyzed separately in Section 2.3 because it resides inside a structurally validated antibody footprint, displaying a residue-level SASA of 52.4 Å² for the isolated P116 chain of the 9GVG model (antibody-free context) and remaining partially exposed (31.9 Å²) in the antibody-bound state. Independent recomputation on the AlphaFold model of P116 reproduced the exposed-candidate rankings obtained on the experimental dimer, indicating that the AlphaFold-versus-cryo-EM cross-validation is mutually consistent for the exposed candidates, while only the experimental dimer resolves the dimer-context exposure of the phenol oxygens. These analyses indicate that proximity-triggered engagement of P116 is geometrically feasible at the Tyr671-centered subsidiary pocket for a small-molecule conjugate and at Tyr398 (and Tyr188) for antibody-format conjugates.

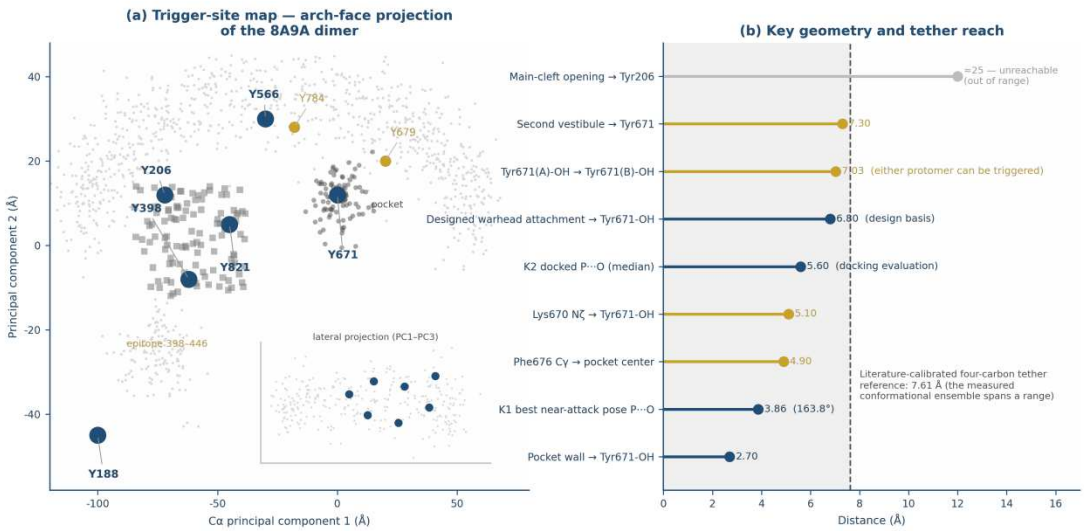


Figure 2. Spatial distribution of prioritized trigger tyrosines on the P116 dimer.

(a) Arch-face projection (Cα principal components) of the 8A9A lipid-bound homodimer (dimer span ~240 Å), with a side-projection inset; trigger tyrosines are shown as circles sized by phenol-hydroxyl SASA in the dimer (Tyr671 43.5, Tyr566 41.9, Tyr188 36.8, Tyr206 35.2, Tyr398 19.0, Tyr821 17.3 Å²), the Tyr671-adjacent 533

Å³ subsidiary pocket in dark gray, and the 9GVG mAb epitope (residues 398–446) as squares. (b) Key distances of the design geometry, ranked on a 0–16 Å scale with the main-cleft opening–Tyr206 distance (~25 Å) shown as an unreachable reference: pocket wall–Tyr671-OH 2.70 Å; Phe676 Cγ–pocket center 4.90 Å; Lys670 Nζ–Tyr671-OH 5.10 Å; docked K2 P···O median 5.60 Å; design anchor–Tyr671-OH 6.00 Å (design basis); cross-subunit Tyr671-OH···OH 7.03 Å; second vestibule–Tyr671 7.30 Å. Vertical markers indicate the 6.00 Å design basis and a literature calibration reference of 7.61 Å for a four-carbon tether.

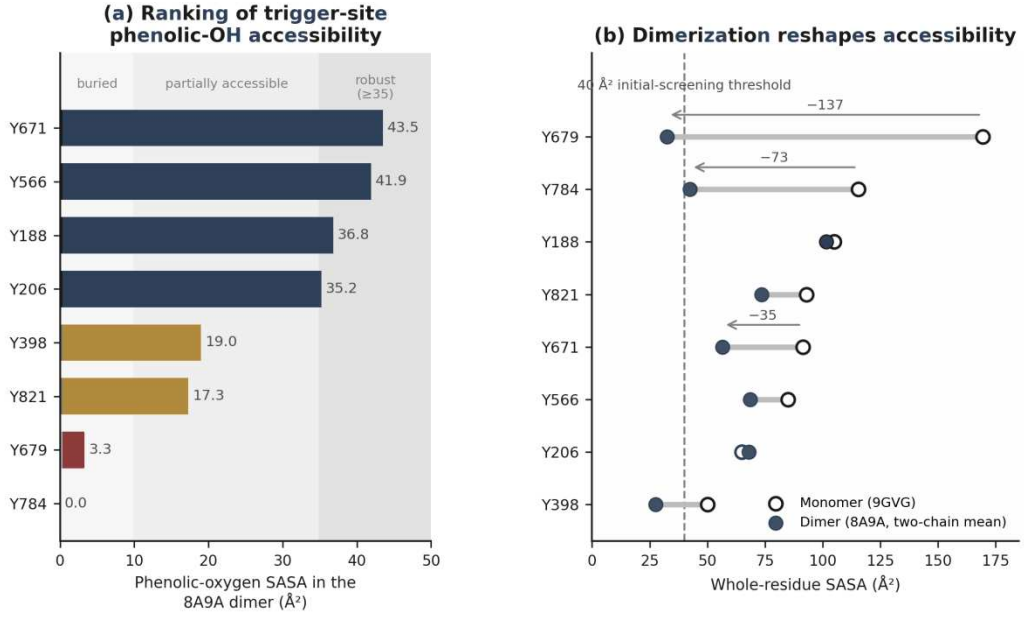


Figure 3. Quantitative assessment of P116 trigger-tyrosine accessibility.

(a) Phenol-hydroxyl SASA of the six candidate trigger tyrosines computed on the 8A9A dimer, rank-ordered against the 35 Å² decision threshold: Tyr671 43.5, Tyr566 41.9, Tyr188 36.8, Tyr206 35.2, Tyr398 19.0, and Tyr821 17.3 Å². (b) Corresponding whole-residue SASA values on the dimer (Tyr671 56.7, Tyr566 74.3, Tyr188 99.5, Tyr206 66.6, Tyr398 26.5, Tyr821 73.3 Å²), illustrating that residue-level and atom-level exposure rankings diverge and that the atom-level measurement is the relevant criterion for warhead accessibility. The 9GVG-derived values in this panel were computed on the P116 chain alone, after removal of the Fab heavy/light chains and of the 654–700 P116 fragment present in that deposit. A companion data table reproducing all values is provided in Table 2.

2.3 Tyr398 Resides within a Validated Antibody Epitope

A proximity-triggered covalent conjugate requires not only an accessible trigger residue but also a delivery vehicle that positions a warhead adjacent to it. The 9GVG structure, a complex of P116 with a monoclonal Fab (resolution 5.06 Å, used here for epitope localization and exposure context), provides exactly such a template for the Tyr398 region [4]. The epitope maps to residues 398–446, with Tyr398 located within this region. Tyr398 is the N-terminal-most residue of this epitope and, at 52.4 Å² residue-level SASA for the isolated P116 chain of the 9GVG model (31.9 Å² in the antibody-bound state, i.e., still partially exposed within the bound footprint), is the only trigger candidate that lies directly inside a structurally validated antibody footprint (Table 2; Figure 2). Although its phenol hydroxyl is only partially exposed in the unliganded dimer (19.0 Å²), an antibody-bound state could present the aromatic ring in an orientation compatible with warhead exchange; this hypothesis remains to be tested with bound-state measurements.

The functional relevance of antibody binding to P116 has been established independently. The 9GVG-deposited antibody has been described in the literature as the strongest inhibitor in a panel of anti-P116 antibodies, though with incomplete growth suppression [4], and a more extensive recent study of the C-terminal-domain-directed monoclonal antibody 3G9 showed that epitope engagement on P116 inhibits cholesterol acquisition from radiolabeled LDL, suppresses mycoplasma growth in culture, and reduces localization of the pathogen to human atherosclerotic plaque tissue ex vivo [5]. The same study localized the 3G9 epitope by cryo-EM to four α-helices forming two of the fingers that define the hydrophobic cavity of P116, and demonstrated by site-directed mutagenesis that the W681A substitution reduces LDL-derived cholesterol uptake by approximately 60% [5]. Taken together, these observations indicate that two independent, experimentally validated binders place their contact surfaces directly on trigger tyrosines identified by our screen—Tyr398 within the 398–446 Fab epitope, and the Tyr671/Trp681 region within the functionally critical C-terminal domain.

These precedents carry a direct design implication. Reversible binders such as 3G9 achieve only partial, occupancy-dependent inhibition, whereas the same targeting elements, if equipped with a proximity-triggered PhoPEx warhead, could convert mild reversible suppression into irreversible covalent inhibition by locking onto Tyr398 or Tyr671 upon binding [7]. The feasibility of such conversion is supported by the PERx (proximity-enabled reactive therapeutics) platform, in which genetically encoded latent bioreactive nanobodies cross-linked a non-catalytic tyrosine (Tyr369) of the SARS-CoV-2 spike RBD within 5–10 min at physiological pH, transforming reversible nanobodies into covalent inhibitors with up

to ~40-fold potency gains [9]. The P116 system differs in that the warhead would be installed by total chemical synthesis rather than genetic code expansion, and the target is an essential bacterial surface protein rather than a viral antigen.

2.4 Pocket-Aperture Analysis Imposes an Architecture Inversion and Yields the Lead Conjugate K2

For a small-molecule (as opposed to antibody-format) trigger engagement, the warhead-bearing ligand must first bind within tether reach of the trigger phenol. An initial cavity survey of the P116 dimer identified the main interdomain cleft, whose grid-estimated volume of ~11,600 Å³ is of the same order as the huge solvent-accessible hydrophobic cavity described experimentally in the lipid-filled state (~1.8×10^4 Å³) [4]. However, the cleft opening lies ~25 Å from the nearest exposed trigger, Tyr206—far beyond the reach of any short tether—so the main cleft was judged geometrically unreachable for trigger-directed design. A focused pocket search (1.5 Å grid, six-axis occlusion count ≥5, 10 Å search radius around each validated trigger) resolved a 533 Å³ subsidiary pocket whose nearest wall lies only 2.7 Å from the Tyr671 phenol oxygen (Figure 4), as well as a second, shallower vestibule 7.30 Å from Tyr671, a ~139 Å³ cavity at Tyr566, and a ~272 Å³ pocket 3.1 Å from Tyr206. The two protomers' Tyr671 phenol oxygens sit 7.03 Å apart across the subunit interface, so a single bound warhead can in principle be triggered from either chain. The pocket region offers several anchoring features: Lys670 Nζ at 5.10 Å from the Tyr671 phenol oxygen, Phe676 Cγ at 4.90 Å from the pocket center, and a design anchor point at 6.00 Å from the Tyr671 phenol oxygen that was adopted as the design basis for warhead placement.

Aperture analysis invalidates direct pocket-wedge designs. Before committing to a classic wedge-in design—an aromatic recognition group inserted into the buried pocket—we computed the pocket entrance geometry. The entrance admits a probe radius of only 2.0–2.25 Å, corresponding to an aperture diameter of 4.0–4.5 Å (narrowest-section free area 2.56–3.19 Å²), and the narrowest section is near-circular (axis ratio <1.5), so a traversing group must fit in two orthogonal dimensions simultaneously. van der Waals envelope calculations on aromatic fragments (RDKit; ETKDGv3 embedding, MMFF force field, 200 conformers per fragment) show that even the smallest aromatic ring, benzene, has an in-plane width of 7.02 Å, and naphthalene 7.36 Å—both far exceeding the widest computed aperture. Because circular-aperture passage requires satisfying both orthogonal directions at once, no aromatic ring can cross the entrance in the static structure, and any direct wedge-in architecture is therefore infeasible.

Architecture inversion. This geometric impasse motivated an inversion of the conjugate architecture: the recognition group was moved from the anchor end of the conjugate to the leaving-group side of the electrophilic warhead, so that the leaving group itself carries the recognition group. The binding site

thereby relocates from the buried pocket to a solvent-side groove adjacent to Tyr671, and the molecule no longer needs to cross any aperture. The resulting lead compound, **K2** (Figure 4), comprises a 4-sulfobenzamide anchor, a two-carbon (C2) tether, and an ethyl phosphinate PhoPEx warhead bearing a 3-fluoro-4-(naphthalen-2-yl)phenoxy leaving group whose biaryl arm provides the groove-recognition element: *N*-(2-(4-sulfobenzamido)ethyl)-2-(ethoxy(3-fluoro-4-(naphthalen-2-yl)phenoxy)phosphoryl)acetamide ($C_{29}H_{27}FN_2O_8PS^-$; monoisotopic mass 613.1215 Da; racemic at phosphorus). Tether length was treated as a design variable rather than a fixed constant: conformer-ensemble sampling (~400 conformers per tether) of straight-chain alkylene tethers of two to six carbons showed that each tether spans a continuous distance interval—the four-carbon tether covers 4.06–10.67 Å with a 95th percentile of 10.24 Å—so the tether was selected against the computed target geometry (most key distances fall within 2.70–7.30 Å, clustered near the 6.00 Å design basis), yielding the C2 tether for K2; a literature calibration reference of 7.61 Å for a four-carbon tether is shown in the figures for orientation only [7].

Docking evaluation and pose-fraction ranking. Candidate conjugates were docked against the 8A9A homodimer with AutoDock Vina 1.2.7 (25 Å box, exhaustiveness 16, four independent random seeds, 20 poses per seed; Section 4.7), and candidates were ranked by the fraction of qualifying poses—warhead phosphorus within 6.0 Å of the Tyr671 phenol oxygen—rather than by docking score. The architecture-inverted lead K2 achieved a qualifying-pose fraction of 0.85 (range 0.80–0.90 across seeds), reproduced at 0.78 (range 0.70–0.90) on an independent seed set (42/7/1234/2024), with a median P···O distance of 5.60 Å. Two controls validate the readout. First, the warhead-inactivated analog K2-N, in which the phosphorus is joined to the aryl group through a phosphorus–carbon bond and no exchangeable phosphorus–aryloxy bond exists, gave a qualifying-pose fraction of 0.00. Second, the two compounds with the best Vina docking scores in the candidate set (−10.62 and −10.45 kcal/mol) produced zero qualifying poses, demonstrating that docking score and reactive feasibility can diverge and justifying ranking by qualifying-pose fraction (pass threshold ≥0.5) rather than by score. The parallel candidate K1, which retains the conventional anchor-end architecture (a 6-carboxynaphthalen-2-yloxyacetyl recognition element at the anchor end, four-carbon tether, ethyl phosphinate warhead with a 3-fluorophenoxy leaving group; $C_{27}H_{30}FN_2O_8P$, monoisotopic mass 560.1724 Da), achieved a qualifying-pose fraction of 0.60 and produced one near-attack conformation (P···O 3.86 Å, O–P–O angle 163.8°, approximating in-line SN2@P attack); K1 is retained as the anchor-end-architecture comparator. We note explicitly that the Tyr671-adjacent groove has no ligand-complex structure and that its ligandability, while suggested by the 533 Å^3 subsidiary volume, the docking pose statistics reported here, and the documented anti-infective chemical space catalogued in ChEMBL [16], remains a rational, shape-based hypothesis pending experimental confirmation.

Reporters, controls, and analogs. Two fluorescent reporter analogs complete the series (Table 3). K3 ($C_{26}H_{30}N_2O_{10}PS^-$, 593.1364 Da) carries a 4-methylumbelliferone-7-oxy leaving group: exchange with the target releases 4-methylumbelliferone, providing a 360/460 nm fluorescence readout of release kinetics. K4 ($C_{30}H_{38}FN_4O_7P$, 616.2462 Da) carries an ethylenediamine handle on the leaving arm for conjugation of near-infrared dyes, serving as a conjugation-tracing probe for cell-surface labeling studies. K4 was designed as a near-infrared imaging analogue: its benzyl-alcohol handle on the leaving arm was activated as a p-nitrophenyl carbonate and coupled to an ethylenediamine spacer, providing a free amine for NIR-dye conjugation while retaining the phosphorus(V)–aryloxy warhead. K1-N ($C_{23}H_{31}N_2O_8P$, 494.1818 Da), a diethyl phosphonate lacking the aryloxy leaving group, serves as the warhead-inactivated control of the K1 series. Leaving-group meta-substituted analogs—K2-A1 (meta-nitro; $C_{29}H_{27}N_3O_{10}PS^-$, 640.1160 Da), K2-B (meta-methanesulfonyl; $C_{30}H_{30}N_2O_{10}PS_2^-$, 673.1085 Da), K2-A3 ($C_{30}H_{28}F_2N_2O_8PS^-$, 645.1278 Da), and K2-A4 ($C_{29}H_{27}F_5N_2O_8PS_2^-$, 721.0872 Da)—extend the design space around the leaving-group recognition element. Under the mass-accounting convention that exchange increases the target mass by the conjugate monoisotopic mass minus the neutral leaving phenol, exchange of K2 with P116 is predicted to generate a +375.0421 Da adduct (613.1215 − 238.0794 Da) and exchange of K1 a +448.1400 Da adduct (560.1724 − 112.0324 Da); these mass shifts are proposed as readouts detectable by intact-protein electrospray mass spectrometry at ≤20 ppm mass tolerance, analogous to the adduct readout central to validation of the BTR platform on its oncology targets [7].

Table 3. Architecture-inverted conjugate series targeting the Tyr671 region of P116 (formulas and monoisotopic masses).

Compound	Formula	Monoisotopic mass (Da)	Role	Description
K2	$C_{29}H_{27}FN_2O_8PS^-$	613.1215	Lead (architecture inversion)	4-sulfobenzamide anchor, C2 tether, ethyl phosphinate warhead with 3-fluoro-4-(naphthalen-2-yl)phenoxy leaving group; qualifying-pose fraction 0.85
K1	$C_{27}H_{30}FN_2O_8P$	560.1724	Parallel candidate (anchor-end architecture)	6-carboxynaphthalen-2-yloxyacetyl anchor, C4 tether, 3-fluorophenoxy leaving group; qualifying-pose fraction 0.60, one NAC (3.86 Å/163.8°)
K2-N	$C_{29}H_{27}FN_2O_7PS^-$	597.1266	Warhead-inactivated control	P–C aryl bond; no exchangeable P–OAr bond; qualifying-pose fraction 0.00

Compound	Formula	Monoisotopic mass (Da)	Role	Description
K1-N	$C_{23}H_{31}N_2O_8P$	494.1818	Warhead-inactivated control (K1 series)	Diethyl phosphonate; no aryloxy leaving group
K3	$C_{26}H_{30}N_2O_{10}PS^-$	593.1364	Fluorescent release reporter	4-methylumbelliferone-7-oxy leaving group; exchange releases 4-methylumbelliferone (360/460 nm)
K4	$C_{30}H_{38}FN_4O_7P$	616.2462	Fluorescent probe (NIR conjugation handle)	Ethylenediamine handle on the leaving arm for near-infrared dye conjugation
K2-A1	$C_{29}H_{27}N_3O_{10}PS^-$	640.1160	Leaving-group meta-substituted analog	meta-NO_2
K2-B	$C_{30}H_{30}N_2O_{10}PS_2^-$	673.1085	Leaving-group meta-substituted analog	meta-SO_2CH_3
K2-A3	$C_{30}H_{28}F_2N_2O_8PS^-$	645.1278	Leaving-group meta-substituted analog	meta-difluoromethyl (CHF_2)
K2-A4	$C_{29}H_{27}F_5N_2O_8PS_2^-$	721.0872	Leaving-group meta-substituted analog	meta-pentafluorosulfanyl (SF_5)

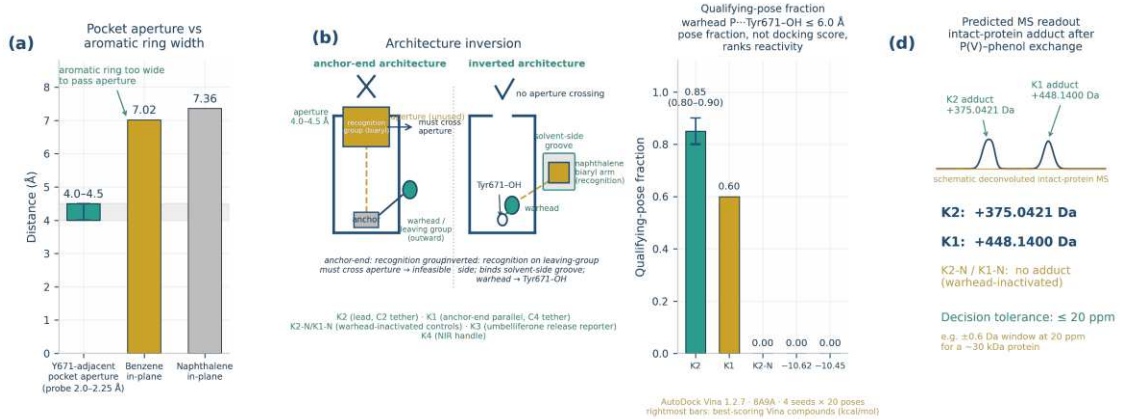


Figure 4. Pocket-aperture analysis, architecture inversion, and docking-based selection of the K-series conjugates.

(a) The 533 Å³ Tyr671-adjacent subsidiary pocket (nearest wall 2.70 Å from Tyr671-OH) with its computed entrance aperture of 4.0–4.5 Å (probe radius 2.0–2.25 Å), compared with the in-plane widths of benzene (7.02 Å) and naphthalene (7.36 Å), illustrating why direct wedge-in architectures are infeasible. (b) Architecture inversion: the recognition group (naphthalene biaryl arm) is carried on the leaving-group side of the phosphorus(V) warhead, relocating binding to the solvent-side groove; shown are K2 (lead; C2 tether), K1 (anchor-end parallel candidate; C4 tether), the warhead-inactivated controls K2-N/K1-N, and the reporters K3 (umbelliferone release) and K4 (NIR handle). (c) Qualifying-pose fractions (warhead P···Tyr671-OH ≤ 6.0 Å; AutoDock Vina 1.2.7, 8A9A homodimer, 25 Å box, exhaustiveness 16, four random seeds, 20 poses per seed): K2 0.85 (0.80–0.90), K1 0.60, K2-N 0.00; the two best-scoring compounds by Vina score (−10.62, −10.45 kcal/mol) give zero qualifying poses, motivating pose-fraction ranking. (d) Key design distances (0–16 Å scale; ~25 Å main-cleft opening as unreachable reference): pocket wall–Tyr671-OH 2.70 Å; K1 best near-attack pose P···O 3.86 Å (angle 163.8°); Phe676 Cγ–pocket center 4.90 Å; Lys670 Nζ–Tyr671-OH 5.10 Å; K2 docked P···O median 5.60 Å; design anchor–Tyr671-OH 6.00 Å; cross-subunit Tyr671-OH···OH 7.03 Å; second vestibule–Tyr671 7.30 Å; vertical markers at the 6.00 Å design basis and the 7.61 Å literature calibration reference. (e) Predicted intact-protein mass-spectrometry readout: +375.0421 Da adduct for K2 (+448.1400 Da for K1) upon phosphorus(V)–phenol exchange.

2.5 Trigger Sites Are Strain-Invariant across P1 Genotypes

Covalent-target durability requires that trigger residues resist sequence drift. We therefore extracted the coding sequences of the four complete *M. pneumoniae* genomes—M129 (P1 type 1; NC_000912.1), FH (P1 type 2; NC_017504.1), 309 (NC_016807.1), and PI1428 (NC_020076.1)—from GenBank [17] and performed window-anchored pairwise comparisons of the three trigger-bearing proteins. This strain panel spans the dominant P1-1 genotype and the P1-2 lineage that together account for essentially all recent clinical isolates, including the 2023–2024 Chinese resurgence in which 97.1% of 421 multicenter isolates were macrolide-resistant [2]. Full-length protein identity was near-absolute for P116 (MPN213: 99.51–100%) and the CARDS toxin (MPN372: 99.83–100%), whereas MPN142 was, as expected, type-variable, with only 70–75% of 20-residue windows matching perfectly between P1 types (the FH and 309 orthologues were recovered as P1-2-type truncations). Despite this global variability, all eighteen trigger sites—seven in P116 (Tyr188, Tyr206, Tyr398, Tyr566, Tyr671, Tyr821, Tyr942), six in P40/P90 (Tyr639, Tyr759, Tyr760, Tyr858, Tyr878, Tyr897), and five in the CARDS toxin (Tyr68, Tyr344, Tyr453, Tyr464, Tyr504)—are identically conserved in all four genomes, with zero mismatches in the ±30-residue neighborhoods flanking every site (Figure 5). These analyses suggest that escape by point mutation at the trigger positions would require changes within sequence blocks that have remained invariant across the known genotype diversity of the species.

The conservation data also sharpen the tiering logic for the two Tier-2 targets. For P40/P90, the crystallographically defined sialic-acid pocket in the N-terminal domain (6TLZ, 3.1 Å) [13] initially appeared to offer a small-molecule route analogous to the Tyr671 region; however, the nearest pocket-proximal tyrosine, Tyr639, lies 10.18 Å (heavy-atom distance) from the bound ligand—at the high end of the four-carbon tether span distribution (P95 10.24 Å), reachable only by extreme conformations—and is only weakly exposed (SASA 22.1 Å²), ruling out a sialic-pocket-directed warhead. In contrast, the P90 C-terminal region presents an adjacent double-trigger patch, Tyr759 (95.2 Å²) and Tyr760 (93.0 Å²), with Tyr858 (71.0 Å²), Tyr878 (46.0 Å²), and Tyr897 (45.0 Å²) as additional exposed candidates—geometry that favors an antibody-format conjugate engaging two neighboring triggers rather than a pocket-wedge small molecule. We note that MPN142 undergoes extensive surface processing (a 26-residue leader is removed, cleavage at position 455 generates the P40 (residues 27–454) and P90 (residues 455–1218) products, and at least fifteen fragments have been mapped by mass spectrometry) [18], so the stoichiometry and surface lifetime of any P90-directed conjugate must account for this proteolytic turnover. For the CARDS toxin, the five conserved triggers are retained as candidates pending dimer-context or complex-context structural validation analogous to that performed here for P116.

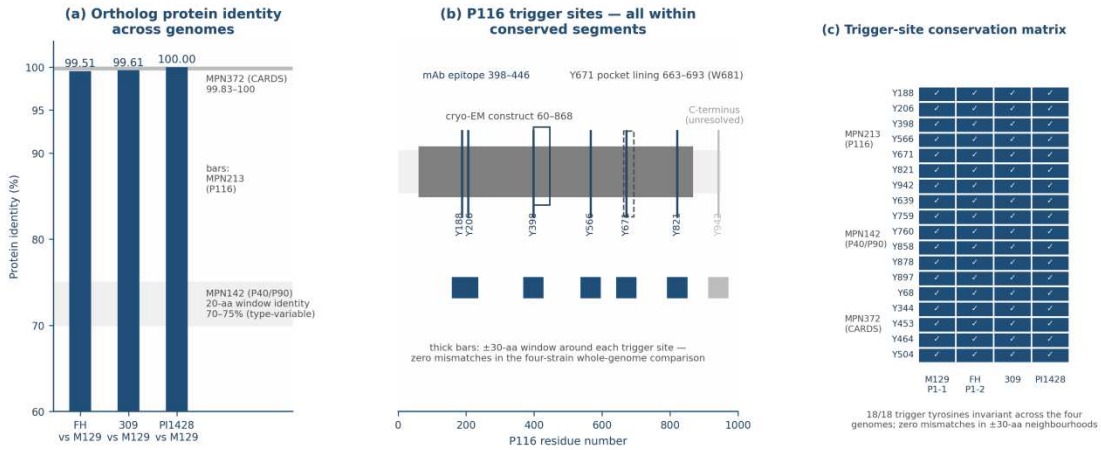


Figure 5. Strain conservation of trigger sites across four complete M. pneumoniae genomes.

Window-anchored protein identity is shown for MPN213 (P116; 99.51–100%), MPN372 (CARDS toxin; 99.83–100%), and MPN142 (P40/P90; type-variable, 70–75% of 20-aa windows identical between P1 types) across M129 (NC_000912.1), FH (NC_017504.1), 309 (NC_016807.1), and PI1428 (NC_020076.1). All eighteen trigger tyrosines (P116: Tyr188, Tyr206, Tyr398, Tyr566, Tyr671, Tyr821, Tyr942; P40/P90:

Tyr639, Tyr759, Tyr760, Tyr858, Tyr878, Tyr897; CARDS: Tyr68, Tyr344, Tyr453, Tyr464, Tyr504) are invariant in all four genomes, with zero mismatches in ±30-residue neighborhoods (highlighted tracks).

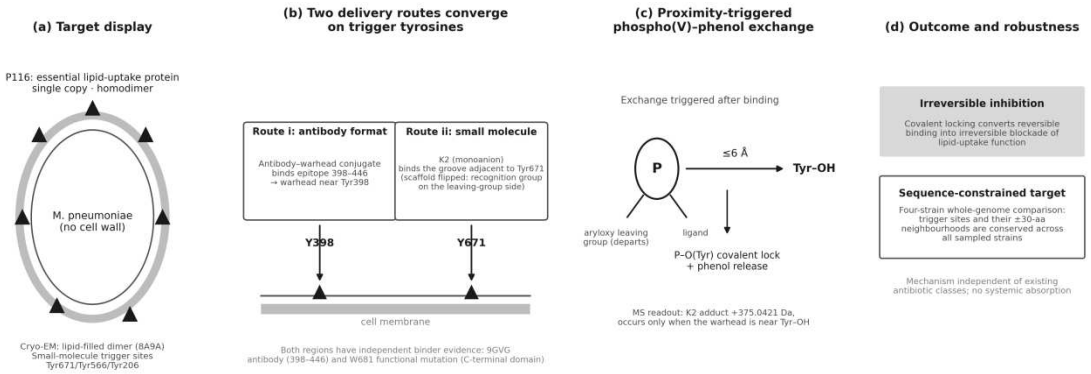


Figure 6. Mechanism schematic of proximity-triggered covalent engagement of P116.

(a) An extracellular M. pneumoniae cell presents the essential lipid-uptake protein P116 at ~34 copies per cell. (b) Two delivery routes converge on validated trigger tyrosines: an antibody-format conjugate binding the 398–446 epitope (route i, Tyr398) and the architecture-inverted small-molecule conjugate K2 binding the solvent-side groove adjacent to the 533 Å³ Tyr671 pocket (route ii, Tyr671; docked P···O median 5.60 Å). (c) Proximity-triggered phosphorus(V)–phenol exchange ejects the leaving arm (which carries the recognition group in the inverted architecture) and locks the conjugate onto the trigger (predicted +375.0421 Da adduct for K2). (d) The covalent lock converts reversible occupancy into irreversible inhibition of cholesterol uptake; conservation of all eighteen trigger sites across P1 genotypes underpins the expected durability of the approach against the current macrolide-resistant resurgence.

3. Discussion

3.1 The First Proximity-Triggered Covalent Strategy for a Bacterial Pathogen

The results presented here position P116 at the intersection of two previously disconnected domains: bacterial surface biology and proximity-triggered covalent chemistry. The binding-to-release (BTR) platform of Wen and colleagues established that a phosphorus(V) electrophile, positioned by noncovalent ligand binding, can be cleaved by a proximal tyrosine phenol to release a payload without receptor-

mediated endocytosis, and validated this mechanism on the tumor-associated targets fibroblast activation protein and programmed cell death ligand 1 [7]. Critically, the scope of that work—and of the associated patent filing (WO 2025/162459), whose targeting-group definitions are confined to tumor- and tissue-associated biomarkers—remains within oncology. The present study extends the same chemical logic, for the first time, to an essential protein on the surface of a bacterial pathogen, thereby defining a mechanism class we term ligand-guided covalent targeting of a bacterial surface protein. The closest experimental precedent is the proximity-enabled reactive therapeutic (PERx) platform, in which genetically encoded fluorosulfate- and fluorothiophosphate-tyrosine warheads converted anti-SARS-CoV-2 nanobodies into covalent inhibitors by cross-linking the non-catalytic receptor-binding-domain residue Tyr369 within 5–10 min at physiological pH [9, 8]. Our approach differs from PERx in two respects that matter for antibacterial development: the warhead is installed by total chemical synthesis on small-molecule or antibody carriers rather than by genetic code expansion, and the target is a bacterial rather than a viral surface antigen. It should also be situated relative to the classical precedent of covalent antibacterials: the β-lactams act by acylation of active-site serines of penicillin-binding proteins, an enzyme-directed covalent mechanism that has dominated antibiotic pharmacology for eight decades [5]. Ligand-guided covalent engagement of a non-enzymatic, surface-exposed lipid-acquisition protein therefore represents, to our knowledge, a new mechanism class for antibacterials—one orthogonal to both the ribosome-targeting formulary and the β-lactam paradigm.

The architecture inversion introduced in Section 2.4 may prove to be the more general lesson of this study. Surface proteins frequently present their most exposed nucleophiles next to shallow, solvent-open grooves rather than deep pockets, and even where a buried pocket exists, its entrance aperture can exclude the aromatic recognition elements that medicinal chemistry typically relies on (here, 4.0–4.5 Å against 7.02–7.36 Å aromatic in-plane widths). Moving the recognition group onto the warhead's leaving arm converts the leaving group from a passive departing fragment into the principal binding determinant, allows the exchange reaction to occur at a solvent-side groove without aperture crossing, and simultaneously provides a built-in structure–activity handle: leaving-group substitution (as in K2-A1, K2-B, K2-A3, and K2-A4) directly modulates recognition, and reporters (K3, K4) can be installed on the same arm. We anticipate that this "recognition-on-the-leaving-group" arrangement will be applicable to other proximity-triggered systems in which the trigger-adjacent cavity is solvent-accessible but aperture-limited.

3.2 Covalent Locking versus Reversible Binding

The pharmacological rationale for covalency on P116 is sharpened by the target's stoichiometry. The validated monoclonal antibody 3G9, whose Fab-bound structure (9GVG) anchors our epitope analysis, impairs cholesterol acquisition yet produces only mild, reversible growth suppression in culture, with the W681A substitution reducing low-density-lipoprotein-derived cholesterol uptake by approximately 60% [5]. Such partial, occupancy-dependent inhibition is the expected ceiling of a reversible binder against a protein present at only ~34 copies per cell: dissociation continuously regenerates functional target, and the small absolute copy number means that even transient unbound intervals restore a substantial fraction of total lipid-uptake capacity [4]. Irreversible engagement inverts this arithmetic. Because P116 is essential, single-copy, and slowly replaced, each covalently locked molecule is permanently removed from the functional pool, so that weak or moderate-affinity binding is converted into durable, cumulative inhibition rather than equilibrium-limited suppression. Experimental precedent supports the magnitude of this conversion: in the PERx system, installing a latent bioreactive warhead on an otherwise reversible nanobody yielded potency gains of up to approximately forty-fold [9]. We propose a washout-and-replate experiment as the decisive test of this mechanism: *M. pneumoniae* cultures exposed to a warhead-bearing conjugate, washed, and replated in conjugate-free medium should fail to recover growth if occupancy is irreversible, whereas recovery after washout of the parent mAb 3G9 defines the reversible baseline. The same logic generalizes beyond P116: the P90 C-terminal double-trigger patch (Tyr759/Tyr760) favors an antibody-format conjugate that locks two adjacent tyrosines, and neutralization of the CARDS toxin through its five conserved trigger tyrosines would convert antitoxin binding from stoichiometric sequestration into permanent inactivation.

3.3 Selectivity by Construction

Four features of the design argue for selectivity by construction rather than by optimization. First, P116 has no human homolog, so the protein target itself is absent from host tissue [4]; more strongly, the P116 domain families (InterPro IPR057145 and IPR054689; Pfam PF23707) have zero members in the human proteome, excluding host off-target triggers at the sequence-family level; consistent with this, the two InterPro entries comprise exclusively Mycoplasmataceae proteins (12 and 25 members, respectively) [19]. Warhead reactivity toward off-target host proteins (plasma or cell-surface tyrosines) was not assessed in this study. Second, the lead conjugate K2 carries no cytotoxic payload; its intended action is covalent blockade of an essential bacterial lipid transporter, not delivery of a cell-killing warhead, which removes the principal toxicity liability of the ADC/SMDC format. Third, the PhoPEx warhead is intrinsically latent: phosphorus(V)–phenol exchange proceeds at appreciable rates only when ligand

binding enforces proximity to a phenol, as demonstrated by the target-dependence of payload release in the BTR system [7]; consistently, the warhead-inactivated control K2-N showed a qualifying-pose fraction of 0.00 in our docking evaluation, confirming that the geometric readout tracks the exchangeable phosphorus–aryloxy bond specifically. Fourth, the anionic 4-sulfobenzamide anchor of K2 provides aqueous solubility compatible with formulation for airway-local delivery. We therefore expect random surface tyrosines on host epithelial proteins to react orders of magnitude more slowly than the guided engagement of P116 triggers. This is, however, an expectation requiring measurement, not a demonstrated property: the fluorescent reporters K3—whose 4-methylumbelliferone leaving group reports exchange by a 360/460 nm release readout—and K4, bearing a near-infrared conjugation handle, are designed for exactly this purpose and should be profiled against human respiratory epithelial cell surfaces to quantify off-target labeling before any in vivo commitment.

3.4 Fit with Airway-Local Delivery

The therapeutic setting of *M. pneumoniae* infection aligns unusually well with a proximity-triggered conjugate. The organism is an extracellular colonizer of the respiratory epithelium, and its essential surface machinery is continuously exposed to the airway lumen rather than sequestered intracellularly [5]. Inhaled or nebulized delivery of a conjugate could therefore achieve high local concentrations at the site of colonization while limiting systemic exposure, which simultaneously addresses two concerns raised by the design. The low-copy-number stoichiometry of P116 (~34 copies per cell) is a liability for systemically dosed agents requiring sustained plasma occupancy, but becomes an advantage for a locally concentrated covalent agent, because few molecules per cell must be locked. Likewise, plasma stability and systemic off-target distribution—central constraints on the BTR oncology program, where the warhead must survive circulation—are largely sidestepped when the route of administration places the conjugate directly on the target-bearing mucosa [7]. This alignment is clinically pertinent precisely where the formulary gap is widest: with macrolide resistance exceeding 90% across most surveyed provinces and the principal patient group restricted from tetracyclines and fluoroquinolones, a topically delivered, mechanism-diverse agent would address the population least served by systemic options [4, 7].

3.5 Limitations

Several limitations qualify the conclusions. (i) All findings reported here are computational and structural analyses; no conjugate has been synthesized, and no wet-laboratory experiment has yet tested trigger engagement, groove occupancy, or growth inhibition by the K series. (ii) PhoPEx exchange kinetics have been validated only on the oncology targets of Wen and colleagues [7]; whether the warhead exchanges at useful rates with P116 tyrosines—or with any target outside that system—is unknown. A model-warhead

kinetic study with a small model warhead is planned as a mechanism gate, because the entire design collapses if proximity-driven exchange does not occur on this protein. (iii) The occupancy–response relationship at ~34 copies per cell is unknown: covalent locking of a fraction of P116 molecules may be bacteriostatic, bactericidal, or phenotypically silent, and the fraction required cannot be estimated from structure alone. (iv) The Tyr671-adjacent solvent-side groove and subsidiary pocket have no ligand-bound structure; the 533 Å³ dimer-interface volume, the aperture measurements, and the architecture-inverted design of K2, now supported by docking pose statistics, constitute a shape-based hypothesis that requires biophysical confirmation by biolayer interferometry or surface plasmon resonance, followed by hit optimization. (v) Warhead off-target reactivity is unmeasured; the selectivity argument of Section 3.3 rests on mechanism-based expectation and awaits experimental quantification with the K3/K4 probes on human cells. (vi) With the exception of P116 and the P40/P90 N-terminal domain, all screen hits—including the lipoprotein MPN474, the nuclease MPN133, and the CARDS-toxin candidates—derive from AlphaFold models and await experimental structures. The conservation analysis spans only four complete genomes and, although it covers both dominant P1 genotypes, cannot exclude rare escape variants in the wider population. (vii) Grid analysis admits only a 2.0–2.25 Å-radius probe at the pocket entrance (4.0–4.5 Å aperture); the architecture inversion sidesteps aperture crossing by design, but groove-side association kinetics and any conformational gating of the entrance cannot be validated on a static structure, and the association kinetics of the first BLI measurement will be the earliest test of groove accessibility. (viii) The structurally robust trigger set is bounded by pocket availability: three sites—Tyr671 (533 Å³ subsidiary pocket), Tyr566 (~139 Å³ cavity), and Tyr206 (~272 Å³ pocket, 3.1 Å away)—pass both the phenol-OH SASA ≥ 35 Å² criterion and the requirement of a detectable pocket within 6 Å, whereas Tyr188, although solvent-exposed, has no adjacent pocket within 6 Å in any parameter combination tested and therefore remains a valid anchor for antibody-format engagement rather than small-molecule design. (ix) The dimer-interface binding site is state-dependent. Which state surface-resident P116 adopts on the bacterial membrane is unknown; binding comparisons between lipid-filled and emptied P116 (or HDX-MS) are required, and if the lipid-emptied state predominates, both the groove-based and pocket-based engagement models would need re-evaluation. (x) For the twelve targets resting on AlphaFold models, pLDDT is a backbone-confidence metric and does not guarantee side-chain geometry; their phenol-hydroxyl SASA values accordingly carry lower confidence than the P116 values, which were computed on experimental structures. All results reported here are computational. The proposed conjugates have not been synthesized, and neither target engagement nor antimicrobial activity has been tested experimentally; the designs are presented as testable hypotheses that require chemical synthesis and in vitro validation.

3.6 Conclusions

This study assembles a complete computational chain from screen to design: a proteome-scale survey of thirteen surface proteins identified P116 as the sole Tier-1 target; corroboration against the experimental cryo-EM structure 8A9A defined a robust trigger set (Tyr671, Tyr566, Tyr188, Tyr206) plus the epitope-resident Tyr398; pocket-aperture analysis excluded direct wedge-in designs and motivated an architecture inversion that yielded the synthesizable K-series conjugates, with the sulfobenzamide-anchored, leaving-group-recognition lead K2 (qualifying-pose fraction 0.85; median P···O 5.60 Å), the anchor-end parallel candidate K1, the warhead-inactivated control K2-N (0.00), and the fluorescent reporters K3 and K4; and four-genome analysis showed all eighteen trigger sites to be invariant across P1 genotypes. The durable deliverable of this work is the trigger map itself—a validated, conservation-anchored catalog of covalent engagement sites that will outlast any individual conjugate. The near-term experimental path is equally clear: the planned model-warhead kinetic study to gate PhoPEx exchange on P116, intact-protein mass spectrometry to detect the predicted +375.0421 Da K2 adduct, K3/K4-probe profiling on human epithelium, and the washout-and-replate test that will determine whether covalent locking converts the mild, reversible suppression of antibody binding into durable antibacterial inhibition.

4. Methods

4.1 Structure and sequence acquisition

Experimental coordinates for P116 were downloaded from the RCSB Protein Data Bank [15]: entry 8A9A (lipid-filled homodimer, 3.3 Å resolution, chains A and B spanning residues 60–868) and entry 9GVG (P116 in complex with mAb 3G9 Fab, 5.06 Å resolution); entry 6TLZ (3.1 Å X-ray structure of the P40/P90 adhesin N-terminal domain in complex with 3′-sialyllactose) was retrieved for cross-target corroboration [2, 4]. Structural confidence was tracked on two rails: AlphaFold models were used only where the local Cα pLDDT was ≥70; experimental structures were qualified by resolution, with entries no worse than 3.5 Å admitted for SASA geometry measurements (8A9A, 3.3 Å) and entries no worse than 5.5 Å admitted for epitope or binding-region localization (9GVG, 5.06 Å), and antigen–antibody complex structures used only for epitope localization and exposure-context analysis, not for SASA geometry qualification. AlphaFold models (database version 4) were obtained programmatically through the AlphaFold Protein Structure Database API [20, 21] by querying the endpoint api/prediction/{accession} and following the canonical modelUrl field, for 13 UniProt accessions corresponding to the *M. pneumoniae* surface proteome panel, including MPN213/P116 (P75556) and MPN309/P65 (P0CJ81) [12]. Reference sequences in FASTA format were retrieved from UniProt [12]. Complete genome records

were downloaded from NCBI GenBank [17] for strains M129 (NC_000912.1), FH (NC_017504.1), 309 (NC_016807.1), and PI1428 (NC_020076.1).

4.2 Custom Shrake–Rupley SASA implementation

Solvent-accessible surface area (SASA) was computed with a custom, zero-dependency Python implementation of the Shrake–Rupley rolling-probe algorithm [10] that parses both PDB and mmCIF formats directly, without external structural-biology libraries. van der Waals radii were taken from a standard element-wise radii table, and a solvent probe radius of 1.4 Å was used. Each atom was decorated with 192 test points distributed by a Fibonacci lattice on the expanded sphere (atomic radius plus probe radius); a point was classified as accessible if no neighboring atom excluded it, and per-atom SASA was accumulated as the accessible point fraction multiplied by the sphere area. Residue-level SASA was obtained by summation over constituent atoms, and atom-level SASA was recorded separately for the tyrosine phenol hydroxyl oxygen, because this atom is the prospective covalent attachment site. The implementation was benchmarked by cross-context comparison: SASA profiles computed on monomeric AlphaFold models were compared with those computed on the experimental P116 cryo-EM homodimer (8A9A), and concordant exposure rankings of the exposed candidates were taken as evidence of correctness. For measurements on the 9GVG complex, only the P116 chain was retained for computation, after removing the Fab heavy and light chains and the 654–700 P116 fragment present in that deposit; values reported for 9GVG therefore describe the uncomplexed P116 chain.

4.3 Trigger screen criteria

A tyrosine was classified as a candidate covalent trigger when it satisfied three simultaneous criteria: (i) residue-level SASA of at least 40 Å², indicating substantial solvent exposure; (ii) a Cα pLDDT confidence score of at least 70 in the AlphaFold model, restricting the screen to confidently predicted local structure [21]; and (iii) location outside the N-terminal 25-residue segment, which was masked to exclude signal peptides and lipobox motifs that are removed or lipidated during maturation. A glycosylation-shielding filter was applied at the sequence level: candidate triggers adjacent to potential N-linked glycosylation sequons (N-X-S/T, X ≠ P) would be excluded as potentially glycan-shielded. Because *M. pneumoniae* lacks a classical PglB-type N-linked glycosylation pathway and reported mycoplasma protein modifications are motif-less, stochastic mono-hexosylations that do not form shielding glycan chains, no candidate trigger was excluded by this filter in the present panel; the filter is retained for applications of the screening method to other pathogens. For proteins with qualifying experimental structures, SASA was applied directly to deposited coordinates, and phenol hydroxyl SASA

with an atom-level decision threshold of 35 Å² was used as the more stringent measure of warhead accessibility.

4.4 Pocket detection and aperture measurement

Pockets were mapped with a focused variant of the LIGSITE grid algorithm. Grid points were laid at 1.5 Å spacing; a grid point was retained as a free (non-protein) point when it lay between 2.6 and 12.0 Å from the nearest protein atom. Along each of the six Cartesian axes, a 10 Å scan with a 2.0 Å half-width was performed, and a direction was counted as blocked when protein density was encountered; free points with at least five blocked directions were defined as pocket points. Pocket points were clustered using 26-connectivity (face, edge, and corner adjacency), and pocket volume was computed as the clustered voxel count multiplied by the voxel volume. Geometry was verified against a second, independent grid parameterization. This procedure recovered the large interdomain cleft of P116 (estimated at approximately 11,600 Å³, of the same order as the ~1.8×10^4 Å³ lipid-filled cavity reported experimentally), an auxiliary pocket of 533 Å³ whose nearest point lies 2.70 Å from the Tyr671 phenol oxygen, a second shallower vestibule 7.30 Å from Tyr671, a ~139 Å³ cavity at Tyr566, and a ~272 Å³ pocket 3.1 Å from Tyr206. Entrance apertures were determined by probe-radius scans at the pocket mouth: the Tyr671-adjacent pocket admits a probe radius of 2.0–2.25 Å (aperture diameter 4.0–4.5 Å; narrowest-section free area 2.56–3.19 Å²), and sections with a long-to-short axis ratio below 1.5 were treated as isotropic apertures for which a traversing group must satisfy both orthogonal dimensions. Aromatic in-plane widths were computed from van der Waals envelopes of RDKit-generated conformer ensembles (ETKDGv3 embedding, MMFF force field, 200 conformers per fragment), giving 7.02 Å for benzene and 7.36 Å for naphthalene.

4.5 Epitope and distance measurements

The mAb 3G9 epitope was defined from 9GVG coordinates as the set of P116 residues with any atom within 5 Å of any Fab atom, yielding the segment spanning residues 398–446. Inter-residue distances were computed both between Cα atoms (CA–CA) and between the nearest heavy atoms, as appropriate to the question. Warhead attachment points were evaluated by the distance from the modeled attachment atom to the target tyrosine phenol oxygen, with 6.00 Å adopted as the design basis for the anchor-to-Tyr671-OH span. Tether reach was estimated from conformer ensembles of straight-chain alkylene tethers of two to six carbons (~400 conformers per tether); each tether spans a continuous distance interval rather than a single threshold (the four-carbon tether covers 4.06–10.67 Å, 95th percentile 10.24 Å), so tether length was treated as a design variable selected against the computed target geometry. A

literature calibration reference of 7.61 Å for a four-carbon tether [7] is indicated in the figures for orientation only.

4.6 Conservation workflow

Coding sequences (CDS) were extracted from the four GenBank records by a three-probe procedure: matching by locus_tag, by old_locus_tag, and by product-name homology probes, using "P116 family lipid acquisition surface protein" for *mpn_213* orthologs and "cytadherence protein" for *mpn_142* orthologs in strains FH and 309, in which the locus identifiers differ from the M129 reference. Positions of interest were mapped between strains by a window-anchored procedure: a ±12-amino-acid window centered on each trigger residue was searched in the orthologous sequence, allowing at most three mismatches, and the mapped position was accepted at the best-scoring alignment. Local conservation was then quantified by counting mismatches within a ±30-amino-acid neighborhood of each mapped site. Whole-protein identity was computed over the full alignment.

4.7 Molecular design, docking, and in silico verification

Candidate conjugates were built and verified with RDKit; all structures were parsed from SMILES and checked for valence and formula consistency. The warhead architecture was modeled on compound 2 of Wen and colleagues [7]: an ethyl phosphinate electrophile bearing a fluoroaryloxy leaving group, designed to undergo phosphorus(V)–phenol exchange with a proximal tyrosine phenol upon binding. The architecture-inverted lead K2 ($C_{29}H_{27}FN_2O_8PS^-$, monoisotopic mass 613.1215 Da) carries a 4-sulfobenzamide anchor, a two-carbon tether, and a 3-fluoro-4-(naphthalen-2-yl)phenoxy leaving group bearing the groove-recognition element; the anchor-end parallel candidate K1 ($C_{27}H_{30}FN_2O_8P$, 560.1724 Da), the warhead-inactivated controls K2-N (597.1266 Da) and K1-N (494.1818 Da), the umbelliferone-release reporter K3 (593.1364 Da), the near-infrared-handle probe K4 (616.2462 Da), and the leaving-group meta-substituted analogs K2-A1 (640.1160 Da), K2-B (673.1085 Da), K2-A3 (645.1278 Da), and K2-A4 (721.0872 Da) were constructed analogously. All masses are monoisotopic. Expected covalent adduct masses were computed from monoisotopic masses under the convention that exchange increases the target mass by the conjugate monoisotopic mass minus the neutral leaving phenol: exchange of K2 predicts a +375.0421 Da shift (613.1215 − 238.0794 Da) and exchange of K1 a +448.1400 Da shift (560.1724 − 112.0324 Da), proposed as readouts accessible to intact-protein electrospray ionization mass spectrometry at ≤20 ppm tolerance.

Molecular docking. Candidate ligands were built as three-dimensional structures in RDKit (ETKDGv3 embedding, MMFF94 minimization); K2, K3, and the K2-A series were modeled as the physiological

sulfonate anions, whereas K1 was modeled as the corresponding carboxylate, and K1-N and K4 were modeled in their neutral forms (as a cell-surface labeling probe, K4 does not require an anionic group). The receptor comprised the P116 cryo-EM homodimer (PDB 8A9A). Docking used AutoDock Vina 1.2.7 with a 25 × 25 × 25 Å box centered on the Tyr671-adjacent site, exhaustiveness 16, four independent random seeds, and 20 poses per seed (80 poses per ligand). Because Vina scores carry an uncertainty of several kcal/mol and the scoring function lacks explicit electrostatics (a limitation particularly relevant to the anionic K series), pose populations were treated as a geometric filter rather than as affinity estimates, and ligands were compared on pose-population statistics rather than single top scores: the qualifying-pose fraction, defined as the fraction of poses placing the warhead phosphorus within 6.0 Å of the Tyr671 phenol oxygen (pass threshold ≥0.5, computed over no fewer than four independent random seeds), and a near-attack-conformation (NAC) criterion combining distance (≤4.0 Å) with the angle between the Tyr hydroxyl oxygen, the phosphorus, and the leaving-group oxygen (≥160°, approximating in-line SN2@P attack). Robustness was verified on an independent seed set (42/7/1234/2024). Ligand and receptor PDBQT files, all pose coordinates, and per-pose metrics are provided in the accompanying data package (multiseed_analysis.json).

4.8 Data and code availability

All analyses reported here are computational and are fully reproducible from public structures and sequences (RCSB PDB entries 8A9A, 9GVG, and 6TLZ; AlphaFold Protein Structure Database models; UniProt FASTA records; and the four NCBI GenBank genomes listed in Section 4.1). Screen outputs (per-residue SASA values, trigger classifications, pocket geometries, conservation mappings), compound structures as SMILES with computed properties, and the complete docking package (receptor and ligand PDBQT files, all pose coordinates across random seeds, and per-pose metrics including warhead-P···Tyr671-OH distances and approach angles) are available from the corresponding author on reasonable request. No wet-laboratory experiments were performed in this study; all mass-spectrometric adducts and kinetic readouts described are predicted or proposed.

Author Contributions

C.S.: Conceptualization; investigation, formal analysis, data curation, visualization, writing—original draft; supervision; writing—review & editing. D.L.: Writing—review & editing. All authors read and approved the final manuscript.

Conflict of Interest

The authors declare the following competing interest: a patent application covering the conjugate compounds and methods described in this work has been filed. The authors declare no other competing interests.

Funding

This work was supported by the Beijing Science and Technology Nova Program Interdisciplinary Project (20230484439) and the Beijing High-Level Public Health Technical Talents Project (2023-02-33). The funders had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

References

1. Sun Y, Li P, Jin R, Liang Y, Yuan J, Lu Z, Liang J, Zhang Y, Ren H, Zhang Y, Chen J, Huang Y, Lin C, Li Y, Zhou J, Wang X, Li Y, Huang S, Xu J, Qin T. Characterizing the epidemiology of *Mycoplasma pneumoniae* infections in China in 2022-2024: a nationwide cross-sectional study of over 1.6 million cases. *Emerg Microbes Infect*. 2025;14(1):2482703. doi:10.1080/22221751.2025.2482703

2. Zhang M, Wen W, Ma Y, Hua Y, Cao Q, Lai H, Ye G, Wu L, Chen G, Ke Q, Sun X. Multicenter analysis of the *Mycoplasma pneumoniae* resurgence in China 2023-2024. *Sci Rep*. 2026;16:76. doi:10.1038/s41598-025-29135-7

3. Waites KB, Xiao L, Liu Y, Balish MF, Atkinson TP. *Mycoplasma pneumoniae* from the respiratory tract and beyond. *Clin Microbiol Rev*. 2017;30(3):747-809. doi:10.1128/CMR.00114-16

4. Sprankel L, Vizarraga D, Martín J, Manger S, Meier-Credo J, Marcos M, Julve J, Rotllan N, Scheffer MP, Escolà-Gil JC, Langer JD, Piñol J, Fita I, Frangakis AS. Essential protein P116 extracts cholesterol and other indispensable lipids for mycoplasmas. *Nat Struct Mol Biol*. 2023;30(3):321-329. doi:10.1038/s41594-023-00922-y

5. Vizarraga D, Marcos M, Rotllan N, Martín J, Santos D, Camacho M, Soto B, Velasco-Reniu L, Guerra P, Pareja F, Collantes M, Wu W, Rodríguez-Arce I, Serrano L, Piñol J, Fita I, Escolà-Gil JC. Sources of essential lipids for *Mycoplasma pneumoniae* via P116 to target liver and atherosclerotic lesions. *Nat Commun*. 2025;16. doi:10.1038/s41467-025-66129-5

6. Singh J, Petter RC, Baillie TA, Whitty A. The resurgence of covalent drugs. *Nat Rev Drug Discov*. 2011;10(4):307-317. doi:10.1038/nrd3410

7. Wen Z, Xu M, Yan Z, Kong Z, Wang Y, Liu P, Chen J, Liu Y, Cui X-Y, Cen T, Li X, Zheng Y, Gu Z, Xu D, Xu Q, Luo Y, Wang C, Liu Z. A binding-to-release strategy for targeted anticancer drug delivery. *Nature*. 2026. doi:10.1038/s41586-026-10971-0

8. Li Q, Chen Q, Klauser PC, Li M, Zheng F, Wang N, Li X, Zhang Q, Fu X, Wang Q, Xu Y, Wang L. Developing covalent protein drugs via proximity-enabled reactive therapeutics. *Cell*. 2020;182:85-97. doi:10.1016/j.cell.2020.05.028

9. Yu B, Li S, Tabata T, Wang N, Cao L, Kumar GR, Sun W, Liu J, Ott M, Wang L. Accelerating PERx reaction enables covalent nanobodies for potent neutralization of SARS-CoV-2 and variants. *Chem*. 2022;8:2766-2784. doi:10.1016/j.chempr.2022.07.012

10. Shrake A, Rupley JA. Environment and exposure to solvent of protein atoms. *J Mol Biol*. 1973;74(4):351-372. doi:10.1016/0022-2836(73)90011-9

11. Varadi M, Bertoni D, Magana P, Paramval U, Pidruchna I, Radhakrishnan M, Tsenkov M, Nair S, Mirdita M, Yeo J, Kovalevskiy O, Tunyasuvunakool K, Laydon A, Žídek A, Tomlinson H, Hariharan D, Abrahamson J, Green T, Jumper J, Birney E, Steinegger M, Hassabis D, Velankar S. AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences. *Nucleic Acids Res*. 2024;52(D1):D368-D375. doi:10.1093/nar/gkad1011

12. The UniProt Consortium. UniProt: the universal protein knowledgebase in 2023. *Nucleic Acids Res*. 2023;51(D1):D523-D531. doi:10.1093/nar/gkac1052

13. Vizarraga D, Kawamoto A, Matsumoto U, Illanes R, Pérez-Luque R, Martín J, Mazzolini R, Bierge P, Pich OQ, Espasa M, Sanfeliu I, Esperalba J, Fernández-Huerta M, Scheffer MP, Pinyol J, Frangakis AS, Lluch-Senar M, Mori S, Shibayama K, Kenri T, Kato T, Namba K, Fita I, Miyata M, Aparicio D. Immunodominant proteins P1 and P40/P90 from human pathogen *Mycoplasma pneumoniae*. *Nat Commun*. 2020;11:5188. doi:10.1038/s41467-020-18777-y

14. Kannan TR, Baseman JB. ADP-ribosylating and vacuolating cytotoxin of *Mycoplasma pneumoniae* represents unique virulence determinant among bacterial pathogens. *Proc Natl Acad Sci USA*. 2006;103(17):6724-6729. doi:10.1073/pnas.0510644103

15. Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, Burkhardt K, Feng Z, Gilliland GL, Iype L, Jain S, Fagan P, Marvin J, Padilla D, Ravichandran V, Schneider B, Thanki N, Weissig H, Westbrook JD, Zardecki C. The Protein Data Bank. *Nucleic Acids Res*. 2000;28(1):235-242. doi:10.1093/nar/28.1.235

16. Gaulton A, Hersey A, Nowotka M, Bento AP, Chambers J, Mendez D, Mutowo P, Atkinson F, Bellis LJ, Cibrián-Uhalte E, Davies M, Dedman N, Karlsson A, Magariños MP, Overington JP, Papadatos G, Smit I, Leach AR. The ChEMBL database in 2017. *Nucleic Acids Res*. 2017;45(D1):D945-D954. doi:10.1093/nar/gkw1074

17. Sayers EW, Cavanaugh M, Clark K, Pruitt KD, Schoch CL, Sherry ST, Karsch-Mizrachi I. GenBank. *Nucleic Acids Res*. 2022;50(D1):D161-D164. doi:10.1093/nar/gkab1135

18. Widjaja M, Berry IJ, Pont EJ, Padula MP, Djordjevic SP. P40 and P90 from Mpn142 are targets of multiple processing events on the surface of *Mycoplasma pneumoniae*. *Proteomes*. 2015;3(4):512-537. doi:10.3390/proteomes3040512

19. Mitchell AL, Attwood TK, Babbitt PC, Blum M, Bork P, Bridge A, Brown SD, Chang HY, El-Gebali S, Fraser MI, Gough J, Haft DR, Huang H, Letunic I, Lopez R, Luciani A, Madeira F, Marchler-Bauer A, Mi H, Natale DA, Necci M, Nuka G, Orengo C, Pandurangan AP, Paysan-Lafosse T, Pesseat S, Potter SC, Qureshi MA, Rawlings ND, Redaschi N, Richardson LJ, Rivoire C, Salazar GA, Sangrador-Vegas A, Sigrist CJA, Sillitoe I, Sutton GG, Thanki N, Thomas PD, Tosatto SCE, Yong SY, Finn RD. InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res*. 2019;47(D1):D351-D360. doi:10.1093/nar/gky1100

20. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, Yuan D, Stroe O, Wood G, Laydon A, Žídek A, Green T, Tunyasuvunakool K, Petersen S, Jumper J, Clancy E, Green R, Vora A, Lutfi M, Figurnov M, Cowie A, Hobbs N, Kohli P, Kleywegt G, Birney E, Hassabis D, Velankar S. AlphaFold Protein Structure Database. *Nucleic Acids Res*. 2022;50(D1):D439-D441. doi:10.1093/nar/gkab1061

21. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021;596:583-589. doi:10.1038/s41586-021-03819-2

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SI.zip