Biological background and rationale: My rationale for this project began with the idea of finding a protein unique to bacteria, or excluded from eukaryotes. Arachaea was not a consideration or a criterion in the search for a novel protein. The cell wall has long been a target of antibiotic targeting because of its exclusivity to prokaryotes (Dik, 2018). Pursuing that idea, I looked for a newly characterized or discovered protein that is essential to the cell wall and came across a very recent paper submitted in late 2022 that discussed the hypothesized function of two previously uncharacterized proteins that are highly involved in the maintenance of the cell wall in prokaryotes (Sit, 2023). Of the two proteins in the paper, I researched the DUF368-domain containing protein because of its relative universality in bacteria and distinct exclusion from eukaryotes. Targeting proteins unique to bacteria minimizes the risk of harming human cells, reducing potential side effects of the antibiotic.

The DUF368-containing protein is believed to be essential for the transmembrane transport of undecaprenyl phosphate lipids and is necessary for peptidoglycan maintenance. It is hypothesized that the DUF368 protein might be involved in C55-P recycling, specifically in C55-P re-internalization. C55-P, also known as undecaprenyl phosphate, is a lipid carrier molecule that transports peptidoglycan precursors across the cytoplasmic membrane. Disruption of C55-P recycling can inhibit the synthesis of the bacterial cell wall, leading to cell death. The exact mechanism of DUF368-containing proteins and their interaction with C55-P recycling is unknown (Sit, 2023).

DUF368-containing proteins are multi-pass inner membrane proteins with a high degree of conservation across species. The DUF368 domain is found almost universally in bacteria (a notable exception is the Escherichia coli strain the authors used), both Gram-negative and Gram-positive. It is also found ubiquitously in archaeal species. There are no experimentally determined structures of the protein, but it is predicted to have a large putative cleft, which is a fundamental characteristic of domains with ligand binding and/or transport activity (Sit, 2023)

In the paper cited above, the researchers exposed E. coli strain *lptD4213* to amphomycin, which is a dual inhibitor of peptidoglycan and wall teichoic acid biosynthesis (Singh, 2016). In modified E. coli cells that heterologously expressed the DUF368-containing protein in the outer membrane, the cells developed resistance to the antibiotic. The control bacteria was weak, malformed, and died when exposed to the antibiotic. As a result of this, it is hypothesized that DUF368-containing proteins confer resistance to antibiotics that target C55-P, such as laspartomycin C, friulimicin B, bacitracin, and amphomycin (Wood, 2022; Stone, 1971).

- Dik DA, Fisher JF, Mobashery S. Cell-Wall Recycling of the Gram-Negative Bacteria and the Nexus to Antibiotic Resistance. Chem Rev. 2018 Jun 27;118(12):5952-5984. doi: 10.1021/acs.chemrev.8b00277. Epub 2018 May 30. PMID: 29847102; PMCID: PMC6855303.
- Sit B, Srisuknimit V, Bueno E, Zingl FG, Hullahalli K, Cava F, Waldor MK. Undecaprenyl phosphate translocases confer conditional microbial fitness. Nature. 2023 Jan;613(7945):721-728. doi: 10.1038/s41586-022-05569-1. Epub 2022 Nov 30. PMID: 36450355; PMCID: PMC9876793.
- Singh M, Chang J, Coffman L, Kim SJ. Solid-state NMR characterization of amphomycin effects on peptidoglycan and wall teichoic acid biosyntheses in Staphylococcus aureus. Sci Rep. 2016 Aug 19;6:31757. doi: 10.1038/srep31757. PMID: 27538449; PMCID: PMC4990924.

- Wood TM, Zeronian MR, Buijs N, Bertheussen K, Abedian HK, Johnson AV, Pearce NM, Lutz M, Kemmink J, Seirsma T, Hamoen LW, Janssen BJC, Martin NI. Mechanistic insights into the C_{55}-P targeting lipopeptide antibiotics revealed by structure-activity studies and high-resolution crystal structures. Chem Sci. 2022 Feb 21;13(10):2985-2991. doi: 10.1039/d1sc07190d. PMID: 35382464; PMCID: PMC8905900.
- Stone KJ, Strominger JL. Mechanism of action of bacitracin: complexation with metal ion and C 55 - isoprenyl pyrophosphate. Proc Natl Acad Sci U S A. 1971 Dec;68(12):3223-7. doi: 10.1073/pnas.68.12.3223. PMID: 4332017; PMCID: PMC389626.

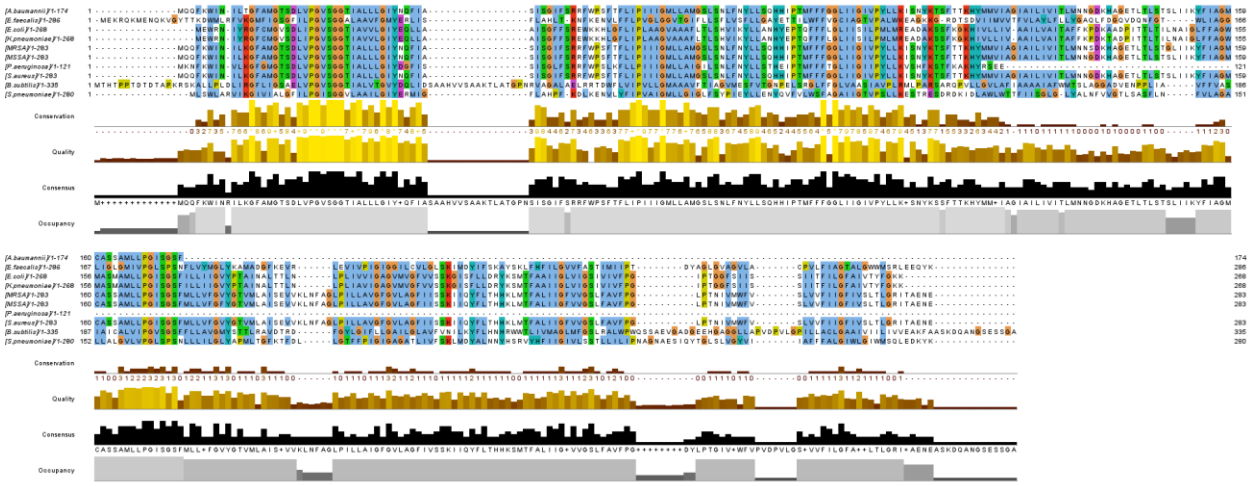
Evolutionary analysis: I began my analysis by using the Vibrio cholerae DUF368 protein's FASTA file as the BLASTp query for the required organisms. I downloaded the FASTA files for each bacteria with the highest sequence similarity and appeared most relevant and correctly categorized. There were three unique results: 1. There were no BLASTp or Interpro domain results for humans and a DUF368 domain. There is no human homolog based on sequence. 2. Acinetobacter baumannii is only a partial protein with 174 aa. 3. Pseudomonas aeruginosa is also only a partial protein with 121 aa. After searching for preexisting AlphaFold structures that contained the DUF368 domain via InterPro, I used new FASTA files for:

- Staphylococcus aureus (strain NCTC 8325 / PS 47) UniProt: Q2FZY8
- Streptococcus pneumoniae (strain ATCC BAA-255 / R6) UniProt: Q8DQ19
- Enterococcus faecalis UniProt: A0A3N3SF47
- Staphylococcus aureus (strain MRSA252) UniProt: A0A7U7EUN1

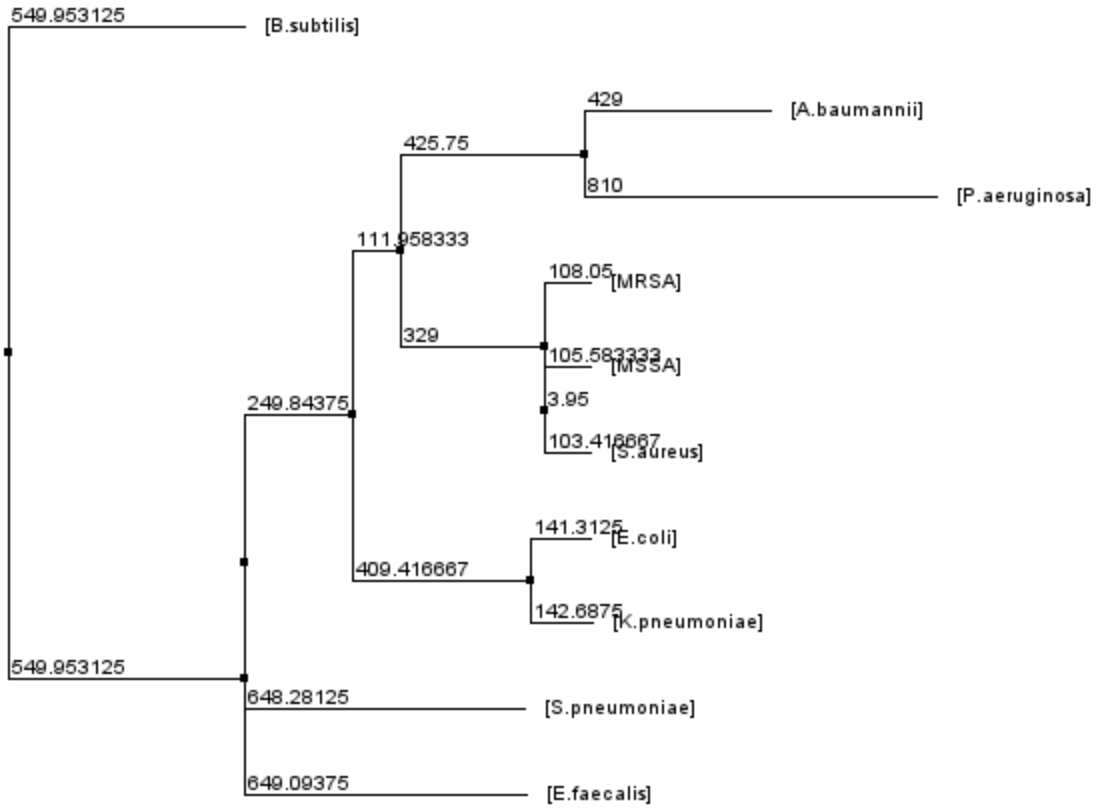
The above list is also the unique species that I will use for structural analysis and drugability of their DUF368 proteins.

When I calculate the overall MSA file's identity and similarity using a custom Python script, I get the following:
Average percent identity: 44.24534360950546
Average percent similarity: 39.63029318702376

Here is the MSA shown using Jalview with Clustal with defaults



Here is the evolutionary tree using neighbor-joining



We can see that B. subtilis is the biggest outlier. If we remove it, we find that the average percent similarity of the MSA increases to 45.93650429737414. The average percent identity marginally increases only to 44.4806369887803. Notwithstanding, we can now hypothesize that overall there are likely to be conserved regions/domains/sites for a novel molecule to bind. We explore this in the next section with **S. aureus, S. pneumoniae, MRSA, and E. faecalis,**

Comparative structural analysis: After consulting InterPro, DALI, and the PDB, there are no experimentally derived structures for DUF368 proteins. Using InterPro, I searched for AlphaFold predicted structures based on either the (DUF368) PF04018 domain or the Undecaprenyl phosphate translocase-like IPR007163 domain, which are functionally equivalent. I debated whether I should use MRSA and S. aureus in the same comparison, but when I looked at their sequences I found that the sequence identity was 281/283(99%). S. aureus has a 136 Gly whereas MRSA has a 136 Ser. S. aureus also has a 149 Ser whereas MRSA has a 149 Gly. I thought the substitutions of a polar serine for a nonpolar glycine in two locations might be worth exploring. There were only four preexisting AlphaFold structures of the ten possible bacteria species, and I knew from prior assignments that AlphaFold was likely to yield the most likely structures. Therefore, the mandated selection criteria funneled me into using these four species and corresponding theoretical structures.

Using the Dali server's AF-DB search for Humans and the AlphaFold structure of Vibrio cholerae's DUF368 protein, I found the closest structural human homolog - ZINC

TRANSPORTER ZIP4, UniProt: Q6P5W5. I used the AF-DB search because it was the fastest way to search for a human structural homolog. There was no gene with a significant sequence similarity in the human genome for DUF368, and DALI doesn't offer an easy way to parse PDB structures by species with their conventional search of RCSB's PDB.

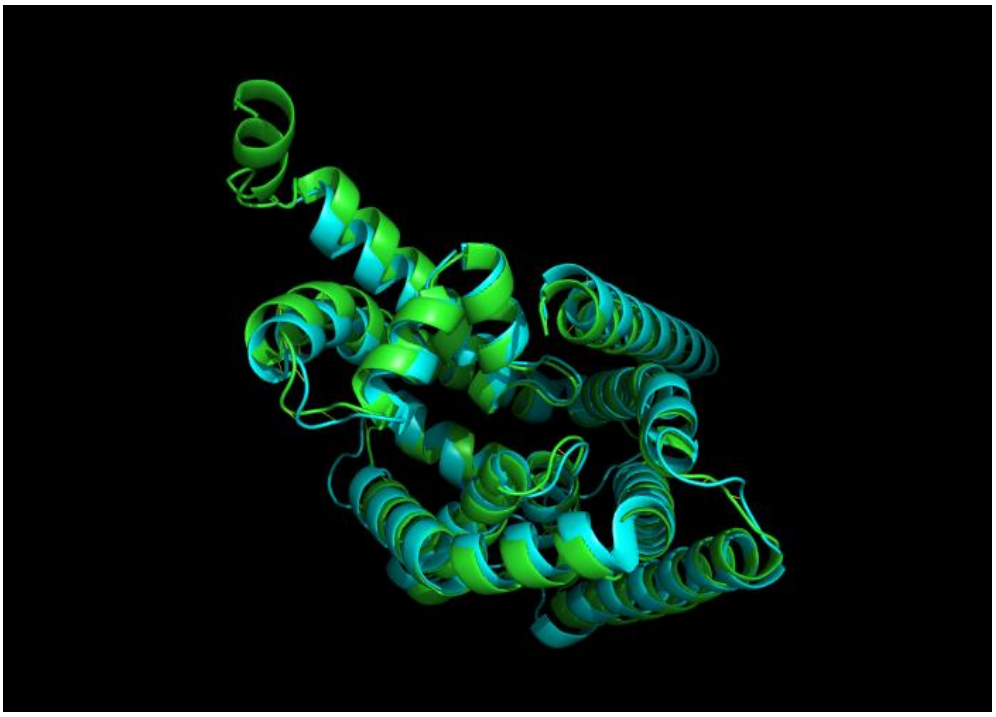
Using PyMOL I was able to calculate the RMSD values of the five proteins to give a rough sense of how similar every protein is by (predicted) structure:

	E. faecalis	H. Sapiens	MRSA	S. pneumoniae	S. aureus	
S. aureus	5.9	35.4	0.38	6.6		S. aureus
S. pneumoniae	1.2	29.8	6.9		6.6	S. pneumoniae
MRSA	6.27	35.3		6.9	0.38	MRSA
H. Sapiens	12.7		35.3	29.8	35.4	H. Sapiens
E. faecalis		12.7	6.27	1.2	5.9	E. faecalis
	E. faecalis	H. Sapiens	MRSA	S. pneumoniae	S. aureus	

Table 1. RMSD Values of all proteins compared via PyMOL

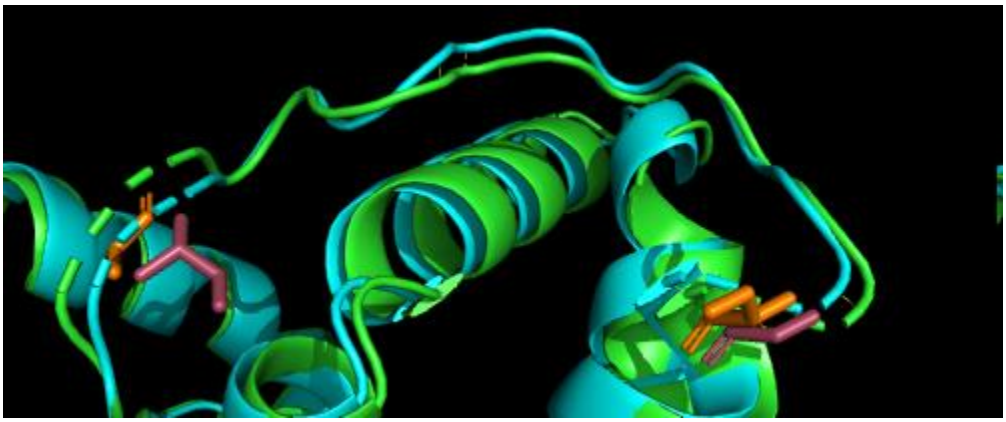
The notable comparisons are between S. pneumonia and E. faecalis (RMSD 1.2) and, of course, between MRSA and S. aureus (RMSD 0.38). The average RMSD of every structure (not including H. sapiens) is 4.54.

S. pneumonia and E. faecalis (shown aligned below blue and green, respectively) only have a sequence similarity of 177/279(63%) and a sequence identity of 123/279(44%) (aligned via BLASTp). Yet, we can see the structures (at least those predicted via AlphaFold) show remarkable overall structural conservation.

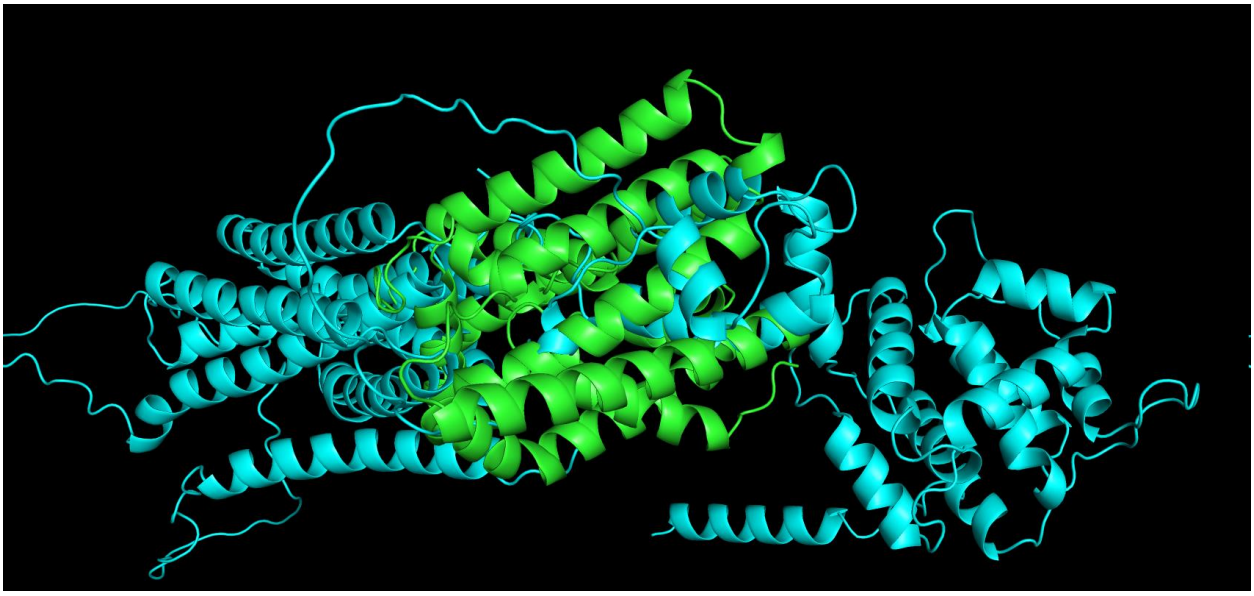


MRSA (green with red residues) and S. aureus's (blue with orange residues) differences are relatively unexciting. The two swapped amino acids appear in the largest unstructured loop of

the protein and are both near the beginning and the end of this loop.



The last structural comparison is between MRSA (green) and H. sapiens (blue). I am choosing to treat MRSA and Staphylococcus aureus's DUF368 protein as structurally the same but erring on the side of caution. I will assume that the possible drug and drug target need to be focused on matching an antibiotic-resistant bacteria and that we, therefore, need to be aware that any ligand that binds to MRSA's protein will not be toxic to humans or bind to a protein in humans.



We can see that MRSA and H. sapiens have very little conservation of structure, especially compared to our other proteins.

Ligands, Binding sites, druggability assessment: I used CastP's pocket prediction software on MRSA's DUF368 predicted structure, downloaded the predictions via the CastP plugin on PyMOL, and found the center of the largest pocket (613.605 angstroms cubed) for ligand analysis via PyMOL. The x,y,z coordinates of the MRSA protein's pocket was 7.038, -0.854, -6.143. Mcule.com identified a ligand for that pocket, "MCULE-7381215206-0-1". This was

expected. Even though Mcule searches for any and all binding sites, the top binding site for the highest docking score ligand was the pocket we identified earlier. It has a Vina Docking score of -7.4, is nontoxic, and has metrics that satisfy Lipinksi's rule of five: Mass: 436.4859; logP: 4.6495; H-bond acceptors: 7; H-bond donors: 2, Refractivity: 117.0439.

I next wanted to find if this ligand would selectively bind to MRSA's DUF368 protein exclusively. I used CastP again to find the general pocket of H. sapien's theoretical zinc protein and then used Pyrx to dock the Mcule.com ligand in the general pocketed area of H. sapien's protein. The preliminary data suggest that this is possibly a promiscuous molecule as it will bind to a pocket of the zinc protein with a Vina Docking score of -7.8.

Using PLIP's web tool for bond analysis, MCULE.com's best alignment only gives the ligand 6 hydrophobic interactions within the pocket of MRSA's protein. There are no hydrogen bonds or other interactions. From my other research, this leads me to believe this is a weak overall interaction between the ligand and protein. Conversely, PLIP gives the ligand 4 hydrophobic interactions, 5 hydrogen bonds, and 1 pi-cation interaction with the human zinc protein. The docking score of -7.8 and these stronger and more numerous bonds with the human zinc protein do not inspire confidence in using this MCULE molecule as a potent antibiotic. However, when I used Pyrx to more thoroughly explore the docking of the ligand to MRSA's protein, I found that the docking score improved to -7.7 with 3 hydrogen bonds and 6 hydrophobic interactions. I've found the optimal docking position, score, and predicted bonds for all 4 DUF368 proteins, shown in table 2 below.

MRSA	***S. aur***	***S. pneu***	***E. fae***
Hydrogen bonds			
Asp 19	Asp 19		
		22 Gly	
			Gly 37
Asn 77			
		156 Gly	
		162 Leu	
Ser 173	Ser 173		
			Gly 176
			Ser 178
Pi stacking bonds			
	Phe 214	Phe 18	
Docking Score			
-7.7	-8.5	-9.7	-10

MRSA	***S. aur***	***S. pneu***	***E. fae***
Hydrophobic interactions			
Asp 19			
Leu 20			
	Ile 43		
	Ile 43		
	Ile 46		
		Thr 96	
		100 Leu	
			Ala 114
			Leu 115
		157 Val	
Val 185			
Ile 189			
Val 206			
Val 210	Val 210		
	Val 210		
	Val 210		
	Leu 211		
			Glu 282

Table 2. All scores, bonds, and interactions between theoretical ligand (MCULE-7381215206-0-1) and DUF368 proteins.

The figure to the left represents each docked configuration of the ligand in its respective DUF368 protein (removed for clarity). S. aureus in blue, MRSA in orange, S. pneumoniae in yellow, and E. faecalis in

pink. Despite the extreme similarity in structure, Pyrx found different optimal conformations for the ligand in the two S. aureus species. I don't believe this is a mistake but rather two different conformations the ligand can take within the same general region. The blue conformation is the better of the two (compared to orange), but we don't yet know which one, if either of the positions, will have any biologically significant effect on the protein. The yellow and pink ligands are bound in a slightly different region in the DUF368 protein, and it's unclear at this point if that binding site is unique to E. faecalis and S. pneumoniae or if, once again, this is just another possible binding site identified by Pyrx. Both locations are within the pocket identified by CastP. We have compelling reasons to believe that all the DUF368 have the same possible binding pocket/region, so I'm left to speculate that there are two possible binding sites for the drug (MCULE-7381215206-0-1) within the same pocket.

Recommendations: The theoretical drug (MCULE-7381215206-0-1) has been shown to have reasonably strong interactions with the DUF368 protein found in four Gram-positive bacteria's theoretical DUF368 structures. S. aureus, MRSA, S. pneumoniae, and E. faecalis are all Gram-positive. The drug is not toxic, but it might promiscuously bind to other similar human proteins. However, all the proteins are only theoretically structured by AlphaFold, and while AlphaFold is a robust protein prediction software, we cannot wholly rely upon it. I recommend that further dockings be performed on AlphaFold predicted Gram-negative structures of DUF368 proteins to identify any bias of bacteria based on their Gram status. It is unknown if the drug will have similarly high binding scores with Gram-negative DUF368 structures. After this, I would recommend testing the drug directly (in vitro) on many bacteria species and determining its efficacy directly. I would also recommend testing close homologs of the drug, though one homolog that I could test, MCULE-8617850052-0, had a worse docking score of -5.9 when tested on MRSA's DUF368 protein. A biochemist or similar might be able to identify a way to improve the drug and obtain even higher docking scores and more atom-to-atom interactions.

Based on my analysis, the DUF368 protein is an attractive target for antibacterial drug development due to its high conservation across bacterial species and its role in bacterial cell wall maintenance. There are several risks associated with targeting the DUF368 protein. First, we lack a comprehensive understanding of its exact function and mechanism of action in bacterial cell wall maintenance. While the disruption of C55-P recycling could inhibit cell wall synthesis and lead to bacterial cell death, the precise mechanisms by which DUF368 proteins interact with C55-P recycling remain unknown. Second, while no human homolog based on sequence similarity was identified, we found that a human protein, the zinc transporter ZIP4, shares some structural similarity with the DUF368 protein, which raises potential concerns about the selectivity and side effects of any drug developed against the DUF368 protein. The ligand that was identified to potentially bind the DUF368 protein has been shown to possibly interact with the human zinc protein as well. But we do not know the effect of the ligand binding on the zinc protein or if the site identified is even exposed for a ligand to bind to it.

Given these factors, my recommendation to Pharmaceutical company X would be to proceed with caution. The DUF368 protein presents a compelling target for antibacterial drug development, but further research is necessary to fully understand its function and its potential interactions with human proteins. It would be worth investigating other potential ligands that could selectively bind the DUF368 protein while avoiding off-target effects with human proteins. In addition, a more thorough exploration of the druggability of the DUF368 protein, potentially including in vitro and in vivo studies, would be necessary before proceeding with a drug discovery campaign.

In conclusion, the feasibility of developing a broad-spectrum antibacterial drug that targets the DUF368 protein is uncertain at this stage. While it offers a novel target for antibacterial drugs, its function and potential off-target effects require further study. Developing a drug against this target is a promising but high-risk venture. Should Pharmaceutical company X decide to proceed, I would recommend a strong focus on achieving specificity and selectivity, minimizing potential liabilities, and conducting rigorous preclinical and clinical studies to assess safety and efficacy.