Deep mutational scanning of essential bacterial proteins can guide antibiotic development

Liselot Dewachter[1,2], Aaron N. Brooks[3], Katherine Noon[3], Nandini Krishnamurthy[3], Wim Versées[4,5,7], Wim Vranken[4,5,6,7], Jan Michiels[1,2,7]

[1] Centre of Microbial and Plant Genetics, KU Leuven, Leuven, Belgium

[2] VIB-KU Leuven Center for Microbiology, Leuven, Belgium

[3] Inscripta, Inc., Boulder, CO 80301, USA

[4] Structural Biology Brussels, Vrije Universiteit Brussel, VUB, Brussels, Belgium

[5] VIB-VUB Center for Structural Biology, Brussels, Belgium

[6] Interuniversity Institute of Bioinformatics in Brussels, ULB-VUB, Brussels, Belgium

[7] Co-senior authors

Correspondence to: jan.michiels@kuleuven.be

Keywords

FabZ, LpxC, MurA, CRISPR, saturation editing, antibiotic resistance, *E. coli*

Abstract

Deep mutational scanning is a powerful approach to investigate a wide variety of research questions including protein function and stability. We performed deep mutational scanning on three essential *E. coli* proteins (FabZ, LpxC and MurA) involved in cell envelope synthesis using high-throughput CRISPR genome editing. This allowed us to study the effect of the introduced mutations in their original genomic context. Using the more than 17,000 variants of FabZ, LpxC and MurA from the saturation editing libraries constructed here, we have interrogated protein function and the importance of individual amino acids in supporting viability. Additionally, we have exploited these libraries to study resistance development against antimicrobial compounds that target the selected proteins. Our results show that, among the three proteins studied, MurA is the superior antimicrobial target due to its low mutational flexibility which decreases the chance of acquiring resistance-conferring mutations that simultaneously preserve MurA function. Additionally, we were able to rank anti-LpxC lead compounds for further development guided by the number of resistance-conferring mutations against each compound. Our results show that deep mutational scanning studies can be used to guide drug development, which we hope will contribute towards the development of urgently needed novel antimicrobial therapies.

Introduction

Deep mutational scanning is a powerful way to study protein function[1-3], stability[4], amino acid roles[5], evolvability[6], epistasis[7] and more. The power of deep mutational scanning approaches relies on the construction of large mutant libraries that contain a wide variety of gene variants, followed by the selection, evaluation and identification of these variants[8,9]. Currently, such large mutant libraries are mostly constructed through error-prone PCR[3,6], or with degenerate oligonucleotides that can be used as primers[10-12] or tiles for ORF construction[13]. While these approaches have proven successful at generating valuable biological insights, they are limited in that library construction occurs *in vitro* and generates mutant alleles that are mostly studied outside of their natural context. Because expression level, copy number and genomic context can influence phenotypes, it would be advantageous to introduce mutations directly into the genome of interest rather than *in vitro*. High-throughput CRISPR-based editing can be used for this purpose[1]. Indeed, one of the main advantages of CRISPR-based genome editing is its scalability; targeting different DNA sequences can be done in parallel by providing different cells with different sgRNAs[14-16]. Recently, efforts in increasing the throughput of CRISPR-based genome editing have led to the development of dedicated workflows that allow for the simultaneous construction of thousands of targeted genomic edits in large pooled mutant libraries that can be used for deep mutational scanning experiments[17,18].

Using automated high-throughput CRISPR-based editing of the *Escherichia coli* genome, we here perform deep mutational scanning to create full-length saturation editing libraries of three different *E. coli* proteins: FabZ, LpxC and MurA. These proteins are all essential for *E. coli* viability and are involved in the synthesis of different layers of the cell envelope. FabZ is a dehydratase involved in the synthesis of fatty acids that are used for the construction of phospholipids[19]. LpxC is needed for the production of lipid A, the lipid portion of lipopolysaccharide (LPS) which is an essential component of the outer membrane of Gram-negative bacteria[20]. Finally, the MurA enzyme catalyzes the first step in the production of peptidoglycan precursors that are used to build the rigid cell wall that helps maintain cell shape and integrity[21]. Importantly, all these proteins are considered attractive targets for the development of novel antibiotics[22-25]. Since we are currently on the verge of a world-wide health crisis due to the relentless increase in antibiotic resistance, the development of new antimicrobials and exploration of novel antibiotic targets is urgently needed[26-31]. We hope to contribute towards this goal by providing detailed functional insights into the potential drug targets FabZ, LpxC and MurA. Moreover, we use our full-length saturation editing libraries to estimate the likeliness of resistance development against lead compounds, thereby prioritizing both targets and compounds for further drug development.

Results

Saturation editing of *fabZ*, *lpxC* and *murA* using high-throughput CRISPR-Cas genome editing

Three essential genes that are involved in the synthesis of the Gram-negative cell envelope; *fabZ*, *lpxC* and *murA*, were chosen for full-length saturation editing. Saturation editing libraries of these genes were created using the Onyx™ Digital Genome Engineering platform, that automates all the steps of genome-scale strain engineering and has a performance optimized version of CREATE technology at its core[18]. This automated platform allows for high-throughput CRISPR-based editing of the *E. coli* genome using the MAD7 CRISPR nuclease (https://www.inscripta.com/technology/madzymes-nucleases), provided on an inducible plasmid. Both the sgRNA and the repair template are provided on a second plasmid carrying a constitutive promoter for sgRNA expression in addition to a barcode to track the plasmids [18,32]. Repair templates contain the desired genomic edit and display homology to the targeted genomic site so that, upon cutting by the MAD7 enzyme, this oligo – together with the desired mutation – is incorporated into the *E. coli* genome (Figure 1A). Apart from the desired edit, the repair template also contains one or more synonymous mutations that prevent re-cutting by eliminating the PAM site[18]. Repair templates were designed so that, at the protein level, every amino acid would be replaced by every other amino acid. Additionally, every codon was also mutated to a synonymous codon. This way, every amino acid should be targeted 20 times (19 amino acid substitutions and one synonymous change), except for methionine and tryptophan residues, for which no synonymous codons exist. No edits were designed to target the start codons of the different genes. In total, 17,415 edits (20*(150 FabZ residues + 304 LpxC residues + 418 MurA residues) – 25 M/W residues, Figure 1B) were designed.

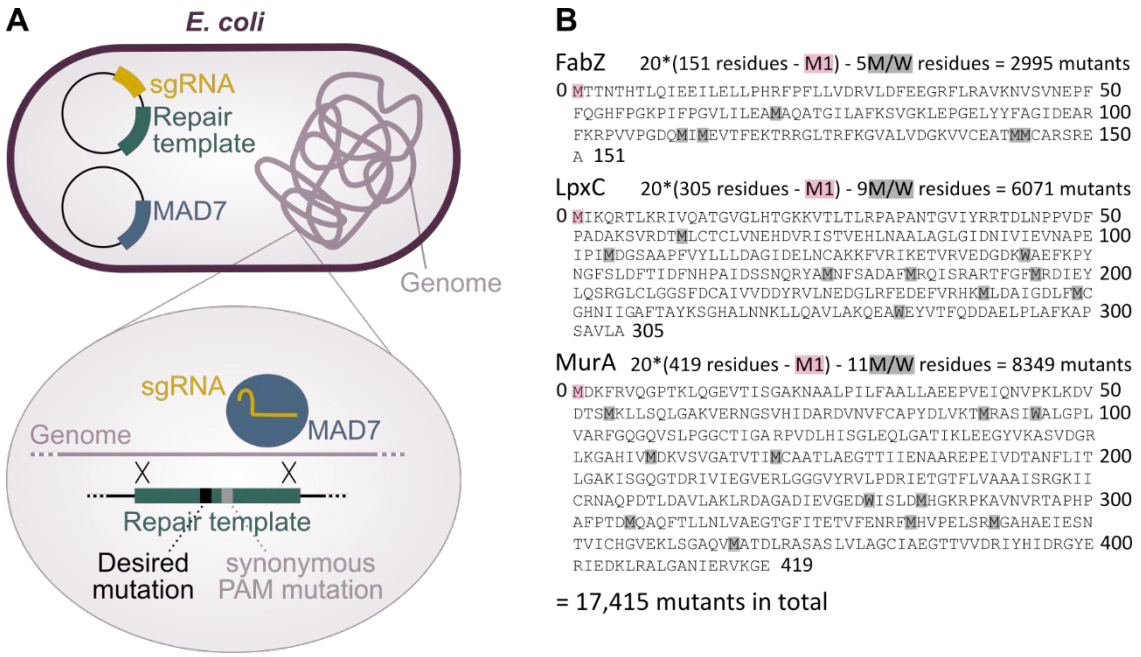

Figure 1: Construction of saturation editing libraries of *E. coli* FabZ, LpxC and MurA using high-throughput CRISPR genome editing. A) For CRISPR-based editing, two plasmids were introduced into individual *E. coli* cells. The first plasmid, the 'engine plasmid' encodes the MAD7 enzyme used for genomic cutting, the second plasmid, the 'barcode plasmid' encodes the sgRNA and repair template. The repair template is incorporated into the *E. coli* genome by homologous recombination and contains the desired edit as well as one or more synonymous mutations that prevent re-cutting by eliminating the PAM site. B) In the FabZ, LpxC and MurA saturation editing libraries, every amino acid was replaced by all 19 other amino acids, except for the start codon which was not mutated. Additionally, as a control, every codon was mutated to a synonymous codon, except for methionine (M) and tryptophan (W) residues for which no synonymous codons exist. This results in a total of 20 (or 19 for M and W residues) mutations per amino acids, leading to 17,415 variants across all three libraries.

After library synthesis and outgrowth of the engineered bacteria, Illumina sequencing was performed to check which of the designed edits could be detected in the *E. coli* genome. Since all three proteins (FabZ, LpxC and MurA) targeted by saturation editing are essential for *E. coli* viability, protein activity can be directly evaluated by checking the presence – and therefore viability – of variants in the constructed libraries. The Onyx technology optimizes for maximum representation of variants during library construction to prevent skews in the population due growth competition. We also restricted growth to what was needed for library construction (see Materials & Methods) to limit competition between constructed variants and retain all variants that support viability. Even though the generated libraries are barcoded[18], we directly sequenced the targeted genes in the chromosome to identify which mutations were present in the pooled mutant library. As a result, cells that received a barcoded plasmid but in which editing did not proceed correctly were not taken into account. Read counts associated with the designed edits are listed in Table S1.

Editing libraries for FabZ, LpxC and MurA are almost fully saturated

We first verified the quality of the generated libraries by estimating the saturation level. Because all three targeted genes are essential, it is to be expected that some edits would not be detected even if they were successfully introduced due to drop-out of non-viable variants. Therefore, instead of looking at all designed edits to determine saturation levels, we focused on the synonymous mutations that – in principle – should have minor effects on cell viability. However, we do note that synonymous mutations are not necessarily neutral and several studies demonstrate that they can have considerable fitness effects[33-36]. It therefore remains plausible that some of the missing synonymous mutations are absent from the libraries due to detrimental effects on fitness. In this case, the saturation levels estimated here would be an underestimation of the true saturation levels. Of all designed synonymous edits, 96.6% were detected for FabZ, 97.3% for LpxC and 96.3% for MurA. Given this estimated saturation level of around 96%, the likelihood of any specific residue not being mutated at all by random chance would be in the order of 10^{-28} (=$(4/100)^{20}$). . Therefore, the absence of a large number of edits at a specific position could point to either biological or technical difficulties in mutating this residue.

To rule out the possibility that technical difficulties, such as inefficient PAM sites, prevent some residues from being mutated, we looked for residues that were not mutated at all, i.e. residues for which none of the 20 designs (including the synonymous design) were detected. We identified one such uneditable residue, MurA R120, which is known to be involved in substrate binding[37-40]. Although we cannot exclude that this residue cannot be mutated by our CRISPR-based editing protocol due to technical reasons, we hypothesize that, because of this residue's important role in substrate binding, many edits at this position did not support viability and that any remaining mutations (such as the synonymous edit) were not introduced or detected due to random chance. Taken together, these data show that the absence of many mutations at a specific position can be used to pinpoint residues that are important for protein function.

Saturation editing libraries identify residues that are important for protein function

Sequencing results revealed the presence or absence of each amino acid substitution in our saturation editing libraries. These data provide a strong indication for whether a specific amino acid substitution can support protein function and viability (Figure 2A-C). Although not all amino acid changes are allowed, the number of tolerated amino acid substitutions is surprisingly high for most positions (Figures 2D-F). In fact, 60% of all residues could be mutated to all or all but one amino acid(s) (Figure 2G), thereby highlighting the robustness of protein function in light of single amino acid changes.

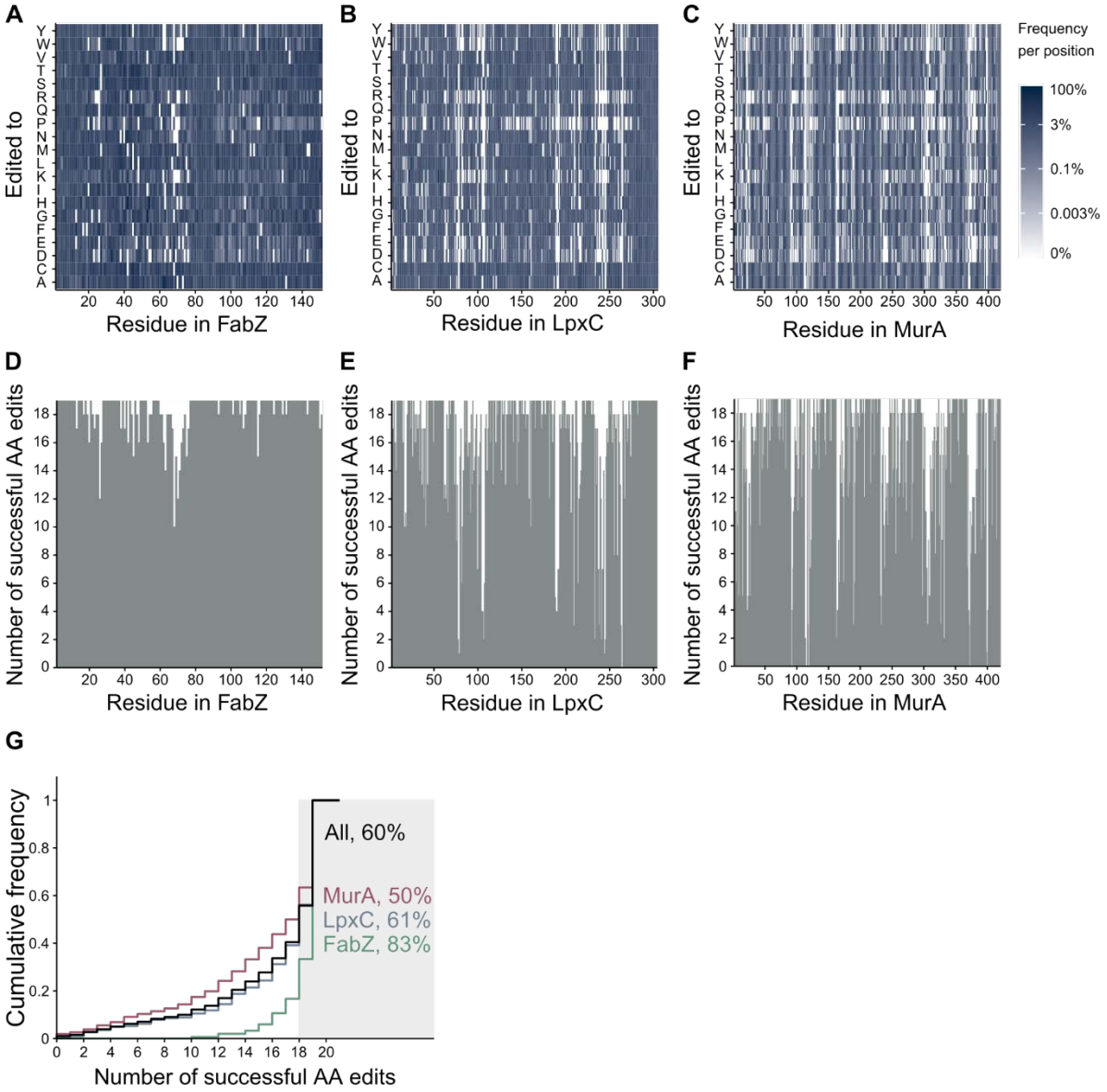

Figure 2: Analysis of the presence and absence of specific mutations in saturation editing libraries can be used to identify important residues. A-C) Heat maps indicating the presence or absence of specific amino acid substitutions in the saturation editing library of FabZ (A), LpxC (B) and MurA (C). The frequency of occurrence was normalized to the sum of the frequency of occurrence of all mutations at the same position and is indicated by a blue color scale. D-F) These plots show the number of successful amino acid changes at each position of the proteins FabZ (D), LpxC (E) and MurA (F). G) The cumulative frequency distribution of successful amino acid substitutions is shown for all libraries taken together (All) or the MurA, LpxC and FabZ library separately. The percentage of residues that tolerates all or all but one amino acid changes (≥ 18) is specifically stated and highlighted in gray. AA, amino acid.

In order to identify positions that are important for protein function, we assigned a tolerance score to each residue based on the number and types of amino acid substitutions that are tolerated. Although complex interpretations exist for assessing amino acid similarities[5], we here use a simple normalized amino acid similarity score[41], where fully tolerant promiscuous sites obtain a score of 1, fully intolerant

ones a score of 0. Tolerated mutations to very biochemically and/or structurally different amino acids (e.g. Gly to Trp) receive higher scores than ones between similar amino acids (e.g. Leu to Ile). A site with few tolerant mutations between very different amino acids might therefore receive a higher tolerance score than a site with more mutations between similar amino acids. Tolerance scores are listed in Table S2 and the distribution of these scores is shown in Figure 3A-C.

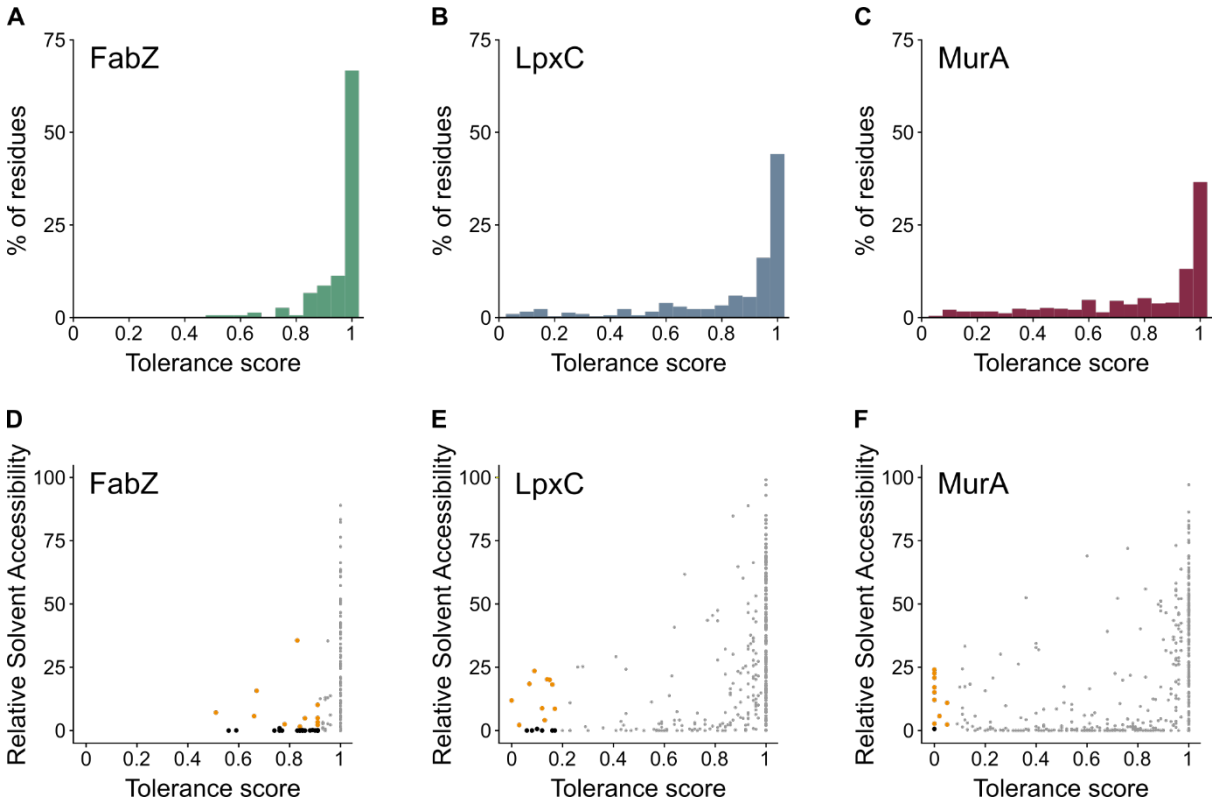

Figure 3: Analysis of the number and types of amino acid substitutions present at each position of the saturation editing libraries can be used to identify important residues. A-C) The distribution of tolerance scores, i.e. how tolerant each residue is towards substitutions with different amino acids, is shown for FabZ (A), LpxC (B) and MurA (C). D-F) The relative solvent accessibility (RSA) is plotted in function of the tolerance score for each residue of FabZ (D), LpxC (E) and MurA (F). Residues with low tolerance scores and low RSA are colored black and are likely essential for protein folding and stability. Residues with low tolerance scores and relatively high RSA are highlighted in orange and likely play an important and direct role in protein function.

Residues with low tolerance scores could be important for protein function due to several different reasons. For example, they could be part of the catalytic site, be involved in protein-protein interactions or influence protein folding and stability. To distinguish between some of these options, we calculated the Relative Solvent Accessibility (RSA) of individual residues, which is a measure for how exposed an amino acid is to the cellular environment. Residues with low RSA values are buried inside the protein and are therefore thought not to play a direct role in protein function, but rather contribute to folding and/or stability. RSA values were extracted from relevant protein structures in the PDB and are displayed in Figure 3D-F and listed in Table S2. For each protein, around 10 residues

that might play a direct role in protein function were selected for further evaluation (Figure 3D-F, Table 1). These are (partially) exposed residues (with RSA > 1%) that display the lowest tolerance scores for their respective libraries. Some of the selected residues were already previously shown to be important for protein function. These include residue H265 of LpxC, which was proposed to act as the general acid required to protonate the amino leaving group in the deacetylase reaction[42-45], and several MurA residues involved in substrate binding such as R91, R120, G164 and R397[37-40,46]. However, we also identify residues that were not yet implicated in protein function, thereby expanding our insight into these essential bacterial proteins. All selected residues together with their previously reported functions are listed in Table 1.

Table 1: Selected surface exposed residues with relatively low mutational tolerance scores for FabZ, LpxC and MurA. For each library, approximately 10 residues with the lowest mutational scores of that library were chosen. The substitutions that are encountered at these positions in the saturation editing libraries are listed as "allowed substitutions".

Protein	Residue	Tolerance score	Allowed substitutions	Previously described function
FabZ	R20	0.83	A,C,D,E,F,G,H,K,L,M,N,Q,R,S,T,V,Y	
FabZ	F23	0.86	A,C,E,F,G,H,I,K,L,M,N,Q,R,S,T,V,W,Y	
FabZ	F55	0.91	A,C,E,F,G,H,I,K,L,M,N,P,Q,R,S,T,V,W,Y	
FabZ	P62	0.78	A,C,D,E,F,G,H,I,L,M,N,P,Q,R,S,T,V	
FabZ	G63	0.67	A,C,E,F,G,K,L,M,N,P,Q,R,S,T,V	
FabZ	I66	0.91	A,C,E,F,G,H,I,K,L,M,N,P,Q,R,S,T,V,W,Y	
FabZ	E68	0.51	A,D,E,F,H,L,M,P,R,S,T	Catalytic residue[25,47,48]
FabZ	A71	0.66	A,C,D,F,G,H,I,L,M,N,Q,S,T,V,Y	
FabZ	Q72	0.85	A,C,D,E,F,G,I,K,L,M,N,P,Q,R,S,T,V	
FabZ	G108	0.91	A,C,D,E,F,G,I,K,L,M,N,P,Q,R,S,T,V,W,Y	
FabZ	L124	0.91	A,C,E,F,G,H,I,K,L,M,N,P,Q,R,S,T,V,W,Y	Substrate binding[49]
LpxC	H79	0.03	H,W	Catalytic Zn^{2+} coordination[42,44,50]
LpxC	D105	0.14	C,D,E,N,T	Protein stability[44]
LpxC	R190	0.17	F,K,M,Q,R	
LpxC	T191	0.15	C,M,N,S,T	Substrate binding[42,50], stabilizing intermediaries[43,51]
LpxC	G210	0.09	A,G,S,T	Substrate binding[42]
LpxC	K239	0.16	C,M,N	Substrate binding[45,50-53]
LpxC	D242	0.12	D,G,H	Catalytic Zn^{2+} coordination[42,50], Substrate binding[45]
LpxC	D246	0.13	D,L,Y	Catalytic activity[42-44]
LpxC	G264	0.07	A,C,G,S	
LpxC	H265	0	H	Catalytic residue[42-45]
MurA	R91	0	R	Substrate binding[37,38]
MurA	G114	0	G	
MurA	C115	0	C	Catalytic residue[38,54-56]
MurA	G118	0	G	
MurA	R120	0	/	Substrate binding[37-40]
MurA	H125	0.05	H,T	
MurA	G164	0	G	Substrate binding[37,38]
MurA	T304	0.05	S,T	
MurA	R397	0.02	K,R	Substrate binding[40], interdomain interactions[38,46]
MurA	G398	0	G	

For LpxC, our analysis revealed one residue that can only be substituted by a synonymous codon: H265. In good agreement, H265 has previously been proposed to act as the general acid required to protonate the amino leaving group in the LpxC-catalyzed deacetylase reaction[42-45]. In addition to this residue, several residues with low tolerance scores and high RSA values - which thus likely directly participate in protein function - were identified in LpxC as listed in Table 1. Not entirely unexpected, most of these residues are located within or around the active site / substrate binding pocket. D246 and G264 are located in close spatial proximity to the essential general acid H265, and mutation of these residues could thus disturb the correct orientation of the imidazole side chain of H265. In addition, G264 is located very close to the pyrophosphate groups of the UDP moiety of the substrate, and substitution with bulky amino acids would likely interfere with substrate binding. D246 directly interacts with the H265 side chain and it has previously been proposed that this interaction is required to keep H265 in the correct protonation state for its role as a general acid[42]. The latter, however, seems to contradict our observation that the D246L and D246Y mutations sustain viability. E78 was previously proposed to act as the general base which deprotonates the Zn^{2+}-bound nucleophilic water molecule[57]. Correspondingly, we find that E78 displays very limited tolerance to mutations, although it is remarkable that the E78R and E78V mutations are retrieved, while it is obvious that especially a Val residue could not maintain the role of general base. Such unexpected tolerated mutations indicate the complexity of protein viability within their *in vivo* cell context, as opposed to *in vitro* experiments. Also the residues involved in coordinating the Zn^{2+} ion (D242, H79 and H238) show a relatively limited tolerance to mutation, with H238 seemingly most tolerant (7 substitutions allowed). Another group of residues that display a limited tolerance to mutations are either located within (R190, T191, F192) or interacting with (D105), the R190-G193 region that directly interacts with the glucosamine moiety of the substrate, and which has been previously shown to be important for catalysis[51]. Also K239 makes a direct hydrogen bond with the glucosamine moiety, and a K239A mutation was previously reported to affect catalysis[51]. In this respect it is remarkable that mutations to Cys, Met and Asn are allowed, while no other mutations are identified in our analysis. A final functional category of residues with low mutational tolerance consists of residues that line the acyl-binding groove (G210 and A215). Mutation of these small residues to amino acids with larger side chains would likely affect substrate binding.

A similar analysis on MurA reveals several residues that cannot be replaced by any other amino acid. These include C115, the proposed general acid required to protonate the C3 atom of the phosphoenolpyruvate (PEP) substrate in the MurA-catalyzed reaction[58], and several other previously

described important residues involved in substrate binding and/or product release, such as R91, G164 and R371[37-40,58]. It is remarkable that we find C115 to be absolutely essential, while it was previously reported that a C115D mutation retains catalytic activity[58]. In this context it is worth mentioning that we also retrieve K22 as a residue with little tolerance to mutations. Eschenburg *et al.* proposed this residue to be the general acid that protonates PEP[39]. However, the latter proposal does not agree with the K22F and K22N mutations that we retrieve as being viable. We also identify essential residues that were until now not described to play an important/essential role in MurA, including G114, G118 and G398. G114 and G118 are located in the P112-P121 catalytic loop harboring the C115 general acid and are most likely crucial to maintain the loop conformation and sustain the required conformational changes within this loop[58]. Within the same P112-P121 loop we also identified P112 and G113 as intolerant to mutations. G398 is located adjacent to the nearly completely essential residue R397. The latter can only accommodate substitutions to R or the closely related K residue, and was proposed previously to play an important role in the product release mechanism[58]. In addition, R397 is located very close to Cys115 and might play a role in tuning the latter's pKa. Three other residues that display very little mutational tolerance are S162, T304 and D305. The side chain of S162 makes a direct hydrogen bond with the pyrophosphate of the UPD moiety of the substrate, which explains its tolerance for Thr substitution. It is quite remarkable though that we also retrieve a S162F mutation, while the latter can obviously not compensate for the hydrogen bond. D305 has previously been found essential for catalysis[59,60], and a role as general base required for deprotonation of the C3 hydroxyl of the UDP-GlcNac substrate has been proposed[38]. In principle the observed (viable) E305E and E305Y variants could take over such a role, although the available space to allow substitution of D305 with a bulkier phenol group seems limited. Finally, T304 is located adjacent to D305 and might be important to maintain the correct orientation of the latter.

The mutational analysis of FabZ presents a more complex and intriguing image, as all FabZ residues are highly tolerant to substitutions. This is particularly remarkable for the residues H54 and E68, which were proposed to act as the general base and general acid, respectively, in the FabZ-catalyzed dehydration of the β-hydroxyacyl-ACP[47,49]. Our observation that H54 can be substituted to all but one (Asp) of the other amino acids and that E68 can be substituted with ten other amino acids (including the non-polar residues A, L, M, P, F) seems hard to reconcile with an essential function for these residues. Several other (partially) surface exposed residues (RSA > 1) also show a somewhat lower tolerance score for substitutions, as listed in Table 1, although still many substitutions are allowed for all of them. A majority of these residues, including F23, F55, P62, G63, I66, A71 and G72, border the surface of the acyl-binding tunnel, and certain substitutions could potentially sterically interfere with

substrate binding. Similarly, L124 is located on the surface of FabZ, at the interaction surface with the ACP protein. Finally, R20 and G108 are located close to the subunit interfaces within the FabZ hexameric "trimer of dimer" arrangement. Despite slightly lower tolerance scores, suggesting important roles for these residues, we were surprised to observe that almost all amino acid substitutions are allowed at these FabZ residues, including substitutions expected to be impactful based on size and/or biochemical properties. Single mutations might not be sufficiently disruptive to interfere with multimer formation, thereby highlighting a fundamental restriction of this approach, which is limited to single isolated mutations and cannot investigate co-occurring mutations that might be synergistic or compensatory.

Taken together, our results demonstrate that saturation editing of essential genes combined with the identification of viable amino acid substitutions can be used to pinpoint important residues and can reveal novel insights into protein function.

Saturation editing libraries can guide efforts for the development of novel antibiotics

Finally, we aimed to exploit the saturation editing libraries of essential *E. coli* proteins FabZ, LpxC and MurA to formulate recommendations for antibiotic development. First, to identify surface exposed regions that are important for protein function and could be targeted by antimicrobial compounds, we plotted the tolerance scores for all residues onto the corresponding protein structures (Figure 4 and S1). As expected, these augmented protein structures reveal the importance of the catalytic site for protein function, but could in theory also reveal sites involved in allosteric regulation, protein-protein interactions, etcetera.

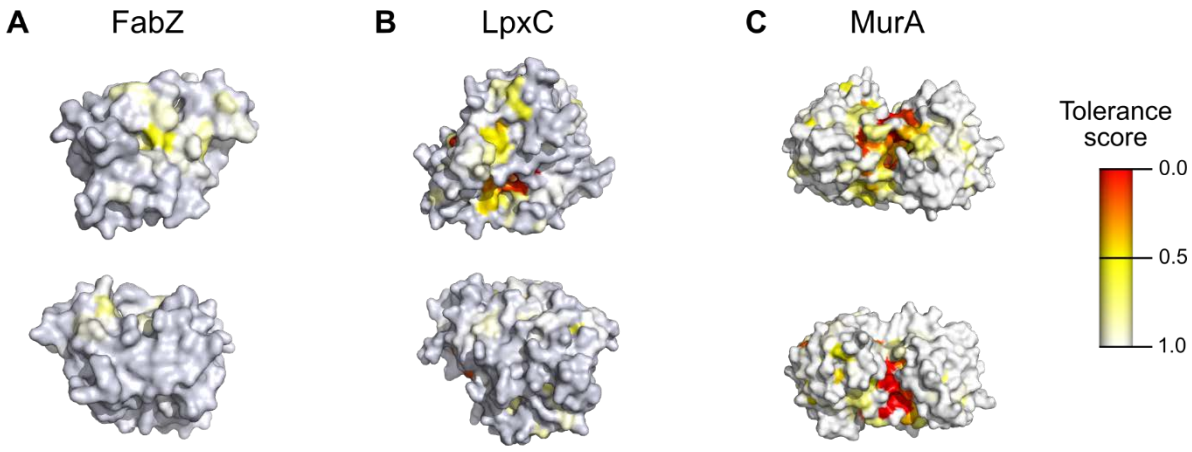

Figure 4: Protein structures colored by each residue's tolerance score reveal regions essential for protein function that can be targeted by antimicrobial compounds. A-C) Tolerance scores calculated here were plotted onto experimentally

determined protein structures for FabZ, PDB 6n3p (A); LpxC, PDB 4mqy (B); and MurA, PDB 1uae (C). For each protein, two different surfaces are shown at the top and bottom. Corresponding cartoon representations are shown in Figure S1.

Additionally, since our saturation editing libraries provide information on the protein's mutational tolerance, they can be used to make predictions regarding resistance development. This way, drug development efforts can be guided towards compounds and targets that are the least susceptible to acquiring resistance mutations. As is clear from Figures 2 and 3, the mutational flexibility of FabZ, LpxC and MurA differs strongly. Whereas the saturation levels for these libraries are almost identical (96-97%), the percentage of mutations that is tolerated varies strongly, with 96 % of mutations tolerated for FabZ (2 877 detected mutations out of 2 998 designed mutations), 86 % for LpxC (5 220 detected mutations out of 6 079 designed mutations) and 80 % for MurA (6 675 detected mutations out of 8 357 designed mutations). The same trend emerges when calculating the percentage of residues that tolerates all or all but one amino acid changes. This number reaches 83 % for FabZ, 61 % for LpxC and only 50 % for MurA (Figure 2G). Taken together, these data indicate that, even though all three proteins are essential for *E. coli* viability, their tolerance to amino acid changes differs widely, with FabZ being the most and MurA the least tolerant. Based on these data, we speculate that MurA is the least likely to develop resistance-conferring mutations when serving as an antibiotic target and is therefore the best target to pursue.

To investigate this hypothesis in more detail, we isolated library variants that are resistant to selected compounds. Fosfomycin, a known antibiotic that targets MurA directly[61], was used to select *murA* resistant variants. LpxC-targeting compounds CHIR-090[62,63] and PF-04753299 (Pfizer) were used to interrogate resistance development through *lpxC* mutations. Additionally, since it has been shown that resistance to anti-LpxC compounds can develop through mutations in *fabZ* that restore the disturbed balance between LPS and phospholipid synthesis[64-66], we also selected the FabZ library against both of these compounds. To select for resistant variants, libraries were plated onto medium containing different concentrations of the selected compounds (4x, 8x and 32x the Minimal Inhibitory Concentration (MIC), see Materials and Methods). Colonies that were able to grow overnight were selected and their *fabZ*, *lpxC* or *murA* gene was sequenced to identify potentially resistance-conferring mutations. All isolated mutations are listed per condition in Table S3, while a condensed form of these data is shown in Table 2.

Table 2: Library variants that are resistant to CHIR-090, PF-04753299 or fosfomycin. Variants were selected by plating the FabZ, LpxC or MurA library on medium with different concentrations of the indicated compounds (4x, 8x or 32x MIC). Colonies that were able to grow were sequenced to identify possible resistance-conferring mutations. This table lists the

residues that were found to be mutated together with the number of times they were targeted and the detected amino acid substitutions (in subscript). For each library-compound combination, 35 resistant clones were sequenced. Wt, wild type.

FabZ			
CHIR-090		PF-04753299	
Mutated Residue	Times found	Mutated Residue	Times found
T3 $_{E}$	1		
L8 $_{R}$	1		
H19 $_{Q}$	1		
		F23 $_{H/T}$	2
		L25 $_{K}$	1
G35 $_{Y}$	1		
		F51 $_{C/P}$	2
		I60 $_{C}$	1
		A71 $_{G}$	2
G75 $_{S}$	1		
		L90 $_{C}$	1
Y92 $_{A/S/T}$	3	Y92 $_{Q/T}$	2
F93 $_{I}$	1	F93 $_{I}$	4
G95 $_{Q/V}$	2		
I96 $_{W}$	1	I96 $_{W}$	1
E98 $_{P}$	1		
A99 $_{L}$	1		
R100 $_{C/S/T/Y}$	6	R100 $_{Y}$	1
F101 $_{V}$	1		
		K102 $_{L}$	1
D109 $_{L}$	1		
R121 $_{G}$	1	R121 $_{F/Q}$	2
R122 $_{W}$	1		
L124 $_{K/N}$	2		
T125 $_{H}$	1		
R126 $_{N/Q}$	2	R126 $_{G/H}$	4
F127 $_{Q/W}$	2		
		G129 $_{I/L}$	6
V138 $_{L}$	1		
		C139 $_{Y/L}$	2
		A141 $_{M/Q}$	3
M144 $_{H/R}$	2		
A146 $_{R}$	1		
wt	0	wt	0

LpxC			
CHIR-090		PF-04753299	
Mutated Residue	Times found	Mutated Residue	Times found
I2 $_{K/Y}$	2	I2 $_{K/Y}$	3
R5 $_{K/N}$	6	R5 $_{K/N/S/V}$	13
R9 $_{H}$	1		
		I10 $_{E}$	1
V11 $_{E/I/K/N/T}$	9	V11 $_{E/F/I/T}$	9
Q12 $_{H/N}$	3		
T14 $_{D/E/F/S/Y}$	8	T14 $_{C/F/Y}$	7
G15 $_{H/N}$	2	G15 $_{H}$	1
V16 $_{E}$	1		
L18 $_{C/V}$	2		
A31 $_{N}$	1		
		T35 $_{D}$	1
wt	0	wt	0

MurA	
Fosfomycin	
Mutated Residue	Times found
V16 $_{M}$	1
D51 $_{A}$	1
V228 $_{I}$	1
R267 $_{E}$	1
I402 $_{W}$	1
wt	30

Whereas all colonies from the FabZ or LpxC libraries selected for resistance to either CHIR-090 of PF-04753299 carry a mutation in respectively *fabZ* or *lpxC*, this is not true for selection of the MurA library against fosfomycin. In this case, the vast majority of selected resistant clones still contain a wild-type *murA* gene, pointing towards the existence of spontaneous resistance mutations that arose elsewhere in the genome. Indeed, when comparing the number of resistant variants present in the libraries to the number of spontaneous resistant variants present in a culture of the wild-type strain, they are highly similar when selecting for fosfomycin resistance (Figure 5A-C). These data thereby confirm that fosfomycin resistance mostly arises through spontaneous resistance mutations that are not located in the *murA* gene. Nonetheless, a few *murA* variants were picked up when selecting the MurA library for fosfomycin resistance. However, each of these mutations was only found once, thereby making us question their role in mediating fosfomycin resistance. To check whether these *murA* mutations are

causal to resistance or are hitchhikers present in a genome that also contains spontaneous resistance mutations elsewhere, we transferred the *murA* mutant alleles to a clean genetic background that has never been exposed to fosfomycin. Since none of these transferred mutations were able to increase MIC levels towards fosfomycin (Table S4), we conclude that also for these selected variants, causal spontaneous mutations are located elsewhere in the genome. In conclusion, not a single *murA* mutation could be found that provides resistance to fosfomycin.

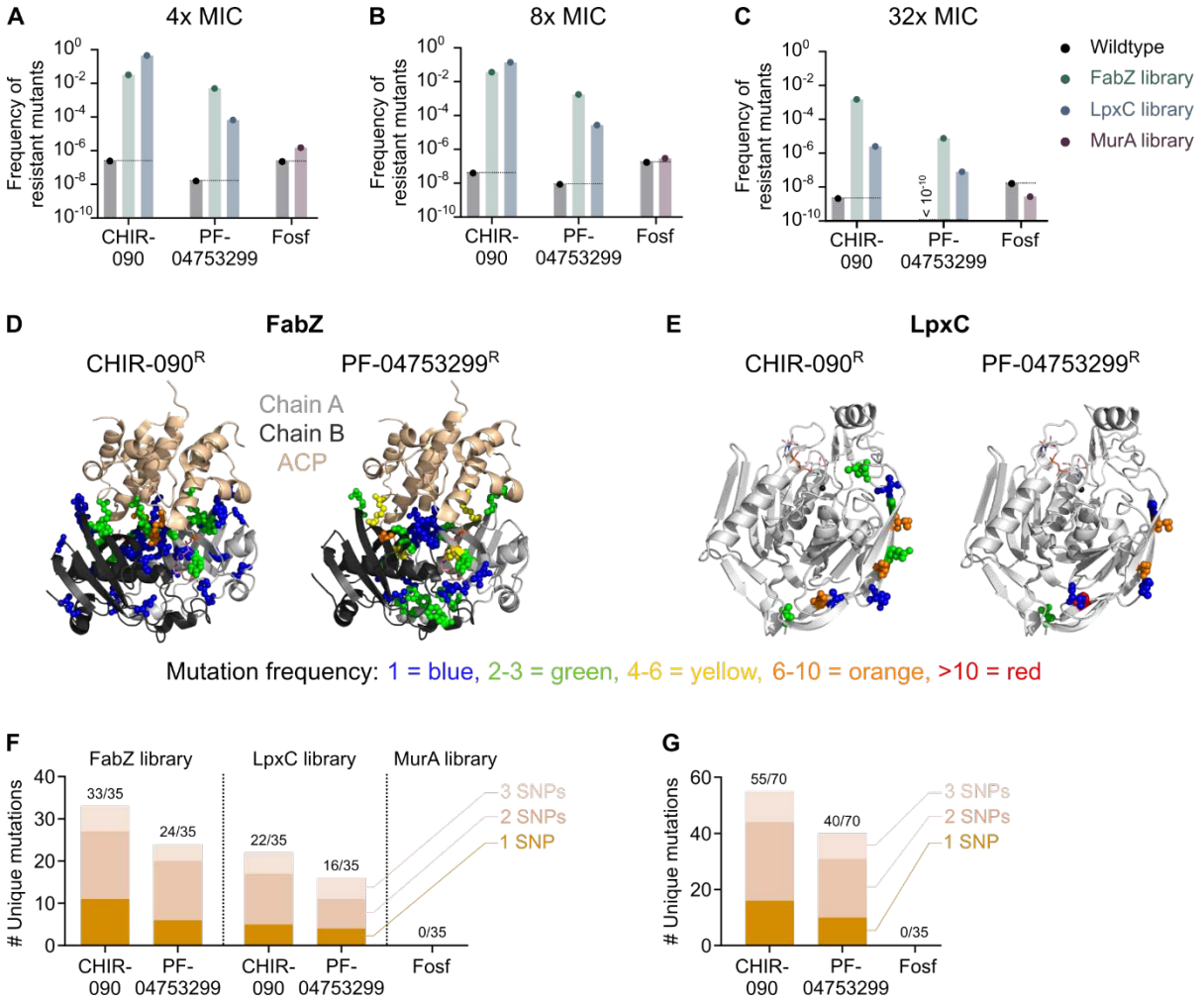

Figure 5: Saturation editing libraries can guide efforts for the development of novel antibiotics. A-C) The frequency of occurrence of spontaneous resistance mutations is compared to the frequency of occurrence of resistant variants in the saturation editing libraries. This was done by plating either a wild-type culture or the different libraries onto medium containing different concentrations of the compound, i.e. 4x MIC (A), 8x MIC (B) or 32x MIC (C), and counting the number of colonies that developed after overnight growth. These numbers were normalized to the total cell numbers present in the wild-type culture or the libraries, respectively. D) The location of targeted residues in the FabZ protein is shown for both CHIR-090 and PF-04753299. Residues are colored according to the number of times they were targeted in isolated resistant variants. Only one dimer of the FabZ hexamer is shown for clarity. Mutations are indicated in both chain A and B. E) The location of targeted residues in the LpxC protein is shown for both CHIR-090 and PF-04753299. Residues are colored according to the number of times they were targeted in isolated resistant variants. F-G) The number of unique mutations in *fabZ* and *lpxC* that provide resistance to CHIR-090 or PF-04753299 and the number of mutations in *murA* that provide

resistance to fosfomycin are shown, either grouped per library and compound (F) or grouped per compound only (G). For each library-compound combination, 35 resistant variants were isolated and their *fabZ*, *lpxC* or *murA* gene was sequenced. The mutations found are subdivided into categories based on the minimal number of SNPs necessary to provide the observed amino acid change. Fosf, fosfomycin; ACP, acyl carrier protein.

On the other hand, Figure 5A-C shows that the number of variants resistant to CHIR-090 or PF-04753299 from either the FabZ or LpxC libraries exceeds the number of spontaneous resistant variants by several orders of magnitude, indicating that the isolated *fabZ* and *lpxC* mutations are likely causal to resistance.

Table 2 and Table S3 show that there is a large variety in possible *fabZ* mutations that provide resistance against either CHIR-090 or PF-04753299. In fact, out of the 35 sequenced variants from the FabZ library that were either resistant against CHIR-090 or PF-04753299, 33 or 24 unique mutations were found, respectively, indicating that the search for resistant variants was not saturated and that additional resistance-conferring mutations probably exist. Mapping the isolated mutations onto the FabZ protein structure demonstrates that the resistance-conferring *fabZ* mutations occur throughout the entire protein with a few preferred hotspots for mutations (Figure 5D).

Similarly, *lpxC* also displays hotspots for resistance-conferring mutations. However, the number of different resistance-conferring mutations in LpxC is much more limited than for FabZ. Out of the 35 sequenced variants from the LpxC library that were either resistant against CHIR-090 or PF-04753299, 22 and 16 unique mutations were found, respectively. Given the relatively large number of *lpxC* mutations that were isolated multiple times, we suspect that our selection for resistant variants was more or less saturated and that most resistance-conferring *lpxC* mutations were identified. Interestingly, these mutations are exclusively found in the N-terminus of the protein (Figure 5E) which for CHIR-090[63], and presumably also PF-04753299, is not where the compound binds.

To estimate how likely spontaneous mutations in *fabZ*, *lpxC* or *murA* are to generate resistance, we classified the number of unique resistance-conferring mutations selected here according to the minimal number of (SNPs) needed to result in the corresponding amino acid substitution (Figure 5F-G). From these data it is clear that there are more 1 SNP mutations located in *fabZ* and *lpxC* that provide resistance against CHIR-090 than PF-04753299, meaning that resistance can likely more easily develop against CHIR-090. PF-04753299 is therefore a more attractive anti-LpxC compound. No mutations in *murA* were identified to provide resistance against fosfomycin, although we have

established that spontaneous resistance mutations to this antibiotic can easily arise elsewhere in the genome.

Discussion

We have performed saturation editing on three essential *E. coli* proteins using high-throughput CRISPR genome editing to identify amino acid residues that are important for protein function. We were able to confirm the role of previously annotated residues while also providing new insights into protein function and formulating recommendations for antibiotic development.

To identify essential residues, we have introduced a tolerance score that reflects how well amino acid changes are tolerated at each position in a protein. By combining this tolerance score with the relative solvent accessibility (RSA) of each residue, we made a distinction between residues that are likely important for protein folding or stability (low tolerance score and low RSA) and residues more directly involved in protein function (low tolerance score and high RSA). This way, we identified several important or essential residues in each protein, some of which were not previously known to play an important role. Notably, our results highlight some key differences with previously obtained results from *in vitro* protein activity tests. For example, in MurA we find C115 to be completely intolerant to mutations *in vivo*, while previously a C115D mutant was shown to retain activity *in vitro*. On the other hand, in FabZ, H54 and E68 were proposed to act as a general base and acid, respectively, while we find these residues to be tolerant to substitutions. These findings stress the importance of complementing any insights obtained *in vitro* with experiments that interrogate behavior in the much more complex *in vivo* setting.

Surprisingly, we found that protein function is very robust in light of mutations. The vast majority of single amino acid substitutions still support cell viability and protein function. No less than 96 % of the designed *fabZ* mutations allow for viable progeny. This number is 86 % for LpxC, and 80 % for MurA. Moreover, 83, 61 and 50 % of residues could be changed to all or all but one amino acid for the FabZ, LpxC and MurA proteins, respectively. Therefore, although all three proteins display a surprisingly high tolerance for amino acid changes, this tolerance level differs widely with FabZ being the most and MurA the least tolerant. This difference could in part be related to the surface-to-volume ratio of different proteins and, related, the percentage of exposed residues. Our data (Figure 2D-F) indeed confirm that surface exposed residues can more easily tolerate mutations than buried residues[67,68]. Larger proteins that tend to have a larger surface-to-volume ratio and a higher percentage of exposed residues are therefore expected to tolerate a higher percentage of mutations than proteins that tend

to have more buried residues. Indeed, FabZ, the smallest protein investigated tolerates the highest number of mutations, while MurA, the largest protein, tolerates the least. Also the percentage of exposed and buried residues fits well with this explanation. The percentage of exposed residues (RSA ≥ 20) for FabZ, LpxC and MurA is 36, 49 and 43 % respectively, while the percentage of fully buried residues (RSA = 0) for these proteins is 10, 10 and 15 %. However, we suspect that this size effect cannot fully explain the remarkably high mutational flexibility of FabZ. In fact, every amino acid of FabZ – including the proposed catalytic residues – could be replaced by at least 10 different amino acids. Nonetheless, the *fabZ* gene was shown to be essential for *E. coli* in multiple studies[69-71]. We therefore hypothesize that either the catalytic activity of FabZ is not the essential function of this protein, that some limited redundancy between FabZ's catalytic activity and other *E. coli* enzymes exist that can partially take over FabZ's function when the activity of this protein is decreased due to impactful amino acid substitutions and/or that no single mutation reduces FabZ activity below a threshold required for viability. The latter explanation would mean that there usually is an excess of FabZ activity in the cell.

The large difference in tolerance to mutations between these three potential drug targets also has implications for drug development. Since FabZ displays such a high mutational flexibility, including for residues located in the catalytic site, FabZ can most likely easily be mutated to block drug-target interactions and thereby obtain resistance against FabZ-targeting compounds. LpxC and MurA on the other hand are much less tolerant to mutations and contain several residues that can only be replaced by a few very specific amino acids. Especially MurA contains many accessible residues that are essential for viability and that could thus be targeted by antimicrobial compounds. Based on our analysis, MurA therefore appears to be the most promising drug target, closely followed by LpxC. FabZ on the other hand does, in our opinion, not hold much potential.

Finally, we have tested these predictions by probing resistance development against existing antibacterial compounds through target modification. Based on the selection for *murA* variants that are resistant to the MurA-targeting antibiotic fosfomycin[61], we confirm that MurA is indeed a highly attractive antibiotic target. Even though several resistant variants could be isolated from the MurA saturation editing library, none of them carried a *murA* mutation that was causal to fosfomycin resistance. Nonetheless, a few resistant *murA* alleles have previously been described. For example, purified MurA C115D or MurA C115E displayed resistance to fosfomycin *in vitro*[54]. It was suggested that, while these mutations to either Asp or Glu delete the target residue for covalent attachment of fosfomycin, they maintain the possibility to take over the role of Cys to act as the general acid in the

enzyme-catalyzed reaction[54]. Indeed, although the activity of MurA was shown to be severely affected by the C115E mutation, MurA C115D retained high activity *in vitro*[54]. However, neither of these mutations were detected here. In fact, these mutations are completely absent from the MurA saturation editing library, suggesting that they could not support viability *in vivo*. This finding highlights the potential differences between *in vitro* and *in vivo* behavior and stresses the need to investigate gene function as close to its natural context as possible, although it remains possible that both of these variants are missing from the library due to a combination of incomplete saturation levels and random chance. Additionally, two other *murA* mutations that provide resistance against fosfomycin were detected in clinical *E. coli* strains, D369N and L370I[72]. The MurA D369N mutation is present in our MurA library but was not picked up when selecting for resistant variants. This could be due to the relatively high frequency of other spontaneous suppressor mutants that were preferentially isolated compared to MurA D369N. The L370I mutation, on the other hand, was not present in the library either because of insufficient saturation levels or because this mutant does not support viability in the *E. coli* lab strain used in this study. Taken together, we conclude that MurA is a very attractive target for new antibiotics because it cannot be easily mutated to overcome direct inhibition by antimicrobial compounds. Moreover, the MurA-fosfomycin combination is excellent in terms of resistance development through target modification. However, other mechanisms that provide resistance against fosfomycin exist and would have to be overcome to fully benefit from the powerful MurA-fosfomycin combination. Indeed, it is known that, *in vitro*, resistance to fosfomycin develops more easily through mutations that limit the import of this antibiotic into the cell than through mutations in the target MurA itself[61]. Thankfully, these import-limiting mutations are rare *in vivo* since they come with a considerable biological cost[73].

Likewise, we selected mutations in *lpxC* and *fabZ* that provide resistance against the LpxC-targeting compounds CHIR-090[62,63] and PF-04753299 (Pfizer). Surprisingly, all identified resistance-conferring mutations in *lpxC* are located in the N-terminal part of the protein. However, based on the experimentally determined structure of *Aquifex aeolicus* LpxC bound to CHIR-090[63], it seems highly unlikely that any of these mutated residues are directly involved in compound binding. Instead, since all the amino-acid substitutions we identify are located in the 5'-end of the gene, it is possible that they provide resistance by altering protein levels. All the identified mutations encode an amino-acid substitution and one or multiple synonymous PAM-site mutations. This change in codon usage at the start of the gene could alter expression levels by a variety of mechanisms, such as an altered speed of translation, changes in transcript stability or others[36]. Rather than influencing compound binding that occurs at an entirely different location, it therefore seems plausible that these 5'-end mutations would

increase resistance levels by influencing protein production. Alternatively, the N-terminal domain could be involved in a previously undescribed regulatory mechanism that influences LpxC activity. It is perhaps unsurprising that no resistant *lpxC* mutations were found at the compound binding site. CHIR-090 is known to bind at the catalytic site[63] and we demonstrate that tolerance scores for residues found at or near this site are very low, meaning that not many mutations are tolerated and that therefore not many potentially resistance-conferring mutations are available at this location. To the best of our knowledge, no mutants with single-amino-acid substitutions in LpxC resistant to CHIR-090 or PF-04753299 have been reported previously.

Although CHIR-090 and PF-04753299 target LpxC, it is known that mutations in *fabZ* can provide resistance against anti-LpxC compounds[64-66]. Many such resistance mutations were detected throughout the entire FabZ protein. These include several residues that are also targeted in previously discovered mutants resistant to anti-LpxC compounds[64-66]. These *fabZ* mutations are believed to provide resistance against anti-LpxC drugs by lowering the activity of FabZ and thereby restoring the balance between phospholipid synthesis and LPS production[65]. It is therefore not surprising that so many different mutations in FabZ were isolated; any mutation that lowers FabZ activity appropriately is expected to provide resistance.

Apart from prioritizing potential antibiotic targets, we can also rank lead compounds based on the likeliness of resistance development. From our experiments using two anti-LpxC compounds, it became clear that PF-04753299 is superior to CHIR-090 from a resistance development point of view. Taken FabZ and LpxC together, there are less mutations – and importantly, less 1 SNP mutations – that confer resistance against PF-04753299 than CHIR-090. We therefore expect that, also *in vivo*, resistance is less likely to develop against PF-04753299, which is an important advantage for further drug development.

Taken together, we here present a deep mutational scanning approach that directly targets the *E. coli* genome and is able to interrogate the effect of selected mutations *in vivo* in its natural genomic context. We have used this approach to study the importance of individual amino acids in the function of three essential proteins involved in *E. coli* cell envelope synthesis. Additionally, we have exploited the CRISPR generated saturation editing libraries to formulate recommendations for antibiotic development based on predictions of the ease of resistance development. Our work may therefore contribute to future endeavors to select and validate targets for the development of new antibiotics.

Methods

Bacterial strains, compounds and growth conditions

Experiments were performed with *E. coli* SX43[74], a derivative of BW25993, except for CRISPR-FRT where *murA* mutations were transferred to *E. coli* BW25113 Δ*sfsB*[69]. Cultures were grown on/in SOB growth medium with/without 1.5 % agar. They were incubated at 37°C with continuous shaking at 200 rpm for liquid cultures, except for performing CRISPR-FRT which was done at 30°C.

Compounds used include CHIR-090 (VWR International), PF-04753299 (Sigma-Aldrich) and fosfomycin (TCI Europe) at different concentrations, as indicated in the text. Additionally, gentamicin (25 µg/ml), kanamycin (40 µg/ml), spectinomycin (50 µg/ml) and anhydrotetracycline (100 ng/ml) were used where appropriate.

Saturation editing library construction

Saturation editing libraries were constructed using high-throughput CRISPR-based editing provided by the Onyx™ Digital Genome Engineering platform. Briefly, repair templates were designed using Inscritpa's Designer software (development version) so that each amino acid would be replaced by every other amino acid and so that every codon would be replaced by a synonymous codon (if a synonymous codon exists). Besides the desired mutation, each oligo may also contains one or more synonymous edits that prevent re-cutting by eliminating the PAM site and/or introducing edits that interfere with cutting. For each mutation present in the repair template, the most frequently used available codon was chosen. 'Barcode Plasmids' containing the repair template, corresponding sgRNA and unique barcode, were cloned in bulk into a high-copy plasmid backbone. These Barcode Plasmid populations were then transformed into cells encoding inducible expression of MAD7, a type V CRISPR nuclease from *Eubacterium rectale*, and the lambda Red Recombination enzymes on an `Engine Plasmid`. Genome editing was performed using developmental reagents and protocols optimized for *E. coli* MG1655 (Onyx™ Engineering Handbook *E. coli* and *S. cerevisiae*. 2022. 1001178, https://inscripta.showpad.com/share/rWxQsFGmJznLKrdfWBHGH). Three libraries were built each targeting a different gene (See Figure 1B). After editing, each library was grown for approximately 8 hours in LB supplemented with 1000 µg/ml carbenicillin and 68 µg/ml chloramphenicol, sampled for sequencing and stored at -80°C in 25 % glycerol to be used in further experiments.

Illumina sequencing for detection of mutations in saturation editing libraries

Genomic DNA was isolated according to standard Inscripta protocols (Onyx™ Genotyping Handbook *E. coli* and *S. cerevisiae*. 2022. 1001182 RevB.). PCR amplification of 2kb genomic regions flanking each gene of interest (*fabZ*, *lpxC*, *murA*) was carried out in a reaction mixture containing 10 ng plasmid DNA,

and 1 µl each of gene-specific forward and reverse primers at 20 µM (Table 3) in a Q5 Hot-Start PCR Master Mix to a total volume of 50 µL. Cycling was carried out in BioRad T100 Thermal cycler instrument as follows: 98°C for 2 min; 17 cycles of 98°C for 10 s, 60°C for 10 s, 72°C for 1 minute with a final extension at 72°C for 5 minutes. PCR fragments were purified and prepared for sequencing according to the Inscripta Onyx™ Genotyping Handbook (Onyx™ Genotyping Handbook *E. coli* and *S. cerevisiae*. 2022. 1001182 RevB. https://inscripta.showpad.com/share/rWxQsFGmJznLKrdfWBHGH).

Table 3: Primers used in this study.

Primer name	Primer sequence
*fabZ*_F	TTGGCGACAATACGGCGGTTG
*fabZ*_R	CCCAGACTGACGGACTGACGTAATG
*murA*_F	CAGGAGTATAGTGATGCTCGACAGAAGAAGTG
*murA*_R	CGCAACTTTGCTCTAAGATGTTTCGCTG
*lpxC*_F	AAGACCGTGCGGAAGAAGCTG
lpxC_R	CCTGAAGAGGCAAAGATTCTTCAGCAACG
SPI12880	GGTTTCGAGGCTCTTTGTGC
SPI12881	AAGAAAACAGCGTTCGCACC
SPI12882	CTACCATGATCCGCAGACCC
SPI12883	GGTCTATGGTCCGCTGATGG
SPI12884	CAACCCAACAAAGTCTGGCG
SPI12885	TTCCACAATGGCGGTTGGAT
SPI14680	GGAAAATAATGAAATTCAGAGCGTGTTGATGAACG
SPI14681	GCCGATGGCGAACAGTTAGGTAAAAATG

Amplified 2 kb genomic regions were sequenced as 150 bp paired-end reads on an Illumina NextSeq. Designs were quantified using Inscripta's proprietary genomic amplicon edit detection pipline. Briefly, this approach uses competitive alignment to determine the origin of each read. Each read was aligned to a set of contigs specifying the unedited reference genome, an alternative set of genomic contigs where, for each design, the repair template (corresponding to Onyx edits) is appended with 1000 nucleotides of flanking genomic sequence on both sides, and a set of contigs containing design sequences flanked by 500 nucleotides of adjacent cassette backbone sequences (accounting for reads attributable to the Barcode plasmid). Alignment was performed using BWA-MEM with default settings, followed by bamsormadup for sorting and duplicate marking. Each design was categorized as providing evidence for the complete intended edit ('Onyx edit' reads), no edit ('reference' reads), or some sequence matching neither the complete intended edit or reference sequence ('other' reads), including ambiguously mapped reads.

Calculating tolerance scores

The tolerance scores were calculated by using a modified version of the Zvelebil similarity score[41], which is based on counting key differences between amino acids. For each sequence position, the score was calculated for each tolerant non-synonymous mutation using the characteristics 'small', 'aliphatic', 'proline', 'negative', 'positive', 'polar', 'hydrophobic' and 'aromatic', whereby for each

difference in each of these characteristics between the original and mutated amino acid, a score 0.1 was added to a starting score of 0.1 (effectively a reversal of the original Zvelebil score). Per sequence position, all scores for tolerant mutations were summed, then normalized by dividing by the maximum difference score possible for the original amino acid (the score obtained when mutated to all other possible amino acids). A score of 1.0 therefore indicates full tolerance in that position, a score of 0.0 no tolerance, and higher in-between scores increasing levels of tolerance for that amino acid type, with mutations to dissimilar amino acids contributing more.

Protein structures & RSA values

Protein structures shown and used in this manuscript were obtained from the PDB, with the codes and corresponding references for the used protein structures: FabZ PDB 6n3p, chain A[49], LpxC PDB 4mqy, chain A[75], MurA PDB 1uae, chain A[38]. RSA (Relative Solvent Accessibility) values were extracted from these single chain protein structures using the PoPMuSiC software[76].

MIC tests

MIC tests were performed according the broth dilution method in SOB medium. Briefly, OD 625 nm of overnight cultures was adjusted to 0.1. Cells were then diluted 200 times and 2-fold dilution series of the tested compounds were added. Cells were incubated for 24h, after which OD 595 nm was measured. MIC values were chosen as the lowest concentration of added compound that led to OD 595 nm values < 10 % of OD 595 nm values of the untreated control. For all MIC tests, 3 biological repeats were performed, each consisting of 2 technical replicates per condition. The MIC value most frequently encountered and/or centered in between all detected values was chosen as the final MIC. MIC values of *E. coli* SX43 for CHIR-090, PF-04753299 and fosfomycin were determined to be 0.032 µg/ml, 0.5 µg/ml and 8 µg/ml, respectively.

Selection and identification of resistant variants

To select variants from the FabZ, LpxC or MurA libraries resistant to CHIR-090, PF-04753299 or fosfomycin, various amounts of frozen library stocks were plated onto SOB agar plates containing one of these compounds at a concentration of 4x, 8x or 32x MIC. After overnight incubation, colonies that were able to form under these conditions were counted to determine the resistant CFUs/ml to each different concentration. Simultaneously, library stocks were also plated on non-selective SOB agar plates to determine total cell concentrations in each of the stocks. The number of resistant CFUs/ml was then normalized to the total cell concentration of the library to determine the frequency of resistance. Additionally, to compare the occurrence of resistance between the saturation editing libraries and a wild-type strain, various amounts of a wild-type SX43 culture were plated onto non-selective SOB plates and plates containing 4x, 8x or 32x MIC of the various compounds. After overnight

incubation, resistant CFUs/ml to each different concentration were determined by counting colonies on selective plates. These numbers were normalized to the total cell concentration in the overnight culture.

To confirm that colonies able to form on selective plates after overnight incubation are indeed resistant, they were transferred to the wells of a microtiter plate containing 200 µl SOB medium supplemented with the compound at the concentration used for initial selection of the specific clones. If clones were able to grow overnight, they were deemed resistant with an MIC value higher or equal to the concentration of compound used for selection of this strain.

To identify mutations in these resistant strains, their *lpxC*, *murA* or *fabZ* gene (dependent on the library the clones were isolated from) was amplified by PCR using primer pairs SPI12880 & SPI12881, SPI12882 & SPI12883 or SPI12884 & SPI12885, respectively (Table 3). PCR products were sent for Sanger sequencing using the same primers that were used for amplification.

Due to the absence of mutations in *murA* for the MurA library selected on fosfomycin, resistance of all selected clones was again confirmed with a MIC test (only 1 biological repeat performed). All isolated clones displayed an increase in MIC that was equal to or greater than the concentration of fosfomycin on which they were selected.

CRISPR-FRT to transfer mutations to new genetic background

Selected *murA* mutations were transferred to a clean genetic background using CRISPR-FRT[77], a modified CRISPR-Cas protocol that targets FRT sites present in the *E. coli* Keio library. The Keio library is a collection of around 4000 mutants that all contain a different single-gene deletion where the gene in question is replaced by an FRT-KmR-FRT cassette[69]. By targeting FRT sites using CRISPR-Cas and providing a rescue oligo that contains flanking homologous region, the deleted gene can be replaced with any desired sequence present on the rescue oligo in between the homologous regions. Selected *murA* mutant alleles, together with extended up- and downstream regions, were amplified from selected resistant variants using primers SPI14680 & SPI14681. These PCR products were used as rescue oligos. They were transformed to an *E. coli* BW25113 Δ*sfsB* mutant that contains the CRISPR-Cas editing plasmids pCas9CR4-Gm and pKDsgRNA-FRT[77]. *sfsB* is a non-essential gene just downstream of the essential *murA* for which no deletion mutant is present in the Keio library. CRISPR-Cas editing was performed as described before[77] and colonies were selected for their loss of kanamycin resistance. Colony PCR on the *murA* gene of selected clones was performed using primers SPI12882 & SPI12883 and PCR products were sequenced with the same primers to confirm the presence of the transferred *murA* mutations.

Acknowledgements

The work was supported by grants from the Research Foundation Flanders (FWO) (G0B0420N, G055517N, G0C4322N, G0I1522N), KU Leuven (C16/17/006), Francqui Research Foundation, VIB Technology Watch and VIB. LD received a postdoctoral mandate from the FWO.

Author contributions

Conceptualization: LD, ANB, KN, NK, WV, WV, JM; Methodology: LD, ANB, KN, NK, WV, WV; Formal analysis: LD, ANB, WV, WV; Investigation: LD, ANB, KN; Writing – Original Draft: LD, WV, WV; Writing – Review & Editing: LD, ANB, KN, NK, WV, WV, JM; Visualization: LD, WV, WV.

Competing interests

ANB, KN and NK are affiliated with Inscripta, Inc. The other authors declare no competing interests.

References

1 Findlay, G. M. *et al.* Accurate classification of BRCA1 variants with saturation genome editing. *Nature* **562**, 217-222, doi:10.1038/s41586-018-0461-z (2018).

2 Sun, Z. & Palzkill, T. Deep Mutational Scanning Reveals the Active-Site Sequence Requirements for the Colistin Antibiotic Resistance Enzyme MCR-1. *mBio* **12**, e0277621, doi:10.1128/mBio.02776-21 (2021).

3 Sarkisyan, K. S. *et al.* Local fitness landscape of the green fluorescent protein. *Nature* **533**, 397-401, doi:10.1038/nature17995 (2016).

4 Matreyek, K. A. *et al.* Multiplex assessment of protein variant abundance by massively parallel sequencing. *Nat Genet* **50**, 874-882, doi:10.1038/s41588-018-0122-z (2018).

5 Dunham, A. S. & Beltrao, P. Exploring amino acid functions in a deep mutational landscape. *Mol Syst Biol* **17**, e10305, doi:10.15252/msb.202110305 (2021).

6 Zheng, J., Guo, N. & Wagner, A. Selection enhances protein evolvability by increasing mutational robustness and foldability. *Science* **370**, doi:10.1126/science.abb5962 (2020).

7 Puchta, O. *et al.* Network of epistatic interactions within a yeast snoRNA. *Science* **352**, 840-844, doi:10.1126/science.aaf0965 (2016).

8 Fowler, D. M. & Fields, S. Deep mutational scanning: a new style of protein science. *Nat Methods* **11**, 801-807, doi:10.1038/nmeth.3027 (2014).

9 Fowler, D. M., Stephany, J. J. & Fields, S. Measuring the activity of protein variants on a large scale using deep mutational scanning. *Nat Protoc* **9**, 2267-2284, doi:10.1038/nprot.2014.153 (2014).

10 Firnberg, E. & Ostermeier, M. PFunkel: efficient, expansive, user-defined mutagenesis. *PLoS One* **7**, e52031, doi:10.1371/journal.pone.0052031 (2012).

11 Jain, P. C. & Varadarajan, R. A rapid, efficient, and economical inverse polymerase chain reaction-based method for generating a site saturation mutant library. *Anal Biochem* **449**, 90-98, doi:10.1016/j.ab.2013.12.002 (2014).

12 Bloom, J. D. An experimentally determined evolutionary model dramatically improves phylogenetic fit. *Mol Biol Evol* **31**, 1956-1978, doi:10.1093/molbev/msu173 (2014).

13 Melnikov, A., Rogov, P., Wang, L., Gnirke, A. & Mikkelsen, T. S. Comprehensive mutational scanning of a kinase in vivo reveals substrate-dependent fitness landscapes. *Nucleic Acids Res* **42**, e112, doi:10.1093/nar/gku511 (2014).

14 Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-821, doi:10.1126/science.1225829 (2012).
15 Hart, T. *et al.* High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. *Cell* **163**, 1515-1526, doi:10.1016/j.cell.2015.11.015 (2015).
16 Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343**, 80-84, doi:10.1126/science.1246981 (2014).
17 Roy, K. R. *et al.* Multiplexed precision genome editing with trackable genomic barcodes in yeast. *Nat Biotechnol* **36**, 512-520, doi:10.1038/nbt.4137 (2018).
18 Garst, A. D. *et al.* Genome-wide mapping of mutations at single-nucleotide resolution for protein, metabolic and genome engineering. *Nat Biotechnol* **35**, 48-55, doi:10.1038/nbt.3718 (2017).
19 White, S. W., Zheng, J., Zhang, Y. M. & Rock. The structural biology of type II fatty acid biosynthesis. *Annu Rev Biochem* **74**, 791-831, doi:10.1146/annurev.biochem.74.082803.133524 (2005).
20 Trent, M. S. Biosynthesis, transport, and modification of lipid A. *Biochem Cell Biol* **82**, 71-86, doi:10.1139/o03-070 (2004).
21 Barreteau, H. *et al.* Cytoplasmic steps of peptidoglycan biosynthesis. *FEMS Microbiol Rev* **32**, 168-207, doi:10.1111/j.1574-6976.2008.00104.x (2008).
22 Erwin, A. L. Antibacterial Drug Discovery Targeting the Lipopolysaccharide Biosynthetic Enzyme LpxC. *Cold Spring Harb Perspect Med* **6**, doi:10.1101/cshperspect.a025304 (2016).
23 Jukic, M., Gobec, S. & Sova, M. Reaching toward underexplored targets in antibacterial drug design. *Drug Dev Res* **80**, 6-10, doi:10.1002/ddr.21465 (2019).
24 Han, H. *et al.* The fungal product terreic acid is a covalent inhibitor of the bacterial cell wall biosynthetic enzyme UDP-N-acetylglucosamine 1-carboxyvinyltransferase (MurA). *Biochemistry* **49**, 4276-4282, doi:10.1021/bi100365b (2010).
25 Zhang, L. *et al.* Structural basis for catalytic and inhibitory mechanisms of beta-hydroxyacyl-acyl carrier protein dehydratase (FabZ). *J Biol Chem* **283**, 5370-5379, doi:10.1074/jbc.M705566200 (2008).
26 Christaki, E., Marcou, M. & Tofarides, A. Antimicrobial Resistance in Bacteria: Mechanisms, Evolution, and Persistence. *J Mol Evol* **88**, 26-40, doi:10.1007/s00239-019-09914-3 (2020).
27 O'Neill, J. Tackling drug-resistant infections globally: final report and recommendations. *The review on antimicrobial resistance* (2016).
28 Cassini, A. *et al.* Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant bacteria in the EU and the European Economic Area in 2015: a population-level modelling analysis. *Lancet Infect Dis* **19**, 56-66, doi:10.1016/S1473-3099(18)30605-4 (2019).
29 OECD. Stemming the superbug tide: just a few dollars more. *OECD Publishing, Paris*, https://doi.org/10.1787/9789264307599-en (2018).
30 Tacconelli, E. *et al.* Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. *Lancet Infect Dis* **18**, 318-327, doi:10.1016/S1473-3099(17)30753-3 (2018).
31 Theuretzbacher, U. *et al.* Critical analysis of antibacterial agents in clinical development. *Nat Rev Microbiol* **18**, 286-298, doi:10.1038/s41579-020-0340-0 (2020).
32 Qi, L. S. *et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173-1183, doi:10.1016/j.cell.2013.02.022 (2013).
33 Zwart, M. P. *et al.* Unraveling the causes of adaptive benefits of synonymous mutations in TEM-1 beta-lactamase. *Heredity (Edinb)* **121**, 406-421, doi:10.1038/s41437-018-0104-z (2018).
34 Lebeuf-Taylor, E., McCloskey, N., Bailey, S. F., Hinz, A. & Kassen, R. The distribution of fitness effects among synonymous mutations in a gene under directional selection. *Elife* **8**, doi:10.7554/eLife.45952 (2019).

35 Firnberg, E., Labonte, J. W., Gray, J. J. & Ostermeier, M. A comprehensive, high-resolution map of a gene's fitness landscape. *Mol Biol Evol* **31**, 1581-1592, doi:10.1093/molbev/msu081 (2014).

36 Liu, Y., Yang, Q. & Zhao, F. Synonymous but Not Silent: The Codon Usage Code for Gene Expression and Protein Folding. *Annu Rev Biochem* **90**, 375-401, doi:10.1146/annurev-biochem-071320-112701 (2021).

37 Skarzynski, T., Kim, D. H., Lees, W. J., Walsh, C. T. & Duncan, K. Stereochemical course of enzymatic enolpyruvyl transfer and catalytic conformation of the active site revealed by the crystal structure of the fluorinated analogue of the reaction tetrahedral intermediate bound to the active site of the C115A mutant of MurA. *Biochemistry* **37**, 2572-2577, doi:10.1021/bi9722608 (1998).

38 Skarzynski, T. *et al.* Structure of UDP-N-acetylglucosamine enolpyruvyl transferase, an enzyme essential for the synthesis of bacterial peptidoglycan, complexed with substrate UDP-N-acetylglucosamine and the drug fosfomycin. *Structure* **4**, 1465-1474, doi:10.1016/s0969-2126(96)00153-0 (1996).

39 Eschenburg, S., Kabsch, W., Healy, M. L. & Schonbrunn, E. A new view of the mechanisms of UDP-N-acetylglucosamine enolpyruvyl transferase (MurA) and 5-enolpyruvylshikimate-3-phosphate synthase (AroA) derived from X-ray structures of their tetrahedral reaction intermediate states. *J Biol Chem* **278**, 49215-49222, doi:10.1074/jbc.M309741200 (2003).

40 Eschenburg, S., Priestman, M. & Schonbrunn, E. Evidence that the fosfomycin target Cys115 in UDP-N-acetylglucosamine enolpyruvyl transferase (MurA) is essential for product release. *J Biol Chem* **280**, 3757-3763, doi:10.1074/jbc.M411325200 (2005).

41 Zvelebil, M. J., Barton, G. J., Taylor, W. R. & Sternberg, M. J. Prediction of protein secondary structure and active sites using the alignment of homologous sequences. *J Mol Biol* **195**, 957-961, doi:10.1016/0022-2836(87)90501-8 (1987).

42 Coggins, B. E. *et al.* Structure of the LpxC deacetylase with a bound substrate-analog inhibitor. *Nat Struct Biol* **10**, 645-651, doi:10.1038/nsb948 (2003).

43 Hernick, M. *et al.* UDP-3-O-((R)-3-hydroxymyristoyl)-N-acetylglucosamine deacetylase functions through a general acid-base catalyst pair mechanism. *J Biol Chem* **280**, 16969-16978, doi:10.1074/jbc.M413560200 (2005).

44 Jackman, J. E., Raetz, C. R. & Fierke, C. A. Site-directed mutagenesis of the bacterial metalloamidase UDP-(3-O-acyl)-N-acetylglucosamine deacetylase (LpxC). Identification of the zinc binding site. *Biochemistry* **40**, 514-523, doi:10.1021/bi001872g (2001).

45 Clayton, G. M. *et al.* Structure of the bacterial deacetylase LpxC bound to the nucleotide reaction product reveals mechanisms of oxyanion stabilization and proton transfer. *J Biol Chem* **288**, 34073-34080, doi:10.1074/jbc.M113.513028 (2013).

46 Schonbrunn, E. *et al.* Crystal structure of UDP-N-acetylglucosamine enolpyruvyltransferase, the target of the antibiotic fosfomycin. *Structure* **4**, 1065-1075, doi:10.1016/s0969-2126(96)00113-x (1996).

47 Kimber, M. S. *et al.* The structure of (3R)-hydroxyacyl-acyl carrier protein dehydratase (FabZ) from Pseudomonas aeruginosa. *J Biol Chem* **279**, 52593-52602, doi:10.1074/jbc.M408105200 (2004).

48 Kostrewa, D., Winkler, F. K., Folkers, G., Scapozza, L. & Perozzo, R. The crystal structure of PfFabZ, the unique beta-hydroxyacyl-ACP dehydratase involved in fatty acid biosynthesis of Plasmodium falciparum. *Protein Sci* **14**, 1570-1580, doi:10.1110/ps.051373005 (2005).

49 Dodge, G. J. *et al.* Structural and dynamical rationale for fatty acid unsaturation in Escherichia coli. *Proc Natl Acad Sci U S A* **116**, 6775-6783, doi:10.1073/pnas.1818686116 (2019).

50 Whittington, D. A., Rusche, K. M., Shin, H., Fierke, C. A. & Christianson, D. W. Crystal structure of LpxC, a zinc-dependent deacetylase essential for endotoxin biosynthesis. *Proc Natl Acad Sci U S A* **100**, 8146-8150, doi:10.1073/pnas.1432990100 (2003).

51 Hernick, M. & Fierke, C. A. Catalytic mechanism and molecular recognition of E. coli UDP-3-O-(R-3-hydroxymyristoyl)-N-acetylglucosamine deacetylase probed by mutagenesis. *Biochemistry* **45**, 15240-15248, doi:10.1021/bi061405k (2006).

52 Buetow, L., Dawson, A. & Hunter, W. N. The nucleotide-binding site of Aquifex aeolicus LpxC. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **62**, 1082-1086, doi:10.1107/S1744309106041893 (2006).

53 Gennadios, H. A. & Christianson, D. W. Binding of uridine 5'-diphosphate in the "basic patch" of the zinc deacetylase LpxC and implications for substrate binding. *Biochemistry* **45**, 15216-15223, doi:10.1021/bi0619021 (2006).

54 Kim, D. H. *et al.* Characterization of a Cys115 to Asp substitution in the Escherichia coli cell wall biosynthetic enzyme UDP-GlcNAc enolpyruvyl transferase (MurA) that confers resistance to inactivation by the antibiotic fosfomycin. *Biochemistry* **35**, 4923-4928, doi:10.1021/bi952937w (1996).

55 Zhu, J. Y. *et al.* Functional consequence of covalent reaction of phosphoenolpyruvate with UDP-N-acetylglucosamine 1-carboxyvinyltransferase (MurA). *J Biol Chem* **287**, 12657-12667, doi:10.1074/jbc.M112.342725 (2012).

56 Mihalovits, L. M., Ferenczy, G. G. & Keseru, G. M. Catalytic Mechanism and Covalent Inhibition of UDP-N-Acetylglucosamine Enolpyruvyl Transferase (MurA): Implications to the Design of Novel Antibacterials. *J Chem Inf Model* **59**, 5161-5173, doi:10.1021/acs.jcim.9b00691 (2019).

57 Robinet, J. J. & Gauld, J. W. DFT investigation on the mechanism of the deacetylation reaction catalyzed by LpxC. *J Phys Chem B* **112**, 3462-3469, doi:10.1021/jp075415m (2008).

58 Jackson, S. G., Zhang, F., Chindemi, P., Junop, M. S. & Berti, P. J. Evidence of kinetic control of ligand binding and staged product release in MurA (enolpyruvyl UDP-GlcNAc synthase)-catalyzed reactions. *Biochemistry* **48**, 11715-11723, doi:10.1021/bi901524q (2009).

59 Rozman, K. *et al.* Discovery of new MurA inhibitors using induced-fit simulation and docking. *Bioorg Med Chem Lett* **27**, 944-949, doi:10.1016/j.bmcl.2016.12.082 (2017).

60 Gautam, A., Rishi, P. & Tewari, R. UDP-N-acetylglucosamine enolpyruvyl transferase as a potential target for antibacterial chemotherapy: recent developments. *Appl Microbiol Biotechnol* **92**, 211-225, doi:10.1007/s00253-011-3512-z (2011).

61 Silver, L. L. Fosfomycin: Mechanism and Resistance. *Cold Spring Harb Perspect Med* **7**, doi:10.1101/cshperspect.a025262 (2017).

62 McClerren, A. L. *et al.* A slow, tight-binding inhibitor of the zinc-dependent deacetylase LpxC of lipid A biosynthesis with antibiotic activity comparable to ciprofloxacin. *Biochemistry* **44**, 16574-16583, doi:10.1021/bi0518186 (2005).

63 Barb, A. W., Jiang, L., Raetz, C. R. & Zhou, P. Structure of the deacetylase LpxC bound to the antibiotic CHIR-090: Time-dependent inhibition and specificity in ligand binding. *Proc Natl Acad Sci U S A* **104**, 18433-18438, doi:10.1073/pnas.0709412104 (2007).

64 Clements, J. M. *et al.* Antibacterial activities and characterization of novel inhibitors of LpxC. *Antimicrob Agents Chemother* **46**, 1793-1799, doi:10.1128/AAC.46.6.1793-1799.2002 (2002).

65 Zeng, D. *et al.* Mutants resistant to LpxC inhibitors by rebalancing cellular homeostasis. *J Biol Chem* **288**, 5475-5486, doi:10.1074/jbc.M112.447607 (2013).

66 Tomaras, A. P. *et al.* LpxC inhibitors as new antibacterial agents and tools for studying regulation of lipid A biosynthesis in Gram-negative pathogens. *mBio* **5**, e01551-01514, doi:10.1128/mBio.01551-14 (2014).

67 Chen, H. & Zhou, H. X. Prediction of solvent accessibility and sites of deleterious mutations from protein sequence. *Nucleic Acids Res* **33**, 3193-3199, doi:10.1093/nar/gki633 (2005).

68 Bhasin, M. & Varadarajan, R. Prediction of Function Determining and Buried Residues Through Analysis of Saturation Mutagenesis Datasets. *Front Mol Biosci* **8**, 635425, doi:10.3389/fmolb.2021.635425 (2021).

69 Baba, T. *et al.* Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2**, 2006 0008, doi:10.1038/msb4100050 (2006).

70 Liang, R. & Liu, J. In-frame deletion of Escherichia coli essential genes in complex regulon. *Biotechniques* **44**, 209-210, 212-205, doi:10.2144/000112687 (2008).

71 Goodall, E. C. A. *et al.* The Essential Genome of Escherichia coli K-12. *mBio* **9**, doi:10.1128/mBio.02096-17 (2018).

72 Takahata, S. *et al.* Molecular mechanisms of fosfomycin resistance in clinical isolates of Escherichia coli. *Int J Antimicrob Agents* **35**, 333-337, doi:10.1016/j.ijantimicag.2009.11.011 (2010).

73 Falagas, M. E., Athanasaki, F., Voulgaris, G. L., Triarides, N. A. & Vardakas, K. Z. Resistance to fosfomycin: Mechanisms, Frequency and Clinical Consequences. *Int J Antimicrob Agents* **53**, 22-28, doi:10.1016/j.ijantimicag.2018.09.013 (2019).

74 Van den Bergh, B. *et al.* Frequency of antibiotic application drives rapid evolutionary adaptation of Escherichia coli persistence. *Nat Microbiol* **1**, 16020, doi:10.1038/nmicrobiol.2016.20 (2016).

75 Lee, C. J. *et al.* Structural basis of the promiscuous inhibitor susceptibility of Escherichia coli LpxC. *ACS Chem Biol* **9**, 237-246, doi:10.1021/cb400067g (2014).

76 Dehouck, Y. *et al.* Fast and accurate predictions of protein stability changes upon mutations using statistical potentials and neural networks: PoPMuSiC-2.0. *Bioinformatics* **25**, 2537-2543, doi:10.1093/bioinformatics/btp445 (2009).

77 Swings, T. *et al.* CRISPR-FRT targets shared sites in a knock-out collection for off-the-shelf genome editing. *Nat Commun* **9**, 2231, doi:10.1038/s41467-018-04651-5 (2018).