

ARTICLE

DOI: 10.1038/s41467-017-02680-0

OPEN

# The evolution of substrate discrimination in macrolide antibiotic resistance enzymes

Andrew C. Pawlowski<sup>1</sup>, Peter J. Stogios<sup>2,3</sup>, Kalinka Koteva<sup>1</sup>, Tatiana Skarina<sup>2,3</sup>, Elena Evdokimova<sup>2,3</sup>, Alexei Savchenko<sup>2,3,4</sup> & Gerard D. Wright<sup>1</sup>

The production of antibiotics by microbes in the environment and their use in medicine and agriculture select for existing and emerging resistance. To address this inevitability, prudent development of antibiotic drugs requires careful consideration of resistance evolution. Here, we identify the molecular basis for expanded substrate specificity in MphI, a macrolide kinase (Mph) that does not confer resistance to erythromycin, in contrast to other known Mphs. Using a combination of phylogenetics, drug-resistance phenotypes, and in vitro enzyme assays, we find that MphI and MphK phosphorylate erythromycin poorly resulting in an antibiotic-sensitive phenotype. Using likelihood reconstruction of ancestral sequences and site-saturation combinatorial mutagenesis, supported by Mph crystal structures, we determine that two non-obvious mutations in combination expand the substrate range. This approach should be applicable for studying the functional evolution of any antibiotic resistance enzyme and for evaluating the evolvability of resistance enzymes to new generations of antibiotic scaffolds.

<sup>1</sup>Michael G. DeGroote Institute for Infectious Disease Research and the Department of Biochemistry and Biomedical Sciences, McMaster University, Hamilton, L8S 4L8 ON, Canada. <sup>2</sup>Department of Chemical Engineering and Applied Chemistry, University of Toronto, Toronto, M5S 3E5 ON, Canada. <sup>3</sup>Center for Structural Genomics of Infectious Diseases (CSGID), Calgary, T2N 4N1 AB, Canada. <sup>4</sup>Department of Microbiology, Immunology and Infectious Diseases, University of Calgary, Calgary, T2N 4N1 AB, Canada. Correspondence and requests for materials should be addressed to G.D.W. (email: [wrightge@mcmaster.ca](mailto:wrightge@mcmaster.ca))

Antibiotic resistance is difficult to overcome because its evolution is persistent and multifactorial<sup>1,2</sup>. In addition to genetic mutations that confer drug insensitivity, mobile genetic elements disseminate resistance genes in pathogens, adding to the complexity and unpredictability of the emergence of resistance in bacterial populations<sup>3–5</sup>. Moreover, the prediction and detection of emerging resistance phenotypes resulting from mutations in resistance genes themselves remains a barrier to clinical surveillance and strategic antibiotic use<sup>6–9</sup>, along with antibiotic drug development<sup>10,11</sup>. A focus of the antibiotic pharmaceutical industry is on identifying new antibiotic scaffolds with decreased susceptibility to current and widespread resistance mechanisms<sup>12–15</sup>. Another fruitful avenue in drug discovery is the development of successive generations of existing antibiotic scaffolds that bypass antibiotic-inactivating enzymes and other resistance mechanisms<sup>13,14,16,17</sup>. For these strategies to be productive in the face of increasing rates of resistance, the functional and evolutionary landscape of resistance enzymes must be interrogated to inform new drug discovery efforts.

Macrolides such as azithromycin are among the most successful and highly prescribed antibiotics in the world. They are first-line treatments for community-acquired respiratory tract infections and gonorrhoea<sup>18</sup>, and increasingly used to treat infections caused by multi-drug resistant Enterobacteriaceae<sup>19</sup>. Actinobacteria produce numerous variants of this antibiotic class but clinical implementation is mostly limited to erythromycin and its semi-synthetic derivatives, azithromycin and clarithromycin. The defining features of macrolide antibiotics include a 12-membered to 16-membered macrolactone ring and a dimethylamino sugar linked to the C5 position, which is essential for interacting with its target;<sup>20–22</sup> the large subunit of the bacterial ribosome (Fig. 1). Some macrolides are decorated with additional sugars at C3 (azithromycin, erythromycin), 4'-OH of the dimethylamino sugar (carbomycin, josamycin), or at multiple positions on the macrolactone (megalomicin, spiramycin, tylosin). Telithromycin is a semi-synthetic derivative of erythromycin and a first generation ketolide with a C11–C12 carbamate and an N-linked alkyl-aryl

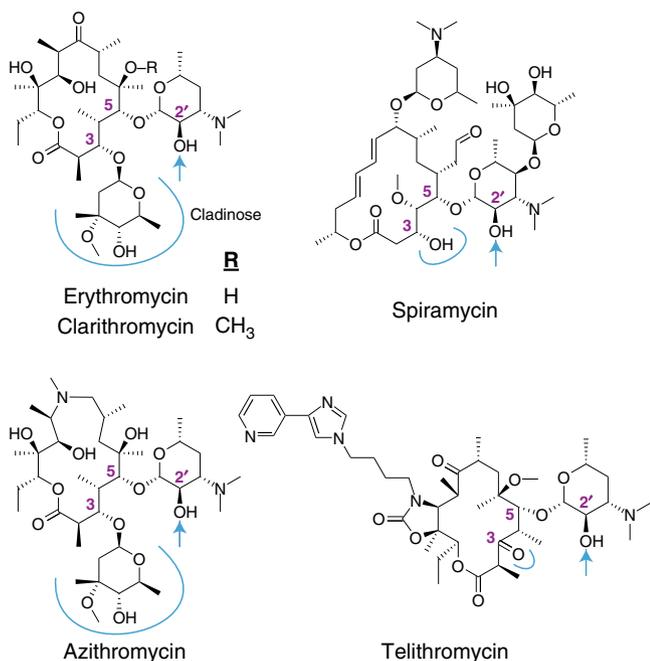
substituent, and a ketone (*ketolide*) at C3 in place of a cladinose. The inspiration for removing the cladinose in telithromycin development was natural ketolides that bypass the induction of macrolide resistance<sup>12</sup>. Nevertheless, bacteria have evolved to overcome to this modification<sup>23–25</sup>. Macrolide resistance is abundant in pathogenic bacteria, and is most often the result of GTP-dependent macrolide kinases (Mph), ribosomal methyltransferases, or efflux pumps<sup>26</sup>. Mph enzymes inactivate macrolides by phosphorylating the 2'-OH of the essential dimethylamino sugar<sup>27,28</sup>, preventing it from binding the ribosome, and providing the chemical rationale for the resistance phenotype.

The Eukaryotic-like kinase superfamily is a structural family related to Eukaryotic protein kinases, and includes antibiotic kinases that modify macrolides (Mph), aminoglycosides<sup>29</sup> (APH), and tuberactinomycins<sup>30</sup> (VPH, CPH). APHs have undergone substantial functional radiation, leading to homologs with different substrate-specificities and regio-specificities<sup>31</sup>. Four Mph homologs (MphA<sup>32</sup>, MphB<sup>33</sup>, MphC<sup>34</sup>, MphE<sup>35</sup>) are mobilized in human pathogens and confer resistance to a wide range of macrolide substrates. In contrast, Mphs from non-pathogenic bacteria are genetically and functionally diverse. We recently described an Mph (MphI) from the environmental bacterium *Paenibacillus* sp. LC231 that does not confer resistance to macrolides with a C3 cladinose<sup>36</sup>. *Paenibacillus* sp. LC231 was isolated from Lechuguilla Cave where it was spatiotemporally separated from surface bacteria for over four million years. MphI shares high sequence identity (94%) to homologs found in related surface *Paenibacillus* sp., indicating the functional divergence of MphI is not recent. The *Bacillus cereus* group have two genetically and functionally distinct Mph enzymes; one that modifies a broad range of macrolides and another that cannot modify macrolides with 16-membered rings<sup>37</sup>. The closest experimentally validated homolog to MphI is MphJ (51% identity), which confers resistance to erythromycin and tylosin, but not to spiramycin or josamycin<sup>38</sup>. Mph enzymes are therefore adaptable resistance enzymes with complex evolutionary paths.

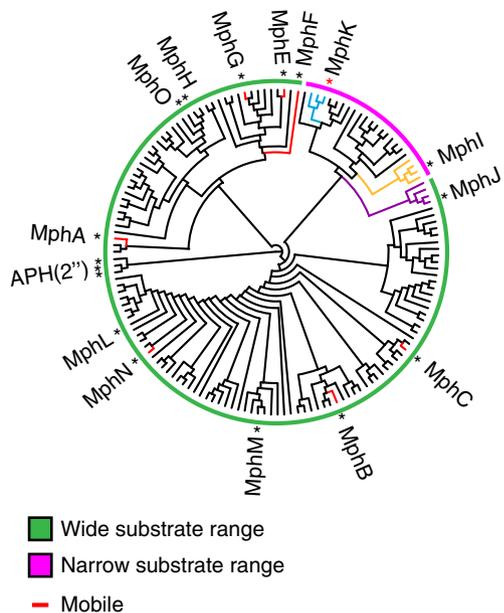
In this study, we show that narrow range Mphs are common in Bacillales, and their phosphorylation rates of cladinose containing macrolides is insufficient to confer resistance. We reconstruct the evolutionary path to Mph functional divergence using ancestral sequence reconstruction and identify four candidate residues responsible for substrate specificity. Applying phylogenetics, structural biology, protein engineering, and in vitro enzyme assays to the Mph family, we evolve increased substrate range in MphI, which results from multiple non-obvious mutations that, in tandem, increase the catalytic rate towards cladinose containing macrolides. This work demonstrates that while the antibiotic kinases have evolved several distinct functions, there exists significant barriers to acquiring a very related, but non-intrinsic function.

## Results

**Multiple independent macrolide kinase mobilization events.** Of the 13 Mph homologs that have been experimentally validated for macrolide kinase activity<sup>27,32–37,39–41</sup>, 7 are captured by mobile genetic elements. This implies that Mphs are at a high risk for mobilization (Supplementary Fig. 1, Supplementary Table 1). These Mphs share low sequence identity and it is therefore likely that each homolog was mobilized independently; if they originated from a single mobilized Mph, then they should cluster together in phylogenetic analysis. However, only MphE, MphF, and MphG are phylogenetically clustered, while MphA, MphB, MphC, and MphN are isolated from other known Mphs (Fig. 2; Supplementary Fig. 2). These results show that Mph genes may



**Fig. 1** Structure of macrolide antibiotics. The blue arrow indicates the site of phosphorylation by Mph enzymes. The C3 position is highlighted with a blue line

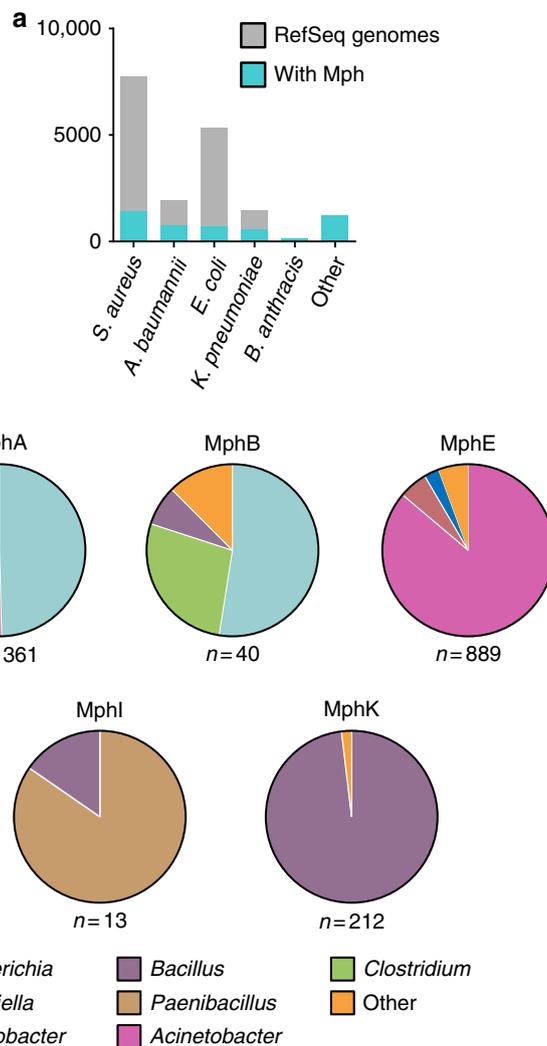


**Fig. 2** Phylogenetic reconstruction of macrolide phosphotransferases. Colored bars represent predicted substrate specificity based on phylogenetic clustering of known Mphs and their homologs. Asterisks represent experimentally validated Mphs, and the red asterisk represents MphK, which was identified from this analysis as a Mph that does not confer C3 cladinose macrolide resistance. Colored portions of the phylogenetic tree highlight clusters of interest. Mphs captured on mobile genetic elements are colored red. The tree is presented as a cladogram, and the root drawn with GTP-dependent APH(2'') enzymes as outgroups. A cladogram with complete labels and bootstrap values can be found in Supplementary Figure 2

have been captured by mobile genetic elements at least five independent times.

The taxonomic distribution Mphs remains unclear despite several studies detailing their phenotypic, biochemical, and structural characterizations. We expanded the known genetic and taxonomic diversity of Mph enzymes by surveying bacterial strains with genomes deposited in the RefSeq database. Mphs were detected in 4732 genomes across 71 genera, and were predominantly found in *Staphylococcus*, *Acinetobacter*, *Escherichia*, *Bacillus*, and *Klebsiella* (Supplementary Fig. 3). Gram-negative pathogens, such as *A. baumannii* (38% of all *A. baumannii* strains), *K. pneumoniae* (38%), and *E. coli* (13%) are significantly enriched for Mphs (Fig. 3a). Ancestral *A. baumannii*, *K. pneumoniae*, and *E. coli* strains did not harbor an Mph homolog, and therefore it is very likely that these were acquired on mobile elements. Analysis of the taxonomic distribution of each Mph homolog reveals that MphA, MphB, and MphE are widespread in Gram-negative bacteria, and that the MphI homologs are limited to *Bacillus* and *Paenibacillus* and have not been mobilized (Fig. 3b; Supplementary Fig. 4). Our analysis demonstrates that Mphs are widely distributed across bacterial taxa, and the mobilized homologs are accumulating in Gram-negative pathogens.

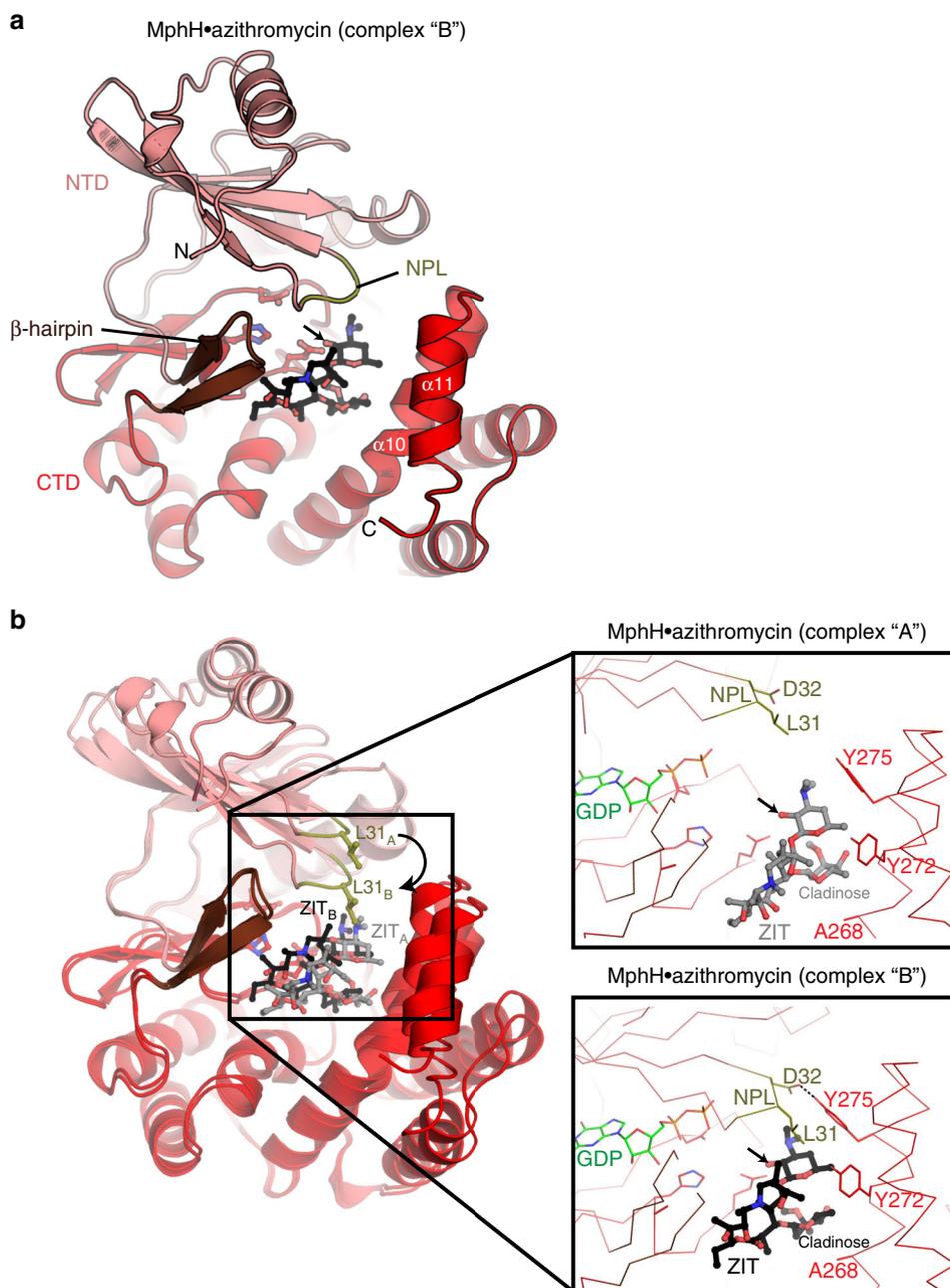
**Substrate specificity correlates with genetic diversity.** MphI is the only known Mph that does not confer C3 cladinose macrolide resistance<sup>36</sup>. If MphI lost the ability to phosphorylate macrolides with a C3 cladinose, then we would expect it to share a common ancestor with Mphs that have wider substrate specificity, and therefore not positioned near the root of a phylogenetic reconstruction. The closest known homolog to MphI, MphJ (51%



**Fig. 3** Taxonomic distribution of Mphs. **a** The top five bacterial species, in the RefSeq genome database, with Mphs by abundance, and presented in a stacked bar graph. **b** Taxonomic distribution of MphA, MphB, MphE, MphI, and MphK by genera

sequence identity), can phosphorylate C3 cladinose macrolides<sup>38</sup> and is an outgroup to the MphI clade (Fig. 2; Supplementary Fig. 2). The most parsimonious explanation is that MphI lost the ability to phosphorylate C3 cladinose macrolides. Adjacent to MphI is another clade with uncharacterized enzymes from *Bacillus* that we reasoned may show MphI-like specificity if phylogeny predicts substrate specificity. We heterologously expressed one of the members of this clade in *E. coli* TOP10 (*mphK/ycbJ*, 54% identical to *mphI*) from *Bacillus subtilis* 168 and verified that it confers resistance to macrolide antibiotics (Supplementary Table 2). Moreover, like *mphI*, *mphK* does not confer resistance to C3 cladinose macrolides, indicating that Mph substrate specificity correlates with phylogeny.

To further investigate Mph substrate specificity, we performed *in vitro* enzyme assays with purified MphB, MphI, and MphK. Azithromycin and telithromycin are semi-synthetic derivatives of erythromycin, with telithromycin lacking a C3 cladinose (Fig. 1). MphI and MphK both phosphorylate telithromycin but also phosphorylate azithromycin (Supplementary Fig. 5, Supplementary Table 3), suggesting that resistance phenotype may not correlate with biochemical analysis of drug modification. To confirm this result, we used tandem mass spectrometry to identify the site of phosphorylation on erythromycin, a macrolide

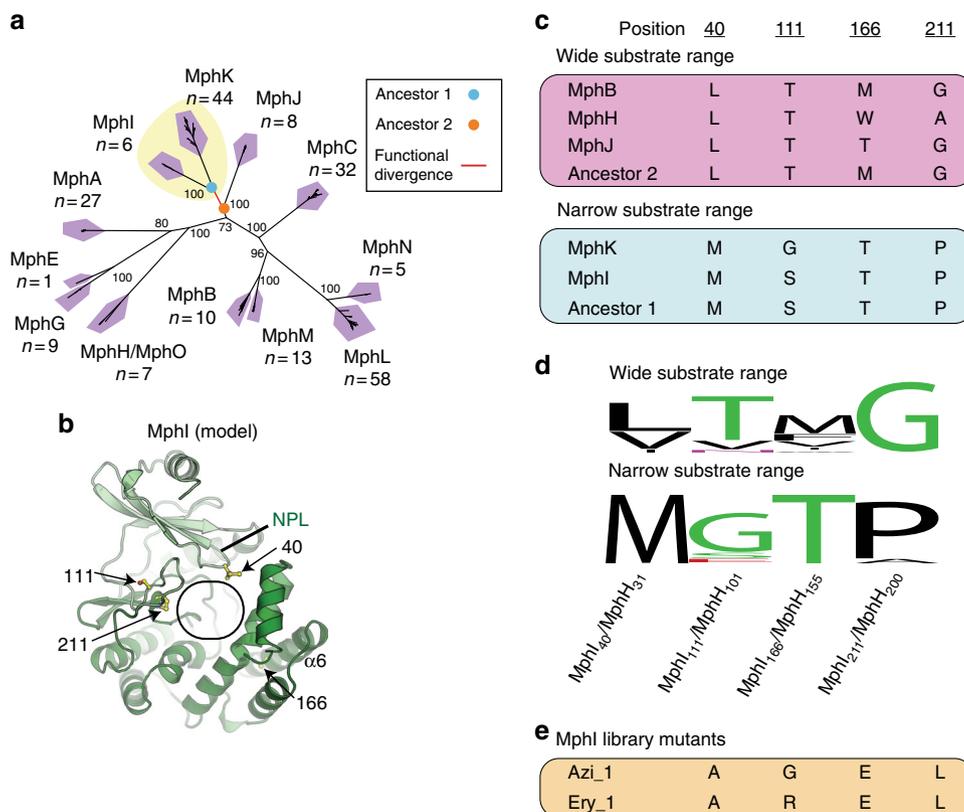


**Fig. 4** Crystal structures of MphH. **a** The MphH•azithromycin (complex "B" of the two complexes in the asymmetric unit of this crystal) is shown in cartoon, with NTD, CTD, NPL, and β-hairpin regions colored differently. Azithromycin (ZIT) is shown in black ball-and-stick representation. α10 and α11 which make up one face of the macrolide-binding site are labeled. Catalytic residues Glu196, His201, and Asp214 are shown in ball-and-stick. Arrow indicates 2'-OH site of phosphorylation on azithromycin. **b** Overlay of MphH•azithromycin complexes "A" and "B", showing conformational differences in NPL (particularly Leu31), position of azithromycin and α11. Inset shows zoom of macrolide-binding sites of each complex, showing key residues that contact the antibiotic compounds. Arrows indicate 2'-OH site of phosphorylation on azithromycin. GDP is shown from crystal structure of MphH•GDP complex

antibiotic with a similar structure to azithromycin and with well-established mass fragmentation patterns<sup>42</sup>. Both MphI and MphK phosphorylate the 2'-OH of erythromycin (Supplementary Fig. 6). Using steady-state kinetics, we demonstrated that MphI and MphK can indeed use C3 cladinose macrolides as substrates, but not efficiently ( $k_{\text{cat}}/K_m$   $10^1$ – $10^2$ ) (Supplementary Tables 4 and 5). The difference in catalytic constant ( $k_{\text{cat}}/K_m$ ) values between C3 cladinose macrolides and descladinose macrolides is about  $10^3$  and  $10^1$ – $10^3$  respectively for MphI and MphK. This poor biochemical specificity translates to the lack of resistance phenotype to these antibiotics when the genes are expressed in *E. coli* TOP10. To confirm whether this low activity against C3

cladinose macrolides impacts resistance phenotypes, we expressed *mphI* and *mphK* in an antibiotic hyper-susceptible strain of *E. coli* (*E. coli* BW25113  $\Delta\text{bamB}\Delta\text{tolC}$ <sup>43</sup>). Both *mph* genes confer only a modest decrease in sensitivity to these antibiotics due to impaired efflux and outer membrane composition in this strain, verifying that the enzymatic activity observed in vitro has minimal impact on resistance phenotype (Supplementary Table 6).

MphB is another homolog also reported to have a narrow substrate range, modifying erythromycin but not azithromycin<sup>44</sup>. In contrast to previous reports, we found that MphB confers resistance to all macrolides tested (Supplementary Tables 2 and 3), and inactivates both telithromycin and azithromycin



**Fig. 5** Ancestral sequence reconstruction of narrow-substrate and wide-substrate range Mphs for identifying possible determinants of substrate specificity. **a** Closely related homologs (>80% identity) to known Mphs were used to construct a phylogenetic tree. The tree is presented as an unrooted radial phylogram. The red tree branch represents the point of functional divergence between Mphs that can phosphorylate C3 cladinose macrolides, and those that cannot. The colored circles represent ancestral nodes with reconstructed sequences used to compare important residues. Bootstrap values are indicated at branch points. Numbers underneath Mph names represent the number of unique sequences in each cluster. The yellow background represents clades that do not efficiently phosphorylate macrolides with C3 cladinose. **b** Model of MphI based on crystal structure of MphH. Oval represents macrolide-binding pocket as determined from MphH•azithromycin crystal structures. Residues identified by ancestral sequence reconstruction of MphI–MphK divergence are colored in yellow. **c** Amino acid residues at select positions of narrow-spectrum and wide-spectrum Mphs, and their reconstructed ancestors. **d** Conservation of residues across the two Mph functional subtypes. Sequence conservation is displayed as a sequence logo using WebLogo. Wide substrate range includes sequences in all clades except MphI and MphK; narrow substrate range includes sequences in only MphI and MphK clades. **e** Amino acid residues at select positions of MphI library mutants with expanded substrate range

**Table 1** Antibiotic susceptibility of *mphI* and two library mutants with expanded resistance to C3 cladinose macrolides expressed in *E. coli* BW25113  $\Delta$ *bamB* $\Delta$ *tolC*

Antibiotic	Empty vector	<i>mphI</i>	<i>azi_1.1</i> <sup>a</sup>	<i>ery_1</i> <sup>b,c</sup>
Erythromycin	0.5	1	16	16
Clarithromycin	0.5	1	4	4
Azithromycin	0.063	0.25	2–4	2–4
Telithromycin	0.13	>16	>16	>16
Spiramycin	2	256	256	256
Tylosin	2	1024	1024	1024
Josamycin	2–4	>64	>64	>64
Kanamycin	1	1	1	1

<sup>a</sup>MphI<sub>M40A/S111R/T166E/P211L</sub>

<sup>b</sup>MphI<sub>M40A/S111G/T166E/P211L</sub>

<sup>c</sup>MphI mutant Ery\_1 was found to contain mutations in the promoter region, potentially altering expression and resistance phenotype. Ery\_1 was therefore subcloned into the wild-type pGDP4 empty vector. MICs in this table are directly comparable.

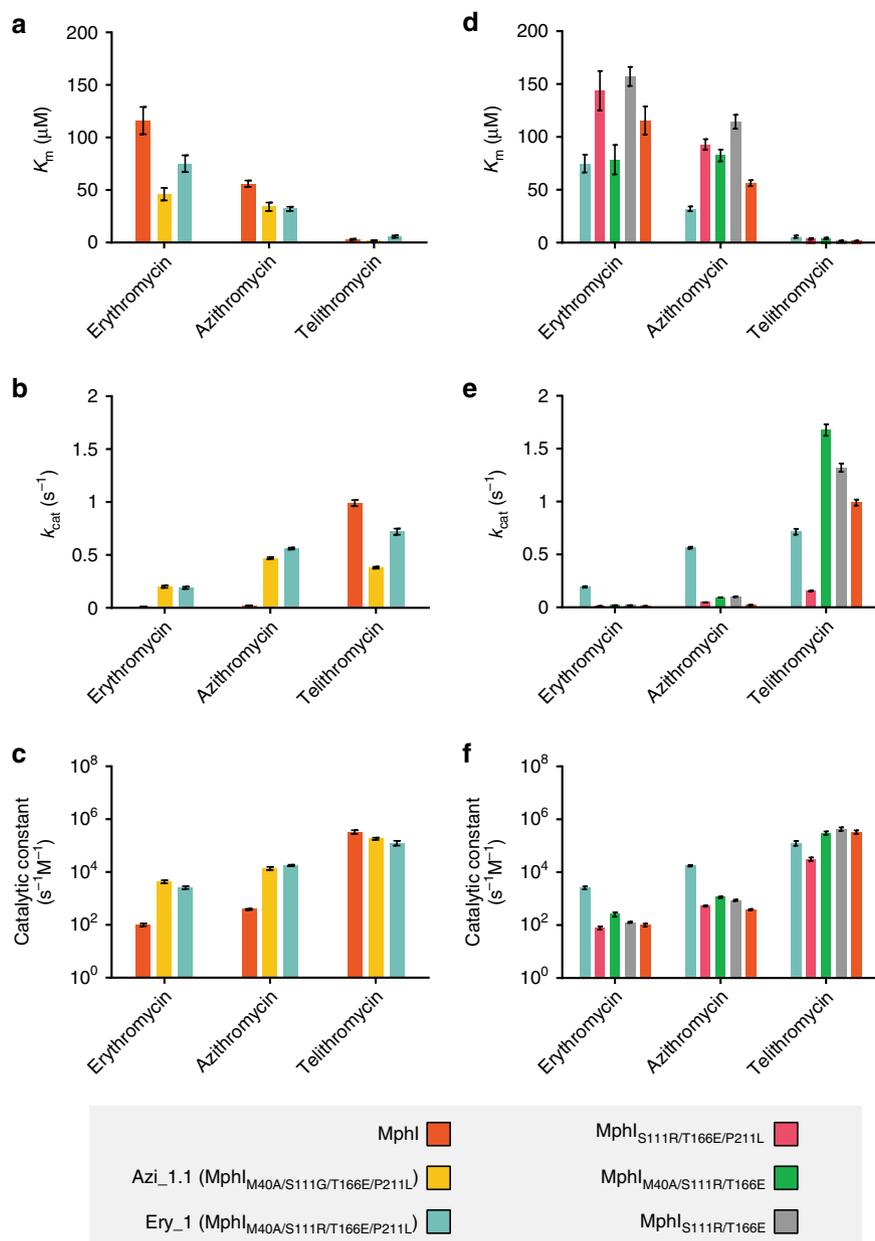
(Supplementary Table 6, Supplementary Fig. 5). Using tandem mass spectrometry, we also confirmed that MphB phosphorylates the desosamine 2'-OH of erythromycin (Supplementary Fig. 6). Moreover, MphB modifies macrolides with efficiencies

comparable to other Mphs ( $k_{cat}/K_m$   $10^4$ – $10^6$ ), and the  $K_m$  for C3 cladinose macrolides is 19–69× lower than MphI (Supplementary Table 8). In aggregate, these results show that C3 cladinose macrolides are good substrates for MphB and other Mphs including MphA<sup>32,44</sup> and MphH<sup>27</sup> (Supplementary Table 2), but are poor substrates for MphI and MphK.

**Structural analysis of macrolide recognition by Mph enzymes.**

To better understand the molecular specificity of Mph enzymes against macrolides, we determined the 3-D structure of MphB as the apoenzyme, and three forms of MphH: (1) apoenzyme, (2) in complex with GDP, and (3) in complex with azithromycin (Supplementary Table 4 details of X-ray crystallographic statistics). While preparing this manuscript, crystal structures of MphA and MphB were reported in complex with various macrolides, which we have included in our analysis of substrate specificity<sup>45</sup>. We also attempted to crystallize MphI and MphK as apoenzymes, with several GTP analogs, and with spiramycin, tylosin, and telithromycin, but we were unsuccessful.

As expected, the structures of MphB (from our study) and MphH showed that the enzymes adopt the bi-lobe kinase fold reminiscent of APH(2'') enzyme crystal structures, with an N-terminal domain dominated by  $\beta$  sheets and a C-terminal domain



**Fig. 6** Kinetic characterization of MphI and mutants with improved activity towards C3 cladinose macrolides. Steady-state enzyme kinetics for MphI library mutants (**a**, **b**, **c**) and intermediate Ery\_1 mutants (**d**, **e**, **f**). **a**  $K_m$ , **b**  $k_{cat}$ , and **c**  $k_{cat}/K_m$  values for MphI and library mutants are presented in clustered bar graph format. **d**  $K_m$ , **e**  $k_{cat}$ , and **f**  $k_{cat}/K_m$  values for intermediate Ery\_1 mutants are also presented. Error bars represent the standard error of the mean (s.e.m.). The y-axis for **c** and **f** are presented in  $\log_{10}$  format

comprised mainly of  $\alpha$ -helices, linked via a  $\beta$ -hairpin insertion (residues 100–110) (Fig. 4a; Supplementary Fig. 7a).

The asymmetric unit of the MphH azithromycin crystal contains two copies of the complex (“complex A” and “complex B”), each showing subtle but important conformational differences (Fig. 4b). In both complexes, the azithromycin molecule is bound in a deep, electronegative cleft with one face formed by the end of C-terminal domain (i.e., helices  $\alpha$ 10 and  $\alpha$ 11), with the other face formed by the apposition of the  $\beta$  hairpin, the highly conserved region of the C-terminal domain containing the catalytic residues, and the nucleotide-positioning loop (NPL; residues 29–33) of the N-terminal domain (Fig. 4a). However, it was evident that the “complex B” encloses the bound macrolide more intimately and more appropriately for catalysis (Fig. 4b) and this complex dominates our analysis.

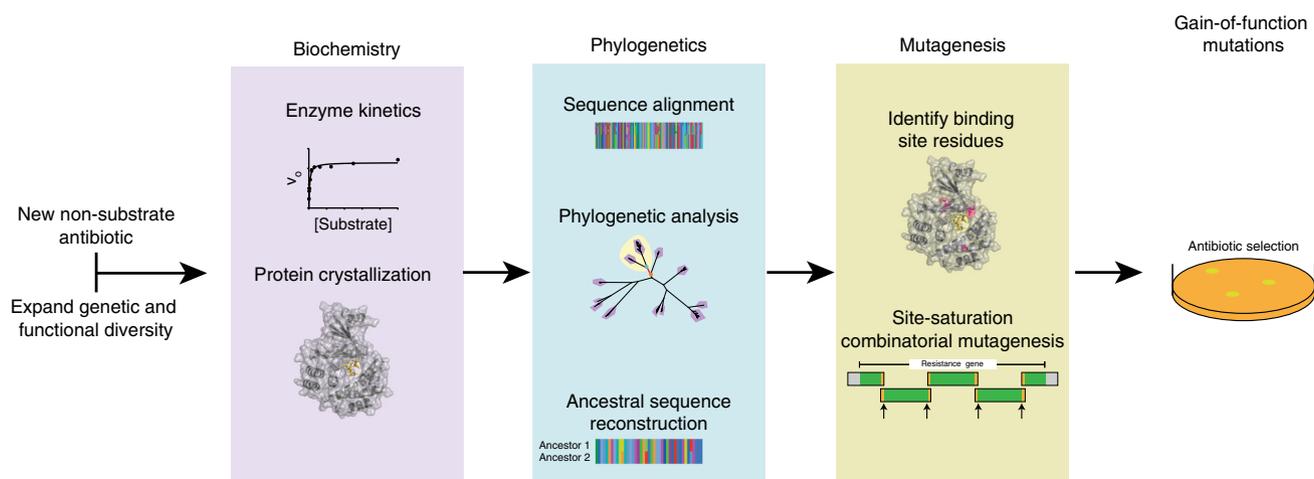
We compared the overall structure of MphH with MphA (5IG1<sup>45</sup>) and MphB (from ref. <sup>45</sup>, 5IGV) to identify possible conformational changes associated with macrolide binding. While the MphA, MphB, and MphH structures are very similar, we observed structural distinctions in all three enzymes that localized to the (a) NPL, (b)  $\beta$ -hairpin insert, and (c) multiple residues in the macrolide-binding cleft (Supplementary Fig. 7a). As we observed in MphH, the NPL, and  $\beta$ -hairpin regions of both MphA and MphB adopt different conformations with macrolide binding. Consistent with the flexibility of MphH in our crystal structure, these observations highlight that conformational changes in these key regions of Mph enzymes play a role in macrolide substrate binding.

We compared the positioning of bound macrolides among the Mph enzymes to gain better insight into recognition of the C3 cladinose ring. The position of azithromycin in the macrolide-

**Table 2 Antibiotic susceptibility of intermediate mutants between MphI and MphI<sub>M40A/S111R/T166E/P211L</sub> expressed in *E. coli* BW25113  $\Delta$ bamB $\Delta$ tolC**

Antibiotic	ery <sub>1</sub> (MphI <sub>M40A/S111R/T166E/P211L</sub> )	mphI <sub>S111R/T166E/P211L</sub>	mphI <sub>M40A/T166E/P211L</sub>	mphI <sub>M40A/S111R/P211L</sub>	mphI <sub>M40A/S111R/T166E</sub>	mphI <sub>S111R/T166E</sub>	mphI <sub>M40A/P211L</sub>	mphI <sup>a</sup>
Erythromycin	32	4	32	32	4	2–4	32	2–4
Clarithromycin	8	1	8	8	2	1	16	1
Azithromycin	8	1	8	8	2	0.5	4–8	0.5
Telithromycin	>16	>16	>16	>16	>16	>16	32	>16
Spiramycin	256	256	256	256	256	256	128–256	>256
Tylosin	1024	1024	1024	1024	1024	>1024	>1024	>1024
Josamycin	>64	>64	>64	>64	>64	>64	>64	>64
Kanamycin	1	1	1	1	1	1–2	1	1

<sup>a</sup>MphI mutant Ery<sub>1</sub> was found to contain mutations in the promoter region, potentially altering expression and resistance phenotype. MphI was subcloned into the Ery<sub>1</sub> vector and therefore MICs in this table are directly comparable.



**Fig. 7** Workflow for assessing the evolvability of resistance enzymes to new antibiotics using biochemistry, phylogenetics, and mutagenesis. First, expand the known functional diversity of an enzyme class by identifying homologs that include the new antibiotic in its substrate range. Characterize the enzymes using steady-state kinetics and solve the protein structure of at least one homolog. Use the biochemical experiments to correlate functional diversity with phylogenetic diversity for identifying the last common ancestor of each functional subtype (e.g., ancestor of narrow range Mphs and wide range Mphs, respectively). The difference in ancestral protein sequences are residues possibly involved in substrate specificity, which should be prioritized based on proximity to the substrate-binding site in the protein structure. Target these residues for mutagenesis, preferably by saturation mutagenesis at each position, in combination (site-saturation combinatorial mutagenesis). Identify residues expanding substrate specificity by sequencing technologies that generate reads covering the whole gene (e.g., Sanger, Nanopore, PacBio)

binding pockets of MphA, MphB, and MphH are similar in that the 2' rings are oriented appropriately for catalysis, and the macrolactone ring interacts with the  $\beta$ -hairpin regions. However, the azithromycin molecule bound to MphH is positioned  $\sim 3$  Å closer to the nucleotide binding region as compared to its position in either the MphA or MphB active sites (Supplementary Fig. 7a). Consequently, the C3 cladinose ring of azithromycin occupies deeper regions of the binding clefts of MphA or MphB. This analysis reveals that the C3 cladinose ring is positioned differently upon binding to individual Mph enzymes.

**Substrate selectivity is an emergent property in Mphs.** Since structural analysis demonstrated that the C3 cladinose occupies slightly different positions in the MphA, MphB, and MphH active sites, we speculated that MphI and MphK may have a large residue facing toward the C3 cladinose that would sterically hinder binding of these macrolides. We therefore compared the sequence composition of the six residues that approach the cladinose in the Mph structures (i.e., in the MphH azithromycin co-

structure—Glu196, Phe229, Phe265, Ala268, Gly271, Tyr272) across a multiple sequence alignment of all known Mphs and their close homologs (>80% identity). We found no significant correlations between residue type and antibiotic substrate specificity (Fig. 4b; Supplementary Fig. 7b, Supplementary Fig. 8). Ultimately, this analysis of the Mph azithromycin molecular contacts shows that the physiochemical properties of residues in proximity with the C3 cladinose ring are not conserved even in wide substrate range Mphs. Therefore, C3 cladinose macrolide specificity does not appear to be directed by nearby residues.

Since Mph substrate specificity is a phylogenetic trait, we reasoned that reconstructing the evolutionary path of functional divergence could identify the molecular determinants influencing C3 cladinose phosphorylation<sup>46–50</sup>. To this end, we reconstructed ancestral amino acid sequences<sup>51</sup> using extant enzymes based on a multiple sequence alignment and phylogeny of related Mph homologs (>80% identity) (Fig. 5a). Both Mph phylogenies (Figs. 2 and 5a) independently indicate MphI and MphK lost activity towards C3 cladinose macrolides, increasing our confidence in the branching order of Mph sequences. Therefore,

the equivalent residues different between Ancestor 1 (MphI and MphK) and Ancestor 2 (broad substrate range) may be determinants of substrate specificity<sup>46</sup>. We mapped these positions onto the MphH•azithromycin crystal structure and found four common sites near the macrolide-binding pocket, thus raising the possibility of their involvement in increased activity towards C3 cladinose macrolides (Fig. 5b–d). According to the MphH•azithromycin structure, one of these sites, MphH<sub>L31</sub> (MphI<sub>M40</sub>) interacts with azithromycin; MphH<sub>L31</sub> faces toward the desosamine and C6–C7 of the macrolactone ring. MphH<sub>A200</sub> (MphI<sub>P211</sub>) is on the short helix  $\alpha$ 8 that approaches MphH<sub>Y198</sub>, which forms a key hydrophobic packing interaction with the macrolactone ring close to the methyl on C13 and the ethyl group on C14. MphH<sub>W155</sub> (MphI<sub>T166</sub>) is in  $\alpha$ 6 that lines the macrolide-binding pocket and lies within 6 Å of the cladinose. The final site, MphH<sub>T101</sub> (MphI<sub>S111</sub>) is more peripheral to the active site and is in the  $\beta$ 1 hairpin but faces away from the macrolide-binding pocket.

We independently mutated each of these positions in MphI to the corresponding residue in the ancestral sequence; none of these variants conferred resistance to C3 cladinose macrolides (Supplementary Table 9). Since engineering new substrate specificity (gain of function) into enzymes is often complex and mutations may require a specific constellation of multiple residues to result in a functional change<sup>52,53</sup>, we created an MphI library with mutations at all four positions, in combination, using combinatorial site-saturation mutagenesis. This enabled effective sampling all 160,000 possible amino acid substitutions at each of the four positions identified in ancestral reconstruction, in combination (1.05 million codon combinations). Using the powerful positive selection for antibiotic resistance to screen this mutant library, we were able to identify gain of function azithromycin or erythromycin resistance. We identified three azithromycin-resistant variants (Azi\_1.1, Azi\_1.2, Azi\_1.3) with identical amino acid sequences, but different codon combinations, and one erythromycin-resistant variant (Ery\_1) (Fig. 5e, Table 1; Supplementary Table 10). All variants with an expanded substrate range were identical at three residue positions: Ala40, Glu166, and Leu211 (mutants of MphI Met40, Thr166, Pro211). This result reveals a very limited number of routes to increase substrate range.

**MphI substrate selectivity is correlated with catalytic rate.** We purified Azi\_1.1 and Ery\_1, and verified azithromycin and telithromycin phosphorylation in an in vitro enzyme assay (Supplementary Fig. 5). Mutations in other antibiotic modifying enzymes are known to alter regiospecificity<sup>54</sup>, and the cladinose hydroxyl of azithromycin is a possible phosphate acceptor. We unambiguously identified the desosamine 2'-OH as the phosphorylation site of Azi\_1.1 and Ery\_1 using tandem mass spectrometry (Supplementary Fig. 6). Using steady-state kinetics, we demonstrated that the  $k_{\text{cat}}/K_m$  of Azi\_1.1 and Ery\_1 for C3 cladinose macrolides increased 35–86 fold and 26–46 fold, respectively (Fig. 6a–c; Supplementary Tables 11 and 12). Rather than a significant decrease in  $K_m$  for C3 cladinose macrolides, the increase in activity towards these substrates was primarily driven by an increase in  $k_{\text{cat}}$ . The  $K_m$  decreased by about half, while the  $k_{\text{cat}}$  value increases 16×(erythromycin), 22–25×(azithromycin), and 48–51×(roxithromycin). Our results suggest that the four mutations in Azi\_1.1 and Ery\_1 increase the rate of product formation, and do not significantly increase catalytically productive C3 cladinose macrolide-binding affinity.

**Expanded substrate specificity requires binary mutations.** Ancestral sequence reconstruction successfully identified residues

implicated in substrate specificity, but the molecular roles of these residues remains unclear. To further explore these, we probed each position's contribution by systematically reversing each Ery\_1 (MphI<sub>M40A/S111R/T166E/P211L</sub>) mutation to wild-type sequence. MphI<sub>M40A/T166E/P211L</sub> and MphI<sub>M40A/S111R/P211L</sub> conferred resistance to C3 cladinose macrolides and retained the same resistance phenotype as Ery\_1 (Table 2). Similarly, MphI<sub>S111R/T166E</sub> had the same resistance phenotype as MphI. These results demonstrate that positions 111 and 166 are not involved in expanding substrate specificity, consistent with their location more distal to the macrolide-binding site. In contrast, MphI<sub>S111R/T166E/P211L</sub> and MphI<sub>M40A/S111R/T166E</sub> lost the ability to confer C3 cladinose macrolide resistance, indicating that a binary combination of both mutations is required. MphI<sub>M40A/P211L</sub> had the same resistance phenotype as Ery\_1 (MphI<sub>M40A/S111R/T166E/P211L</sub>), further supporting the role of both mutations. We confirmed these results using steady-state kinetics of MphI<sub>S111R/T166E/P211L</sub>, MphI<sub>M40A/S111R/T166E</sub>, and MphI<sub>S111R/T166E</sub> (Fig. 6d–f; Supplementary Tables 13–15). Each MphI variant had  $k_{\text{cat}}/K_m$  values similar to wild-type MphI. These results indicate that together, residues Met40 and Pro211, play key roles in macrolide substrate specificity of Mphs. Met40 and Pro211 are located in the nucleotide-positioning loop and in the highly structurally conserved region near GDP-coordinating/catalytic residues, respectively.

## Discussion

New generations of macrolide antibiotics are being developed to bypass resistance<sup>17,55</sup>, which necessitates understanding the evolutionary landscape of Mphs. In this study, we find that two non-obvious mutations in tandem are required for expanding the resistance phenotype of MphI to include C3 cladinose macrolides by increasing the catalytic rate. These results are consistent with a plastic MphI active site that accommodates a wide selection of macrolides, but selectively phosphorylates macrolides without a cladinose. Narrow range Mphs such as MphI lost the ability to modify C3 cladinose macrolides, likely because of lack of selection by diverse macrolide structures in the local environment accompanied by neutral drift.

Amino acid residues positioned near, but not in, substrate-binding sites can impact function by modulating protein dynamics<sup>56,57</sup>. For example, combinations of mutations located outside the  $\beta$ -lactam binding pocket of  $\beta$ -lactamases expand substrate range by increasing conformational flexibility<sup>50,56–58</sup>. Introduction of each mutation individually does not expand the substrate range of  $\beta$ -lactamases<sup>57</sup>. In our study, M40A and P211L are located outside the MphI macrolide-binding pocket and both mutations are required to confer resistance to C3 cladinose macrolides. Altogether, the binary M40A and P211L mutations may increase the conformational flexibility of the macrolide-binding pocket, increasing the phosphorylation rate of C3 cladinose macrolides.

Understanding the evolution of resistance enzymes is of considerable interest because of their wide functional and genetic diversity, and their biological importance in medicine and microbial ecology<sup>2,59,60</sup>. Combining random mutagenesis strategies with antibiotic selection is powerful for interrogating the genetic-phenotypic continuum of resistance enzymes. Indeed, many of the strategies for protein engineering and synthetic biology were first developed using resistance enzymes<sup>61–65</sup> (e.g., TEM  $\beta$ -lactamase), and provide a strong foundation for biochemical investigations of resistance. Our approach builds on previous strategies by identifying target positions using phylogenetic and evolutionary principals, and sampling random genetic diversity at each of the four positions simultaneously. The

strategy we report here offers a single step for efficient site-saturation combinatorial mutagenesis of multiple positions using Gibson Assembly and productive sampling of the mutant library (>1 million variants) using positive antibiotic selection. Our approach successfully identified candidate positions involved in Mph substrate selectivity that were unidentifiable from Mph•macrolide co-structures reported here or by Fong et al.<sup>45</sup>. Despite the macrolactone ring and dimethylamino sugar shared by all macrolides, the M40A and P211L mutations in MphI specifically increase phosphorylation rate of those with a C3 cladinose without affecting catalysis of descladinose macrolides. In the MphH•azithromycin co-structure, the equivalent MphH positions Leu31 and Ala200 do not interact with the C3 cladinose. Our study illuminates the utility of evolutionary guided mutagenesis for studying non-obvious molecular determinants of substrate selectivity.

Ancestral sequence reconstruction has proven to be a valuable tool in drug discovery<sup>49</sup>, protein engineering<sup>46,48</sup>, and molecular evolution studies<sup>47,50,66</sup>. We believe our approach will be widely applicable for uncovering the molecular basis of functional divergence in antibiotic resistance enzymes. New generations of antibiotic scaffolds are being designed to bypass resistance mechanisms mobilized in pathogens<sup>67</sup>. However, if a homolog in environmental bacteria confers resistance to the new antibiotics, the potential for evolving resistance is high. Use of ancestral sequence reconstruction combined with structural and biochemical studies, as performed here, will be useful for predicting the possibility and feasibility of expanded resistance phenotypes to semi-synthetic antibiotics (Fig. 7).

Recent platforms that synthesize diverse macrolides will enable the discovery of newer generations of resistance-proof antibiotics<sup>17,55</sup>. Our study suggest that macrolide binding may be a precursor to phenotypic resistance in Mphs. Therefore, next generation macrolide antibiotics that bypass Mph-mediated resistance should also be investigated for active site binding, and whether homologs from environmental bacteria confer resistance.

## Methods

**Antibiotics and reagents.** All buffers and salts were purchased from BioShop (Burlington, ON, Canada) unless otherwise specified. Macrolide antibiotics, GTP, pyruvate kinase: lactate dehydrogenase (PK:LDH), phosphoenolpyruvate, and NADH were purchased from Sigma-Aldrich (Oakville, ON, Canada). Molecular biology kits and organic solvents were purchased from Fisher Scientific (Ottawa, ON, Canada). Telithromycin was purified from the drug formulation Ketek (400 mg, Sanofi-Aventis US)<sup>36</sup>.

**Phylogenetic reconstruction of macrolide kinases.** Diverse Mph sequences were identified with BLASTp<sup>68</sup> by querying the Mph sequences in Supplementary Table 1 with an e-value cutoff of  $1e^{-30}$ , and the following flags; -max\_target\_seqs 500, and -entrez\_query 'NOT partial'. Note that Mph enzymes were renamed for consistency with established nomenclature<sup>26</sup> and to avoid multiple distinct sequences (<80% amino acid identity) with the same name. A hit was defined as having at least 50% sequence identity and an alignment length >80% of the query length. Sequences were clustered at 80% identity using the cluster\_smallmem algorithm of uclust<sup>69</sup>, and aligned with MAFFT (L-INS-i method)<sup>70</sup>. APH(2'') enzymes are homologs of Mphs<sup>71</sup>, and therefore APH(2'')-IIa, APH(2'')-IIIa, and APH(2'')-Ie were used as outgroups for rooting the tree. The alignment was then weighted by alignment confidence using the transitive consistency score (TCS) function of T-coffee<sup>72</sup>. RAxML was used to generate a maximum likelihood phylogenetic tree with the GAMMA model of rate heterogeneity and LG empirical base frequencies (-m PROTGAMMAAUTO) and using 100 rapid bootstraps<sup>73</sup>.

**Mph cloning and antibiotic susceptibility testing.** *mphB* was codon optimized, synthesized as a gBlock (IDT) and cloned into pET28a with an N-terminal histidine tag. The codon optimized *mphB* sequence is in Supplementary Table 17. *mphI* was previously cloned into pET28a<sup>36</sup>. pGDP4 is a derivative of pET28a with the multiple-cloning site under the control of the constitutive lac promoter<sup>43</sup>. *mphI* was subcloned into pGDP4 using the XbaI and XhoI restriction sites. *E. coli* was routinely cultured on LB-Lennox (Bioshop) at 37 °C overnight, and either 100 µg mL<sup>-1</sup> ampicillin or 50 µg mL<sup>-1</sup> kanamycin was used for plasmid selection.

*E. coli* TOP10 (Invitrogen) was used for routine cloning experiments and susceptibility testing of *mphs* cloned into pGDP4. *E. coli* BL21(DE3) (Stratagene, USA) was used for protein overexpression experiments. *E. coli* BW25113  $\Delta$ *bamB* $\Delta$ *tolC* is an antibiotic hypersensitive<sup>45</sup>, and was used for susceptibility testing of Mphs cloning into pGDP4. Antibiotic susceptibility testing was performed according to the Clinical and Laboratory Standards Institute guidelines<sup>74</sup> for determining minimal inhibitory concentration (MIC) in 96-well plates (Sarstedt, Germany) in duplicate. MIC values were reproducible. A saline suspension of OD<sub>600</sub> of 0.1 was made from 2–3 colonies of a fresh overnight, and diluted 1:200 into Mueller-Hinton Broth (BD Biosciences). MIC values were read after stationary incubation at 37 °C for 24 h. Susceptibility assays were performed at least two independent times.

**Protein overexpression and purification.** Mphs were purified as described previously for MphI<sup>36</sup>. *E. coli* BL21(DE3) was cultured with shaking at 37 °C in 1 L of LB-Lennox until an OD<sub>600</sub> of 0.5–0.6, chilled in an ice bath for 20 min, and protein overexpression induced by the addition of 1 mM IPTG and incubation at 17 °C for 16 h. Cell mass was harvested by centrifuging at 10,000 × g for 20 min, washed with cold saline, and resuspended in 20 mL buffer A (50 mM HEPES pH 7.5, 150 mM NaCl, 5% glycerol, 10 mM imidazole). Cell lysis was performed using a One-shot Cell Disruptor (Constant Systems Limited) at 20,000 psi, and cell debris removed after adding 15 mL buffer A, 5 mg bovine pancreas DNase, and 2.5 mg of bovine pancreas RNase by centrifugation at 50,000 × g for 45 min. Proteins were purified using 1 mL Ni<sup>2+</sup>-nitrilotriacetic acid column (Qiagen) and a linear gradient of 95% buffer A to 100% buffer B (buffer A with 250 mM imidazole) over 20 column volumes. Fractions with pure Mph were pooled and desalted into 50 mM HEPES pH 7.5 using a PD-10 desalting column (GE Scientific). Pure enzyme stocks were stored at 4 °C.

**Characterization of phosphorylated macrolides.** The product of MphB, MphI, MphK, Azi\_1.1, and Ery\_1 catalyzed reactions were confirmed with LC-MS analysis. Each 100 µL reaction consisted of Mph Buffer (50 mM HEPES pH 7.5, 40 mM KCl, 10 mM MgCl<sub>2</sub>), 1 mg mL<sup>-1</sup> azithromycin or telithromycin, 2 mM GTP, and 125–250 µg of enzyme. Control reactions were prepared identically, but devoid of enzyme. Reactions were incubated at 37 °C overnight, stopped with the addition of 100 µL cold methanol and stored at -20 °C. Before analysis, samples were centrifuged at 16,000 × g for 10 min. 20 µL of each reaction was injected onto an Agilent 1100 Series LC system and a QTRAP LC/MS/MS System (ABSciex) using the following HPLC conditions; isocratic 5% solvent B (0.05% formic acid in acetonitrile), 95% solvent A (0.05% formic acid in water) over 1 min, followed by a linear gradient to 97% B over 7 min at a flow rate of 1 mL min<sup>-1</sup> and C18 column (Sunfire, 5 µm, 4.6×50 mm).

**Structural elucidation of phosphorylated erythromycin.** Tandem mass spectrometry was used to identify the site of phosphorylation by MphB, Azi\_1.1, and Ery\_1. Each 100 µL reaction consisted of Mph Buffer, 0.5 mg mL<sup>-1</sup> erythromycin, 2 mM GTP, and 125–250 µg enzyme. Reactions were performed as described above. After centrifugation, samples were diluted 1:1000 into 50% methanol. High resolution electrospray ionization mass spectra were acquired using Agilent 1290 UPLC separation module and qTOF G6550A mass detector in positive ion mode. Fragmentor voltage was set to 365.0 V. Liquid chromatography separation was carried out using Eclipse C18 (3.5 µm, 2.1 × 100 mm) column (Agilent Technologies) and the following pump method: at 0 min 95% solvent A (0.1% v/v formic acid in water), from 1 to 7 min up to 97% solvent B (0.1% v/v of formic acid in acetonitrile), at a flow rate 0.4 mL min<sup>-1</sup>. Targeted MS<sup>2</sup> acquisition mode was used for the fragmentation under 3 different fixed collision induced dissociation (CID) settings: 10, 30 and 40 eV. Targeted list contained the following m/z: 734.4681 (erythromycin [M+H]<sup>+</sup>) and 814.4342 (Mph erythromycin product [M+H]<sup>+</sup>). Fragments were identified based on ref. 42.

**Steady-state kinetics of Mphs.** Steady-state kinetics for Mphs were measured using the continuous PK/LDH-coupled assay<sup>27,32,36,71</sup>. Reactions were performed in 96-well plates (Nunc, Thermo Scientific) using Spectramax Plus384 (Molecular Devices) in duplicate. Individual reactions were 100 µL and contained 50 mM HEPES pH 7.5, 40 mM KCl, 10 mM MgCl<sub>2</sub>, 0.2 mM NADH, 3.5 mM PEP, and 4.8 U PK/LDH (pyruvate kinase/lactate dehydrogenase). For determining the K<sub>m</sub> for GTP, telithromycin was maintained at 100 µM and GTP was varied between 0.49 and 250 µM. For determining the K<sub>m</sub> for macrolides, GTP was maintained at 200 µM and macrolides were varied between 0.49 µM and 1000 µM. Enzyme efficiency varied considerably between substrates, and therefore some enzymes were used at different concentrations depending on the substrate in order to produce a linear range suitable for K<sub>m</sub> determination. These concentrations were as follows; MphB 140 nM (most substrates), MphB 36 nM (josamycin), MphI 45 nM (descladinose macrolides), MphI 1.44 µM (C3 cladinose macrolides), Azi\_1.1 90 nM (all substrates), Ery\_1 (all substrates), MphK 680 nM (all substrates), MphI<sub>S111R/T166E/P211L</sub> 90 nM (descladinose macrolides), MphI<sub>S111R/T166E/P211L</sub> 1.43 µM (C3 cladinose macrolides), MphI<sub>M40A/S111R/T166E</sub> 22 nM (descladinose macrolides), MphI<sub>M40A/S111R/T166E</sub> 430 nM (C3 cladinose macrolides), MphI<sub>S111R/T166E</sub> 22 nM (descladinose

macrolides), MphI<sub>S111R/T166E</sub> 1.43  $\mu\text{M}$  (C3 cladinose macrolides). Enzyme kinetics experiments were performed at least two independent times.

**Ancestral reconstruction of ancestral Mph sequences.** Close Mph homologs were identified with ncbi-blast-2.2.31 + BLASTp<sup>68</sup> by querying the Mph sequences in Supplementary Table 1, e-value cutoff of  $1e-50$ , -max\_target\_seqs 500, and -entrez\_query 'NOT partial'. A close homolog was defined as having >80% identity and an alignment length >85% of the query length. Sequences were aligned using MAFFT (L-INS-i) and the BLOSUM45 substitution matrix. The alignment was weighted using the T-coffee TCS function. A maximum likelihood tree was generated with RAxML with the JTT model (PROTGAMAAUTO) and rapid bootstrap analysis of 100 replicates. Reconstructed ancestral sequences were inferred using an Empirical Bayes approach (PAML 4.9a) with the Jones amino acid replacement matrix. For ancestral reconstruction, the RAxML generated tree was rooted at the point of functional divergence and used the unweighted multiple sequence alignment. The ancestral sequences of each Mph subtype (i.e., those that phosphorylate C3 cladinose macrolide efficiently and those that do it poorly) were contrasted, and these positions were mapped onto the MphH crystal structure to identify residues near the macrolide-binding pocket. The alignment for ancestral reconstruction was also used to identify amino acid conservation for sequence logs.

**Construction of MphI mutant library.** We constructed the MphI mutant library by simultaneously introducing 'NNK' codons at amino acid positions 40, 111, 166, and 211 using a hierarchical approach based on the Gibson Assembly method<sup>75</sup>. This method assembled 4 PCR fragments; 3 sections of *mphI* and another containing part of the gene and the vector. Oligonucleotides were designed to amplify 93–358 ( $f_1$ ) and 471–633 ( $f_2$ ), and incorporate 'NNK' codons at the target positions with an additional 25 bp on each end for Gibson Assembly. Additionally, 334–495 ( $f_3$ ) was amplified, as well as the vector, 1–117, and 634–927 ( $f_4$ ) (Supplementary Fig. 9). PCR fragments were amplified with Phusion Polymerase (Thermo Scientific) and gel purified with 2% low melting point agarose (Thermo Scientific). Fragments  $f_1$  (266 bp),  $f_2$  (162 bp),  $f_3$  (188 bp), and  $f_4$  backbone (3636 bp) were assembled in a 20  $\mu\text{L}$  Gibson Assembly reaction (performed in triplicate) and incubated at 50 °C for 1 h. The Gibson reaction contained; 100 ng  $f_4$  backbone, 20 ng  $f_1$ , 12 ng  $f_2$ , 14 ng  $f_3$ , 0.004 U T5 exonuclease (New England Biolabs), 0.025 U Phusion polymerase, 1 U Taq ligase (New England Biolabs), 0.2 mM each dNTP (FroggaBio), 1 mM NAD (Sigma), 100 mM Tris-HCl pH 7.5 (Bioshop), 10 mM MgCl<sub>2</sub> (Fisher), 10 mM DTT (Bioshop), and 125 mg mL<sup>-1</sup> PEG-8000 (Sigma). Reactions were pooled, purified using PCR purification kit (Thermo Scientific) and eluted in 20  $\mu\text{L}$  water. A volume of 1  $\mu\text{L}$  was used to transform three individual *E. coli* Electromax electrocompetent cell (Invitrogen) reactions. Cells were recovered in a total of 3 mL Super Optimal broth with Catabolite repression (SOC) (Invitrogen) for 1 h at 37 °C with shaking. Library titer was estimated by plating  $10^{-3}$  and  $10^{-4}$  dilutions and determining the CFU mL<sup>-1</sup> of transformation. Plasmids from 10 random colonies were isolated, sent for Sanger sequencing (Mobix, McMaster University), and verified to contain a unique missense mutation at each of the four target positions. In total, the library contained 3.9 million variants. The rest of the transformations were used to inoculate 100 mL LB-Lennox with 100  $\mu\text{g}$  mL<sup>-1</sup> carbenicillin and incubated with shaking for 16 h at 37 °C. 50 mL was centrifuged at 4000  $\times$  g for 1 h at 4 °C, resuspended in 40 mL of fresh LB-Lennox and 10 mL of 80% glycerol, and stored at -80 °C. The library was diluted to OD<sub>600</sub> of 0.25 (~200 million CFU mL<sup>-1</sup>) and 250  $\mu\text{L}$  was plated (~51  $\times$  the library) on LB agar with 20  $\mu\text{g}$  mL<sup>-1</sup> azithromycin and 100  $\mu\text{g}$  mL<sup>-1</sup> carbenicillin. In another experiment, the library was diluted to OD<sub>600</sub> of 0.022 (~17.8 million CFU mL<sup>-1</sup>) and 250  $\mu\text{L}$  was plated (~4.6  $\times$  the library) on 250  $\mu\text{g}$  mL<sup>-1</sup> erythromycin and 75  $\mu\text{g}$  mL<sup>-1</sup> carbenicillin. The concentration of azithromycin and erythromycin was such that *E. coli* Electromax harboring wild-type *mphI* could not grow. Plasmids from azithromycin or erythromycin-resistant mutants were isolated, transformed into *E. coli* TOP10, and confirmed to be resistant to C3 cladinose macrolides. This verified that the resistance phenotype was encoded on the plasmid and not a chromosomal mutation. Missense mutations were identified by Sanger sequencing (Mobix, McMaster University). For protein purification experiments MphI mutants were subcloned into pET22b.

**Site-directed mutagenesis.** Oligonucleotides in Supplementary Table 16 were used for site-directed mutagenesis in the following 50  $\mu\text{L}$  reaction; 10 ng plasmid, 0.5  $\mu\text{M}$  primers, 200  $\mu\text{M}$  dNTPs, 1  $\times$  HF buffer, 1 U Phusion polymerase. The following thermocycler conditions were used; initial denaturation 98 °C 2 min, 98 °C 10 s, 55 °C 20 s, 72 °C 3.75 min for 16 cycles, and 7 min final extension at 72 °C. An aliquot of 2  $\mu\text{L}$  of FastDigest DpnI (Thermo Scientific) was added and incubated at 37 °C for 1 h to remove template DNA. Reactions were cleaned up with PCR purification, and 5  $\mu\text{L}$  was used to transform chemically competent *E. coli* TOP10. Plasmids were isolated and sent for Sanger sequencing to confirm the desired mutations. Wild-type MphI was used for individual mutations predicted with ancestral reconstruction. The Ery\_1 MphI mutant was used to reverse mutations to wildtype. All mutations were made independently in pET28a-*mphI*, pET28a-*ery\_1*, pGDP4-*mphI*, and pGDP4-*ery\_1*.

**Taxonomic distribution of Mphs.** The RefSeq genome database was used to evaluate the taxonomic distribution of Mphs in bacteria. The assembly report for all 82,640 RefSeq genomes (as of 29 March 2017) was used to correlate nucleotide accession numbers with RefSeq genomes. The identical protein report was acquired for each protein accession from the phylogenetic reconstruction of diverse Mphs using the efetch Entrez function. Strains with a Mph were identified if the nucleotide accession of a Mph was within the list of RefSeq genomes above. Strains with 2 Mphs were normalized to 1 by matching the unique 4 letter prefix of each nucleotide accession number for WGS sequences, or by matching the entire accession number for complete genomes. This was performed using custom python code.

**Crystal structure determination of Mphs.** For crystallization, Mph genes were subcloned into the pMCSG53 expression vector that codes for a N-terminal His<sub>6</sub>-tag and a TEV protease cleavage site. Se-Met-substituted versions of MphB and MphH were expressed using the standard M9 high-yield growth procedure according to the manufacturer's instructions (Shanghai Medicilon). Crystallization was performed at room temperature using the sitting drop method and 2  $\mu\text{L}$  protein or protein:ligand mixture plus 2  $\mu\text{L}$  reservoir solution. The MphH concentration was 10 mg mL<sup>-1</sup> and the MphB concentration was 90 mg mL<sup>-1</sup>. Se-Met MphB (apoenzyme) was crystallized with reservoir solution 0.2 M calcium acetate, 0.1 M sodium cacodylate pH 6.5, 18% (w/v) PEG 8 K. Native MphH (apoenzyme) was crystallized with reservoir solution 2 M ammonium sulfate, 0.1 M Hepes pH 7.5, 2% (w/v) PEG 400 and 1% trehalose. Se-Met MphH•GDP complex was crystallized with reservoir solution 2 M ammonium sulfate, 0.1 M Hepes pH 7.5, 2% (w/v) PEG400, 1% trehalose, 5 mM azithromycin and 5 mM GMPPCP. Native MphH•azithromycin complex was crystallized with reservoir solution 2 M ammonium sulfate, 5 mM azithromycin and 5% DMSO. All crystals were cryoprotected in paratone oil prior to data collection. Diffraction data for MphH (native, apoenzyme) was collected at 100 K using a Rigaku Micromax 007-HF rotating anode and a Rigaku R-Axis IV ++ detector. Diffraction data for MphB (Se-Met, apoenzyme), MphH•GDP (Se-Met) and MphH•azithromycin (native) were collected at the Advanced Photon Source, Argonne National Laboratory, Life Sciences Collaborative Access Team beamlines 21-ID-F or 21-ID-G at the selenium absorption edge. Data was processed by HKL-3000<sup>76</sup> or XDS<sup>77</sup> and Aimless<sup>78</sup>. SAD structure determination was completed using Phenix.autosol<sup>79</sup> and Molecular Replacement for native structures was compiled using Phenix.phaser. Refinement was completed with Phenix.refine and Coot<sup>80</sup>. All B-factors were refined as isotropic with TLS parameterization. All geometry was verified using Phenix validation tools and the wwPDB server. Electron density maps are in Supplementary Figure 10.

**Code availability.** The custom python code used to survey taxonomic distribution of Mphs is available from the corresponding author upon request.

**Data availability.** The crystal structures of Mph enzymes were deposited into the PDB; MphB Se-Met apoenzyme (5UXA), MphB native apoenzyme (5UXB), MphH (native) GDP + Pi complex (5UXC), and MphH (Se-Met) azithromycin complex (5UXD). Protein accession numbers for all Mph enzymes included in this study are listed in Supplementary Table 1. Other data are available from the corresponding author upon reasonable request.

Received: 7 September 2017 Accepted: 19 December 2017

Published online: 09 January 2018

## References

- Perry, J. A. & Wright, G. D. Forces shaping the antibiotic resistome. *BioEssays* **36**, 1179–1184 (2014).
- Davies, J. & Davies, D. Origins and evolution of antibiotic resistance. *Microbiol. Mol. Biol. Rev.* **74**, 417–433 (2010).
- Jia, B. et al. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res.* **45**, D566–D573 (2017).
- Gibson, M. K., Forsberg, K. J. & Dantas, G. Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology. *ISME J.* **9**, 207–216 (2015).
- Crofts, T. S., Gasparrini, A. J. & Dantas, G. Next-generation approaches to understand and combat the antibiotic resistome. *Nat. Rev. Microbiol.* **15**, 422–434 (2017).
- Robicsek, A. et al. Fluoroquinolone-modifying enzyme: a new adaptation of a common aminoglycoside acetyltransferase. *Nat. Med.* **12**, 83–88 (2006).
- Chanal, C. et al. Nucleotide sequences of CAZ-2, CAZ-6, and CAZ-7 beta-lactamase genes. *Antimicrob. Agents Chemother.* **36**, 1817–1820 (1992).
- Henquell, C., Chanal, C., Sirot, D., Labia, R. & Sirot, J. Molecular characterization of nine different types of mutants among 107 inhibitor-

- resistant TEM beta-lactamases from clinical isolates of *Escherichia coli*. *Antimicrob. Agents Chemother.* **39**, 427–430 (1995).
9. Casin, I. et al. Aminoglycoside 6'-N-acetyltransferase variants of the Ib type with altered substrate profile in clinical isolates of *Enterobacter cloacae* and *Citrobacter freundii*. *Antimicrob. Agents Chemother.* **42**, 209–215 (1998).
  10. Silver, L. L. Challenges of antibacterial discovery. *Clin. Microbiol. Rev.* **24**, 71–109 (2011).
  11. Sommer, M. O. A., Munck, C., Toft-Kehler, R. V. & Andersson, D. I. Prediction of antibiotic resistance: time for a new preclinical paradigm? *Nat. Rev. Micro* **15**, 689–696 (2017).
  12. Bonnefoy, A., Girard, A. M., Agouridas, C. & Chantot, J. F. Ketolides lack inducibility properties of MLS(B) resistance phenotype. *J. Antimicrob. Chemother.* **40**, 85–90 (1997).
  13. Armstrong, E. S. & Miller, G. H. Combating evolution with intelligent design: the neoglycoside ACHN-490. *Curr. Opin. Microbiol.* **13**, 565–573 (2010).
  14. Papp-Wallace, K. M., Endimiani, A., Taracila, M. A. & Bonomo, R. A. Carbapenems: past, present, and future. *Antimicrob. Agents Chemother.* **55**, 4943–4960 (2011).
  15. Sum, P. E. & Petersen, P. Synthesis and structure-activity relationship of novel glycylicycline derivatives leading to the discovery of GAR-936. *Bioorg. Med. Chem. Lett.* **9**, 1459–1462 (1999).
  16. Rolinson, G. N. Forty years of beta-lactam research. *J. Antimicrob. Chemother.* **41**, 589–603 (1998).
  17. Fernandes, P., Martens, E. & Pereira, D. Nature nurtures the design of new semi-synthetic macrolide antibiotics. *J. Antibiot.* **70**, 527–533 (2017).
  18. Hicks, L. A. et al. US outpatient antibiotic prescribing variation according to geography, patient population, and provider specialty in 2011. *Clin. Infect. Dis.* **60**, 1308–1316 (2015).
  19. Gomes, C. et al. Macrolide resistance mechanisms in Enterobacteriaceae: focus on azithromycin. *Crit. Rev. Microbiol.* **43**, 1–30 (2017).
  20. Kannan, K. et al. The general mode of translation inhibition by macrolide antibiotics. *Proc. Natl Acad. Sci. USA* **111**, 15958–15963 (2014).
  21. Dunkle, J. A., Xiong, L., Mankin, A. S. & Cate, J. H. D. Structures of the *Escherichia coli* ribosome with antibiotics bound near the peptidyl transferase center explain spectra of drug action. *Proc. Natl Acad. Sci. USA* **107**, 17152–17157 (2010).
  22. Tu, D., Blaha, G., Moore, P. B. & Steitz, T. A. Structures of MLSBK antibiotics bound to mutated large ribosomal subunits provide a structural explanation for resistance. *Cell* **121**, 257–270 (2005).
  23. Bailey, M., Chettiath, T. & Mankin, A. S. Induction of *erm(C)* expression by noninducing antibiotics. *Antimicrob. Agents Chemother.* **52**, 866–874 (2008).
  24. Wolter, N. et al. Telithromycin resistance in *Streptococcus pneumoniae* is conferred by a deletion in the leader sequence of *erm(B)* that increases rRNA methylation. *Antimicrob. Agents Chemother.* **52**, 435–440 (2008).
  25. Park, B. & Min, Y. H. Inducible expression of *erm(B)* by the ketolides telithromycin and cethromycin. *Int. J. Antimicrob. Agents* **46**, 226–227 (2015).
  26. Roberts, M. C. Update on macrolide–lincosamide–streptogramin, ketolide, and oxazolidinone resistance genes. *FEMS Microbiol. Lett.* **282**, 147–159 (2008).
  27. Bhullar, K. et al. Antibiotic resistance is prevalent in an isolated cave microbiome. *PLoS ONE* **7**, e34953 (2012).
  28. Ohara, K., Kanda, T. & Kono, M. Structure of a phosphorylated derivative of oleandomycin, obtained by reaction of oleandomycin with an extract of an erythromycin-resistant strain of *Escherichia coli*. *J. Antibiot.* **41**, 823–827 (1988).
  29. Hon, W. C. et al. Structure of an enzyme required for aminoglycoside antibiotic resistance reveals homology to eukaryotic protein kinases. *Cell* **89**, 887–895 (1997).
  30. Skinner, R. H. & Cundliffe, E. Resistance to the antibiotics viomycin and capreomycin in the *Streptomyces* species which produce them. *J. Gen. Microbiol.* **120**, 95–104 (1980).
  31. Magnet, S. & Blanchard, J. S. Molecular insights into aminoglycoside action and resistance. *Chem. Rev.* **105**, 477–498 (2005).
  32. Shakya, T. & Wright, G. D. Nucleotide selectivity of antibiotic kinases. *Antimicrob. Agents Chemother.* **54**, 1909–1913 (2010).
  33. Noguchi, N., Katayama, J. & O'Hara, K. Cloning and nucleotide sequence of the *mphB* gene for macrolide 2'-phosphotransferase II in *Escherichia coli*. *FEMS Microbiol. Lett.* **144**, 197–202 (1996).
  34. Matsuoka, M., Inoue, M., Endo, Y. & Nakajima, Y. Characteristic expression of three genes, *msr(A)*, *mph(C)* and *erm(Y)*, that confer resistance to macrolide antibiotics on *Staphylococcus aureus*. *FEMS Microbiol. Lett.* **220**, 287–293 (2003).
  35. Schluter, A. et al. Erythromycin resistance-conferring plasmid pRSB105, isolated from a sewage treatment plant, harbors a new macrolide resistance determinant, an integron-containing Tn402-like element, and a large region of unknown function. *Appl. Environ. Microbiol.* **73**, 1952–1960 (2007).
  36. Pawlowski, A. C. et al. A diverse intrinsic antibiotic resistome from a cave bacterium. *Nat. Commun.* **7**, 13803 (2016).
  37. Wang, C. et al. Functional characterization and phylogenetic analysis of acquired and intrinsic macrolide phosphotransferases in the *Bacillus cereus* group. *Environ. Microbiol.* **17**, 1560–1573 (2015).
  38. Pawlowski, A. C., Westman, E. L., Koteva, K., Waglechner, N. & Wright, G. D. The complex resistomes of Paenibacillaceae reflect diverse antibiotic chemical ecologies. *ISME J.* <https://doi.org/10.1038/s41396-017-0017-5> (2017).
  39. Szczepanowski, R., Krahn, I., Bohn, N., Puhler, A. & Schluter, A. Novel macrolide resistance module carried by the IncP-1beta resistance plasmid pRSB111, isolated from a wastewater treatment plant. *Antimicrob. Agents Chemother.* **51**, 673–678 (2007).
  40. Nonaka, L., Maruyama, F., Suzuki, S. & Masuda, M. Novel macrolide-resistance genes, *mef(C)* and *mph(G)*, carried by plasmids from *Vibrio* and *Photobacterium* isolated from sediment and seawater of a coastal aquaculture site. *Let. Appl. Microbiol.* **61**, 1–6 (2015).
  41. Yang, J. et al. Characterization of a multiresistant mosaic plasmid from a fish farm Sediment *Exiguobacterium* sp. isolate reveals aggregation of functional clinic-associated antibiotic resistance genes. *Appl. Environ. Microbiol.* **80**, 1482–1488 (2014).
  42. Crowe, M. C., Brodbelt, J. S., Goolsby, B. J. & Hergenrother, P. Characterization of erythromycin analogs by collisional activated dissociation and infrared multiphoton dissociation in a quadrupole ion trap. *J. Am. Soc. Mass. Spectrom.* **13**, 630–649 (2002).
  43. Cox, G. et al. A common platform for antibiotic dereplication and adjuvant discovery. *Cell Chem. Biol.* **24**, 98–109 (2017).
  44. Chesneau, O., Tsvetkova, K. & Courvalin, P. Resistance phenotypes conferred by macrolide phosphotransferases. *FEMS Microbiol. Lett.* **269**, 317–322 (2007).
  45. Fong, D. H., Burk, D. L., Blanchet, J., Yan, A. Y. & Berghuis, A. M. Structural basis for kinase-mediated macrolide antibiotic resistance. *Structure* **25**, 750–761 e755 (2017).
  46. Chen, F. et al. Reconstructed evolutionary adaptive paths give polymerases accepting reversible terminators for sequencing and SNP detection. *Proc. Natl Acad. Sci. USA* **107**, 1948–1953 (2010).
  47. Randall, R. N., Radford, C. E., Roof, K. A., Natarajan, D. K. & Gaucher, E. A. An experimental phylogeny to benchmark ancestral sequence reconstruction. *Nat. Commun.* **7**, 12847 (2016).
  48. Cole, M. F., Cox, V. E., Gratton, K. L. & Gaucher, E. A. Reconstructing evolutionary adaptive paths for protein engineering. *Methods Mol. Biol.* **978**, 115–125 (2013).
  49. Zakas, P. M. et al. Enhancing the pharmaceutical properties of protein drugs by ancestral sequence reconstruction. *Nat. Biotechnol.* **35**, 35–37 (2017).
  50. Risso, V. A., Gavira, J. A., Mejia-Carmona, D. F., Gaucher, E. A. & Sanchez-Ruiz, J. M. Hyperstability and substrate promiscuity in laboratory resurrection of Precambrian beta-lactamases. *J. Am. Chem. Soc.* **135**, 2899–2902 (2013).
  51. Yang, Z. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**, 555–556 (1997).
  52. Raquet, X. et al. Stability of TEM beta-lactamase mutants hydrolyzing third generation cephalosporins. *Proteins* **23**, 63–72 (1995).
  53. Jacquier, H. et al. Capturing the mutational landscape of the beta-lactamase TEM-1. *Proc. Natl Acad. Sci. USA* **110**, 13067–13072 (2013).
  54. Holbrook, S. Y. & Garneau-Tsodikova, S. Expanding aminoglycoside resistance enzyme regiospecificity by mutation and truncation. *Biochemistry* **55**, 5726–5737 (2016).
  55. Seiple, I. B. et al. A platform for the discovery of new macrolide antibiotics. *Nature* **533**, 338–345 (2016).
  56. Zou, T., Risso, V. A., Gavira, J. A., Sanchez-Ruiz, J. M. & Ozkan, S. B. Evolution of conformational dynamics determines the conversion of a promiscuous generalist into a specialist enzyme. *Mol. Biol. Evol.* **32**, 132–143 (2015).
  57. Gonzalez, M. M., Abriata, L. A., Tomatis, P. E. & Vila, A. J. Optimization of conformational dynamics in an epistatic evolutionary trajectory. *Mol. Biol. Evol.* **33**, 1768–1776 (2016).
  58. Tomatis, P. E. et al. Adaptive protein evolution grants organismal fitness by improving catalysis and flexibility. *Proc. Natl Acad. Sci. USA* **105**, 20605–20610 (2008).
  59. Morar, M. & Wright, G. D. The genomic enzymology of antibiotic resistance. *Annu. Rev. Genet.* **44**, 25–51 (2010).
  60. Wright, G. D. The antibiotic resistome: the nexus of chemical and genetic diversity. *Nat. Rev. Microbiol.* **5**, 175–186 (2007).
  61. 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 (2014).
  62. Orenca, M. C., Yoon, J. S., Ness, J. E., Stemmer, W. P. & Stevens, R. C. Predicting the emergence of antibiotic resistance by directed evolution and structural analysis. *Nat. Struct. Biol.* **8**, 238–242 (2001).
  63. Stemmer, W. P. Rapid evolution of a protein in vitro by DNA shuffling. *Nature* **370**, 389–391 (1994).
  64. Haller, G. et al. Massively parallel single-nucleotide mutagenesis using reversibly terminated inosine. *Nat. Methods* **13**, 923–924 (2016).

65. Bratulic, S., Gerber, F. & Wagner, A. Mistranslation drives the evolution of robustness in TEM-1 beta-lactamase. *Proc. Natl Acad. Sci. USA* **112**, 12758–12763 (2015).
66. Gaucher, E. A., Thomson, J. M., Burgan, M. F. & Benner, S. A. Inferring the palaeoenvironment of ancient bacteria on the basis of resurrected proteins. *Nature* **425**, 285–288 (2003).
67. Pawlowski, A. C., Johnson, J. W. & Wright, G. D. Evolving medicinal chemistry strategies in antibiotic discovery. *Curr. Opin. Biotechnol.* **42**, 108–117 (2016).
68. Camacho, C. et al. BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
69. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).
70. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
71. Shakya, T. et al. A small molecule discrimination map of the antibiotic resistance kinome. *Chem. Biol.* **18**, 1591–1601 (2011).
72. Chang, J. M., Di Tommaso, P. & Notredame, C. TCS: a new multiple sequence alignment reliability measure to estimate alignment accuracy and improve phylogenetic tree reconstruction. *Mol. Biol. Evol.* **31**, 1625–1637 (2014).
73. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
74. Clinical and Laboratory Standards Institute. *Methods for dilution: antimicrobial susceptibility testing of bacteria that grow aerobically*. (CLSI, Wayne, PA, USA, 2012).
75. Gibson, D. G. et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
76. Minor, W., Cymborowski, M., Otwinowski, Z. & Chruszcz, M. HKL-3000: the integration of data reduction and structure solution—from diffraction images to an initial model in minutes. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 859–866 (2006).
77. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
78. Winn, M. D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242 (2011).
79. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
80. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).

## Acknowledgments

We thank Nicholas Waglechner (McMaster) for useful discussions on phylogenetics, Mph surveillance, and ancestral sequence reconstruction. We also thank Murray Junop (Western University) for helpful discussion on protein engineering experiments. This research was funded by a Canadian Institutes of Health Research grant (FRN-148463)

and by a Canada Research Chair in Antibiotic Biochemistry (to G.D.W.) and through funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under Contract No. HHSN272201200026C.

## Author contributions

A.C.P. and G.D.W. designed research. A.C.P., P.J.S., and K.K. performed research and analyzed data. A.C.P. performed the following experiments; antibiotic susceptibility testing, phylogenetics, enzyme kinetics, ancestral sequence reconstruction, mass spectrometry, site-saturation combinatorial mutagenesis, and in vitro enzyme assays. P.G.S. performed crystal structure determination of MphH. K.K. performed tandem mass spectrometry experiments. T.S. and E.E. performed protein crystallization. A.S. supervised protein crystallization and contributed to interpreting mechanism of expanded substrate range. A.C.P., P.J.S., and G.D.W. wrote the manuscript with input from all authors.

## Additional information

**Supplementary Information** accompanies this paper at <https://doi.org/10.1038/s41467-017-02680-0>.

**Competing interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017