



OPEN

Integron gene cassettes harboring novel variants of D-alanine-D-alanine ligase confer high-level resistance to D-cycloserine

Md. Ajjur Rahman^{1,2}, Frank Kaiser¹, Shirin Jamshidi³, Marta Freitas Monteiro^{1,6}, Khondaker Miraz Rahman³, Peter Mullany¹ & Adam P. Roberts^{4,5}✉

Antibiotic resistance poses an increasing threat to global health. To tackle this problem, the identification of principal reservoirs of antibiotic resistance genes (ARGs) plus an understanding of drivers for their evolutionary selection are important. During a PCR-based screen of ARGs associated with integrons in saliva-derived metagenomic DNA of healthy human volunteers, two novel variants of genes encoding a D-alanine-D-alanine ligase (*ddl6* and *ddl7*) located within gene cassettes in the first position of a reverse integron were identified. *Treponema denticola* was identified as the likely host of the *ddl* cassettes. Both *ddl6* and *ddl7* conferred high level resistance to D-cycloserine when expressed in *Escherichia coli* with *ddl7* conferring four-fold higher resistance to D-cycloserine compared to *ddl6*. A SNP was found to be responsible for this difference in resistance phenotype between the two *ddl* variants. Molecular dynamics simulations were used to explain the mechanism of this phenotypic change at the atomic scale. A hypothesis for the evolutionary selection of *ddl* containing integron gene cassettes is proposed, based on molecular docking of plant metabolites within the ATP and D-cycloserine binding pockets of Ddl.

Globally, the emergence of antibiotic resistance in clinically important pathogens pose a significant threat to human health. Antibiotic resistance genes (ARGs) carried on various mobile genetic elements (MGEs), including plasmids, transposons and integron gene cassettes (GCs) are responsible for the emergence, and worldwide dissemination, of multidrug-resistance. Identifying the principal reservoirs of ARGs^{1,2} as well as understanding the drivers for their evolutionary selection and spread³ are fundamental to tackling the problem of antibiotic resistance.

Integrons are genetic elements capable of capturing and expressing open reading frames (ORFs) embedded in GCs. They are well known for their role in dissemination of ARGs particularly in gram-negative pathogens. The basic structure of an integron is composed of a gene (*intI*) encoding a DNA integrase, an integron associated recombination site; *attI* and a gene cassette promoter (*Pc*), which expresses the inserted cassettes within the GC array. Usually the integron integrase gene and the genes in the GC array are divergently transcribed, however where the integrase and the GC array are transcribed in the same direction the integron is known as a “reverse integron”^{4,5}. The integron integrase (IntI) catalyses recombination between the cassette associated recombination site; *attC* and the attachment site; *attI*, within the integron platform^{6–8}. More than 130 MGE associated ARGs have been detected in integrons, conferring resistance to most classes of antibiotics including aminoglycosides, chloramphenicol, beta-lactams, trimethoprim and streptothricin^{9,10}. Integrons carry ARGs, plus other genes encoding proteins with more diverse functions including adaptation to stress, virulence and resistance to xenobiotics, and many, particularly those integrons located on chromosomes, encode proteins of

¹Department of Microbial Diseases, University College London, 256 Gray's Inn Road, London WC1X 8LD, UK. ²Department of Pharmacy, University of Rajshahi, Rajshahi 6205, Bangladesh. ³School of Cancer and Pharmaceutical Science, King's College London, London SE1 9NH, UK. ⁴Department of Tropical Disease Biology, Liverpool School of Tropical Medicine, Pembroke Place, Liverpool L3 5QA, UK. ⁵Centre for Drugs and Diagnostics, Liverpool School of Tropical Medicine, Liverpool, UK. ⁶Present address: Division of Immunology and Allergology, University Hospitals and Medical Faculty, Geneva, Switzerland. ✉email: Adam.Roberts@lstmed.ac.uk

Figure 1. Features of the *ddl* gene cassettes, their closest homologues and the phylogenetic analysis. (A) Genetic organisation of 2024 bp inserts in the pGEM-T Easy vector; (B) Comparison of putative *attI* sequence preceding *ddl* in the inserts with the putative *attI* of the integron of *T. denticola* ATCC 35405¹¹. The putative integrase binding sites S1 and S2 as well as DR1 and DR2 are also shown. The putative recombination point G↓TT and the putative transcription start site (TSS) located at the 3'-end of *attI* are also shown. (C) Comparison of partial sequence of *attC* detected at the 3'-end of the 2024 bp insert with a typical complete *attC*-associated with Tde1837 of *T. denticola* integron¹¹. (D) The percentage identity of the flanking sequence of *ddl6/ddl7* (40 bp upstream and 29 bp downstream) with their closest homologue, *ddl(GC)* located on a 5699 bp contig of *T. pedis* B683 genome (GenBank accession: NZ_AOTN01000179). The putative core sites of the *attC* (R" and L"), the simple integrase binding site (S1) are shown. The recombination points located on S1 and core site, R' of the *attC* located upstream of *ddl(GC)* (*attC_{HDI}*) are marked with the blue arrows. The binding sites for the primers, TDIF and MASRS2, used for amplification of the integrons and associated gene cassettes are also shown. The start codons (ATG) as well as the S1, R', R", L" sites are in bold. (E) Phylogenetic tree of *ddl6*, *ddl7* and *ddl(GC)* along with the *ddls* of different strains of *T. denticola* and other *Treponema* species. The *ddlA* of *E. coli* was used as an outgroup. The evolutionary relationship was inferred using the Neighbour-Joining method¹⁸. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches. Evolutionary analyses were conducted in MEGA X¹⁹.

unknown functions⁵. For instance, among the 70 ORFs detected within 47 gene cassettes in the chromosomal integron of *Treponema denticola*¹¹, an oral pathogen associated with periodontitis, only five ORFs matched with proteins of known functions (last confirmed August 2019).

This article reports the detection of two novel reverse integron-located genes encoding D-alanine-D-alanine ligases (Ddls). Ddls are ATP-dependent enzymes that play a critical role in the synthesis of peptidoglycan during the cytoplasmic stage by catalysing the formation of the D-ala-D-ala dipeptide required for cross-linking peptidoglycan strands during cell wall synthesis¹². The gene encoding Ddl is a housekeeping gene and has always been found in the chromosome, never in a MGE. The formation D-Ala-D-Ala dipeptide catalysed by Ddl is critical for bacterial growth^{13,14}. As there is no human homologue, Ddl is considered an excellent target for designing new antibiotics¹⁵. D-Cycloserine is the only inhibitor of Ddl used in the clinic¹³, however, due to the severe neurologic side effects, its use is currently reserved for the treatment of multidrug resistant (MDR) and extensively-drug resistant (XDR) tuberculosis^{16,17}.

We demonstrate D-cycloserine resistance associated with these integron located *ddls* and present an experimentally supported hypothesis as to why *ddl*, a housekeeping gene, has evolved to be on an integron gene cassette.

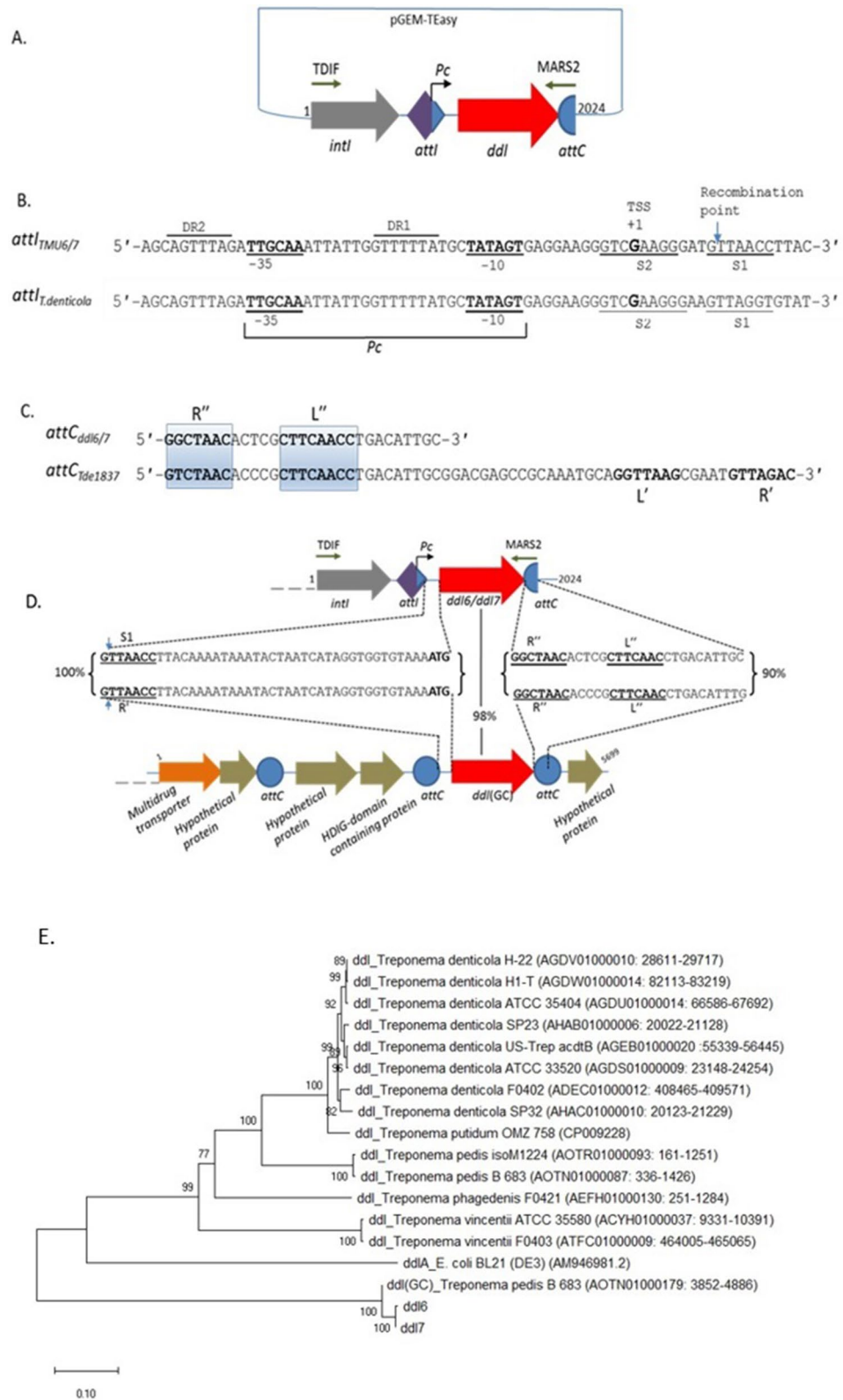
Results

Two novel variants of *ddl* were discovered from the human oral cavity located immediately downstream of the core structure of a reverse integron. During a PCR screen for the presence of integron GCs carrying ARGs in the saliva-derived metagenomic DNA of healthy volunteers from the UK (n = 11) and Bangladesh (n = 12), two natural variants of *ddl* located within the first GC of a reverse integron were discovered. In the libraries of PCR amplicons of integrons and associated GCs, a total of five clones (three from the UK and two from Bangladesh) were recovered carrying *ddl* predicted to encode a D-alanine-D-alanine ligase. Among the three *ddl*-carrying clones obtained from UK metagenomic libraries, one clone carried *ddl6* and two clones carried *ddl7* and among the two clones obtained from BD library, one clone carried *ddl6* and another clone *ddl7*. The genetic organisation of the *ddl* inserts in the clones are shown in Fig. 1A. The *intI* and the *ddl* gene cassettes are oriented in the same direction in reverse integrons (where the integrase is orientated in the opposite direction compared to usual integrons) of *T. denticola*¹¹. Two putative integrase binding sites [L or S2 and R or S1], two direct repeat sequences (DR1 and DR2) and a putative *attI/attC* recombination site were detected (Fig. 1B). The stop codon of the *ddl* ORF was located on the 7 bp *attC* core site, R" (Fig. 1C). The R" site is followed by the L" site-which covers the priming site of the reverse primer MARS2.

Two variants of *ddl* (named as *ddl6* and *ddl7*) were different from each other by two SNPs at c.490 and c.777 positions of the coding sequences of the genes. The CTT (Leu164) and TGG (Trp259) codons of *ddl6* were substituted with TTT (Phe164) and TGT (Cys259) in *ddl7*, respectively.

The closest homologue of *ddl6* and *ddl7* is a *ddl* located within the accessory genome of *T. pedis* B683. The closest homologue of the cassette-located *ddls* having 98% nucleotide sequence identity (97% identity at the amino acid level) was located on a 5699 bp contig of *T. pedis* B683 genome (GenBank accession: NZ_AOTN01000179) containing a total of six ORFs (Fig. 1D). This homologous *ddl* of *T. pedis*, named *ddl(GC)*, was flanked by two *attC* recombination sites. The sequences of the core sites (R" and L") of *attC* located downstream of *ddl(GC)*, were 100% identical with the corresponding sites of *attC* located downstream of *ddl6* and *ddl7* (Fig. 1D). Another *attC* was found upstream of *ddl(GC)* flanking two ORFs encoding hypothetical proteins (Fig. 1D). However, as there was no integrase gene, the association of this contig with an integron, although likely, could not be confirmed. The next closest homologue was Ddl of *Syntrophobotulus glycolicus* (ADY55260) with 55% amino acid identity, followed by *Clostridium* sp. (54%; WP_033164556), *Lachnoclostridium phytofermentans* (53%, WP_029502590) and *Paenibacillus pini* (52%; WP_036650467). Ddl proteins with altered specificity that confer resistance to vancomycin such as VanA (AAA65956), VanB (YP_009076352) and VanC (P29753) also shared low sequence identity (27 to 30%) with Ddl6 and Ddl7.

T. pedis B683 carries two *ddl* genes, one of which (GenBank accession: AOTN01000124: 8689-9912 bp) is most likely the house-keeping *ddl* as it is found in other strains of *T. pedis* including TA4 and TM1 and its



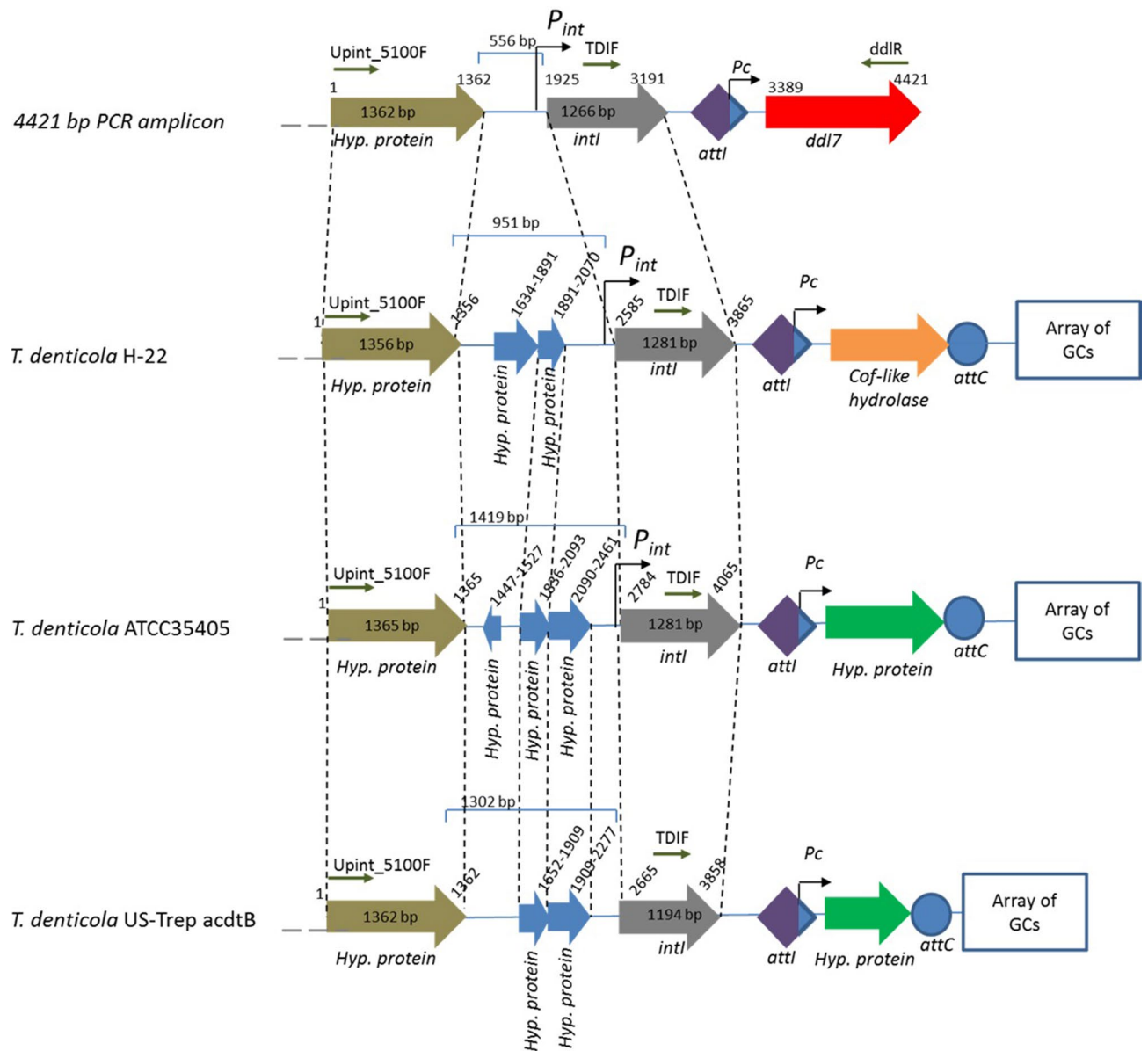


Figure 2. The genetic arrangement of 4421 bp pGEM-T Easy insert carrying *ddl7*, full length *intI* and another partial ORF encoding a hypothetical protein. The primers used to amplify this product are shown with small arrows. The genetic arrangements of the closest homologues including *T. denticola* H-22 (GenBank accession: AGDV01000005:c23656-19631), *T. denticola* ATCC35405 (GenBank accession: AE017226:1873126-1877351) and *T. denticola* US-Trep acdtB (GenBank accession: AGE01000014.1:3405-5167) have also been shown. The most variable region among the strains was found to be the sequence between the ORF for a hypothetical protein (Hyp. Protein in the top row) and *intI*.

G+C content (38.0%) is similar to the genome of *T. pedis* B683 (36.9%). In contrast, the G+C content of the GC located *ddl* is 30.6% suggesting that this second, cassette-located *ddl* may have been acquired by horizontal gene transfer by *T. pedis* B683.

Comparative genomics shows that a likely horizontal gene transfer (HGT) occurred between the *T. pedis* B683 and the host of *ddl6* and *ddl7*. The following three observations suggested HGT between *T. pedis* B683 and the host of *ddl6*/*ddl7*: (i) The identical 40 bp sequences upstream of *ddl6*/*ddl7* and their homologue *ddl*(GC); (ii) the nearly identical (90%) 29 bp downstream sequences containing R" and L" sites of *attC* (Fig. 1D); (iii) the phylogenetic tree constructed using the *ddls* of different species of *Treponema* showed that the integron-located *ddl6* and *ddl7* and *ddl*(GC) formed a separate clade within the tree (Fig. 1E).

Sequence of the upstream region of the integron carrying *ddl7* suggests that the putative host of the integron carrying *ddl6* and *ddl7* is a strain of *T. denticola*. To confirm if the initially

Bacterial strain	Plasmid vector	Inserts and their size	MIC of D-cycloserine ($\mu\text{g/mL}$)	MIC of vancomycin ($\mu\text{g/mL}$)
E. coli α-select				
EC121	pGEM-TEasy	Empty (vector control)	8	ND
EC126	pGEM-TEasy	<i>intI-attI-ddl6</i> (2024 bp)	16	ND
EC127	pGEM-TEasy	<i>intI-attI-ddl7</i> (2024 bp)	64	ND
EC206	pGEM-TEasy	<i>intI-attI-ddl6 c.490 C>T</i> (2024 bp)	16	ND
EC207	pGEM-TEasy	<i>intI-attI-ddl6 c.777G>T</i> (2024 bp)	64	ND
B. subtilis 168				
BS700	pHCMC05	Empty (vector control)	8	<0.125
BS706	pHCMC05	pHCMC05:: <i>ddl6</i>	64	<0.125
BS707	pHCMC05	pHCMC05:: <i>ddl7</i>	64	<0.125

Table 1. MIC of D-cycloserine against different strains of *E. coli* and *B. subtilis* expressing *ddl*. ND not determined.

recovered 2024 bp PCR amplicons were obtained from the *T. denticola* genome, sequence of the *intI* that was further upstream was amplified by PCR. Analysis of a 4421 bp PCR product showed that it contained a partial ORF encoding a hypothetical protein (1–1362 bp), a complete ORF encoding IntI (1266 bp: 1925–3191 bp), a putative *attI*, the *Pc* promoter followed by a GC carrying *ddl7* (3389–4421 bp) (Fig. 2). The complete *intI* had 96.0% nucleotide identity with *intI* of *T. denticola* H-22 (GenBank accession no: AGDV01000005:19792–21072). At the amino acid level, the identity ranges between 81 and 97% with IntI of different *Treponema* sp. Maximum identity was found with the IntI of *T. denticola* AL-2 (97%, EMB42956).

The genetic organisation of the 4421 bp amplicon was different from the closely related hits in the NCBI databases (Fig. 2). The non-protein-coding 556 bp sequence located between the partial ORF and the *intI* harbors the putative promoter for the *intI* (P_{intI}) with a sequence of 5'-ATGAAT [19 bp] TAAACT-3'. The upstream sequence of the *intI* was analyzed in silico for potential LexA binding sites similar to that found in the upstream of *intI* of mobile and chromosomal integrons²⁰. However, no LexA binding motif could be detected. Additionally, this region contains several inverted repeat and direct repeat sequences as well as a putative transcriptional terminator sequence.

A phylogenetic tree was constructed to compare the evolutionary relationship of *intIs* associated with the integron carrying *ddl* with the *intIs* of different species of *Treponema*. The tree showed that the *intI* associated with *ddl7* is closely related with other *intIs* of *T. denticola* and distantly related to *T. pedis* (Supplementary Fig. S1). The results obtained from the *intI* tree as well as the identity of the upstream sequence *intI* with the other strain of *T. denticola* suggest that the likely host of the integron associated with *ddl7* is a species of the genus *Treponema* and most likely a strain of *T. denticola*.

Expression of *ddl6* and *ddl7* confers resistance to D-cycloserine only, not to vancomycin and beta-lactams.

As the overexpression of *ddl* of *M. tuberculosis* and *M. smegmatis* was found to confer D-cycloserine resistance²¹, we hypothesised that the expression *ddl6* or *ddl7* in an integron array will result in D-cycloserine resistance, especially as they are detected in the first position due to the efficient expression by the *Pc*. When the minimum inhibitory concentrations (MIC) of D-cycloserine against the surrogate *E. coli* hosts EC126 and EC127 carrying pGEM-T Easy::*intI-attI-ddl6* and pGEM-T Easy::*intI-attI-ddl7*, respectively were tested, a two to four-fold increase of MIC of D-cycloserine was observed (Table 1). As the upstream region of *ddl6* and *ddl7* cassettes carrying the *attI* and *Pc* were 100% identical, we concluded that the fourfold change in the MIC was due to either the SNP c.490 or c.777 or both.

To determine if the overexpression of *ddl* in gram-positive bacteria can also confer resistance to D-cycloserine, vancomycin and β -lactam antibiotics, the 1032 bp coding sequence of *ddl6* and *ddl7* was cloned into a medium-copy expression vector, pHCMC05 and transformed into *B. subtilis* 168. The MIC of D-cycloserine against the strains BS706 and BS707 which carry pHCMC05::*ddl6* and pHCMC05::*ddl7* vectors, respectively was found to increase by eight-fold (64 $\mu\text{g/mL}$) compared to the BS700 strain that carries the empty plasmid (8 $\mu\text{g/mL}$) (Table 1). However, no difference in the MIC of D-cycloserine was observed between BS706 and BS707 carrying *ddl6* and *ddl7*, respectively, when the expression of the genes was induced by IPTG. No resistance to vancomycin was observed (Table 1). Nor was there any resistance to penicillin G, amoxicillin, oxacillin and ampicillin against both *E. coli* and *B. subtilis* strains. This suggests that overproduction of Ddl does not confer cross-resistance to antibiotics targeting cell-wall biosynthesis.

Site-directed mutagenesis confirmed that the SNP at c.777 was responsible for the alteration of MIC of D-cycloserine.

To identify which SNP of *ddl6* and *ddl7* was responsible for altering the susceptibility of *E. coli* to D-cycloserine, the nucleotides at position c.490 and c.777 of *ddl6* were changed to the corresponding nucleotides of *ddl7* by site-directed mutagenesis. MICs of D-cycloserine against *E. coli* carrying the constructs were determined and no change in the susceptibility was observed by the c.490 C>T substitution in

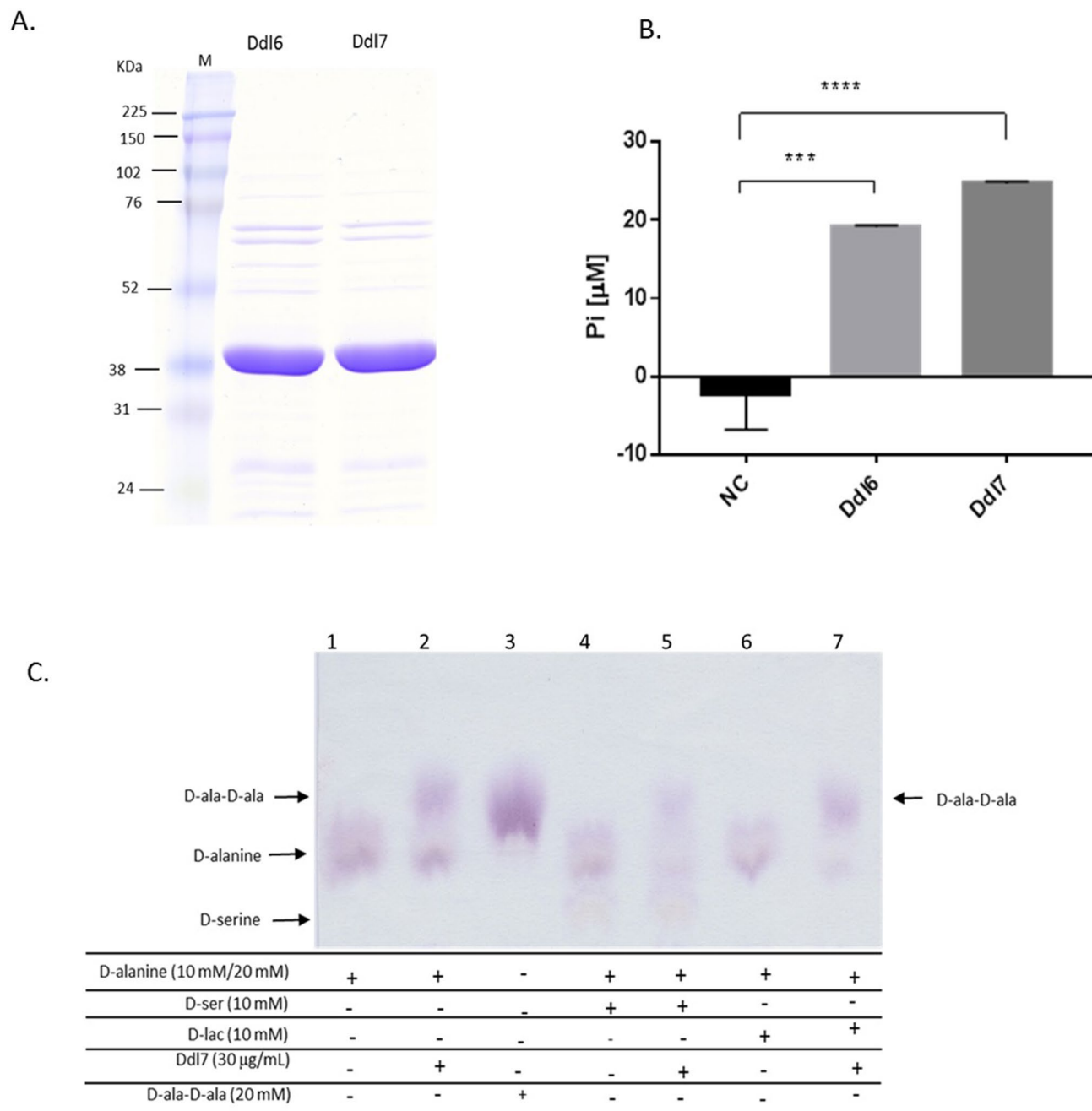


Figure 3. Purification of Ddl and determination of the functional activity Ddl. **(A)** SDS-PAGE of purified Ddl6 and Ddl7. **(B)** Release of inorganic phosphate in the reactions catalysed by Ddl6 and Ddl7 in the presence of 20 mM D-alanine. The differences of the release of Pi in between the negative controls (no Ddl) and Ddl catalysed reactions were analysed by one-way ANOVA (** $P < 0.001$; *** $P < 0.0001$). **(C)** Ascending paper chromatography to detect the formation of D-ala-D-ala and D-ala-D-ser dipeptide as well as D-ala-D-lac depsiptide. In the first two reactions (lane 1 and lane 2) 20 mM D-alanine was used, whereas in other reactions it was reduced to 10 mM and supplemented with 10 mM D-serine or D-lactate.

ddl6, however, the G>T substitution which alters Trp259 to Cys259 (W259C) increased the MIC fourfold which is similar to the MIC of the strain carrying *ddl7* (Table 1).

The integron-encoded Ddls are functional. The biological activity of Ddl6 and Ddl7 was confirmed using purified proteins (Fig. 3A). The release of the inorganic phosphate (Pi) in the Ddl-catalysed reaction was assayed and both variants of Ddl catalyze the release of a significant amount of Pi into the reaction compared to the control ($P > 0.05$) (Fig. 3B). This indicates that ATP is consumed in the reaction and Pi is released. Furthermore, paper chromatography analysis of the Ddl-catalysed reaction product showed that both proteins catalyze

the ligation of D-ala resulting in the synthesis of D-ala-D-ala dipeptide (Fig. 3C). The substrate specificity assay indicated that they could only catalyse the synthesis of D-ala-D-ala dipeptide, and not D-ala-D-ser or D-ala-D-lac.

The predicted 3D structures of Ddl6 and Ddl7 provided information on alteration of resistance phenotype and substrate specificity. To determine if the alteration of Trp259 of Ddl6 to Cys259 of Ddl7 caused a conformational change, 3D models of both Ddl6 and Ddl7 were constructed using I-TASSER²² (Supplementary Fig. S2). When the structures were superimposed using TM-align²³, they superimpose near perfectly with a TM score of 0.97. However, a notable predicted change in the conformation of the omega-loop region was observed likely resulting from the substitutions at 164 and 259 positions of Ddl6 and Ddl7 (Supplementary Fig. S3).

Three putative domains of Ddl6 and Ddl7 were identified by comparing predicted 3D structures of Ddl6 and Ddl7 with the crystal structure of D-ala-D-ser ligase (VanG, PDB code: 4FU0)²⁴, DdlB of *E. coli*²⁵, VanA of *E. faecium*²⁶ and Ddl of *S. aureus*²⁷. The predicted N-terminal domain of Ddl6 and Ddl7 runs from the N-terminus to Gly125, central domain from Ser126 to Gly219 and C-terminal domain from Arg220-Ile340. The putative omega-loop, which plays an important role in catalytic activity and substrate specificity of Ddl proteins, was located on the C-terminal domain of Ddl6 and Ddl7 and runs from Lys240-Ser255. The Trp259 of Ddl6 whose substitution with Cys259 causes a fourfold increase in D-cycloserine resistance was found to be located at the end of the 4th β -sheet (residues Asn256-Ile258) of the C-terminal domain and outside of the omega-loop (Supplementary Fig. S2). The amino acid residues (Ser150 and Tyr216) of *E. coli* DdlB which are conserved in the Ddls that can only produce D-ala-D-ala residues were also conserved in Ddl6 and Ddl7 with the corresponding residues of Ser184 and Tyr250 (Supplementary Fig. S4). This supports the results of the in-vitro assays and explains why it could produce only the D-ala-D-ala dipeptide, not D-ala-D-ser or D-ala-D-lac.

Although the integron-encoded Ddls were not found to confer vancomycin resistance and did not carry the amino acid residues for altered substrate specificity (D-Ser or D-Lac), among the closest homologues in PDB that matched with the predicted 3D structures of Ddl6 and Ddl7, four were related to vancomycin resistance including VanG (PDB code: 4FU0) and VanA (PDB code: 1E4E) with high TM scores (Supplementary Table S1).

Molecular dynamics simulations post-analysis provides insights into the mechanism of alteration of MIC of D-cycloserine due to W259C mutation in Ddl6. The results of 50 ns molecular dynamics (MD) simulations of wild-type Ddl6 and its W259C mutant derivative in complex with D-cycloserine (Fig. 4) showed that the ligand is released from the recombinant protein after approximately 20 ns, however, the complex with the wild-type protein remains stable in the binding site during the course of the MD simulations.

The movement of the ligand from the protein was studied by monitoring the distances (Supplementary Fig. S4) between the centre of D-cycloserine and Trp259 of the wild-type Ddl6 and Cys259 of the Ddl6 W259C mutant during the course of the simulation. The distance between the centre of D-cycloserine and Cys259 of the mutant protein progressively increased suggesting a movement away from the binding site, whereas the distance between the centre of D-cycloserine and Trp259 of the wild-type protein remained relatively stable throughout the simulation (Supplementary Fig. S5).

The relative binding free energy values calculated from the MD simulations trajectories further supported this observation as the complex of wild-type Ddl6 and D-cycloserine was more favorable (-14.47 kcal/mol) than the complex of D-cycloserine and Ddl6 W259C mutant complex (-12.15 kcal/mol) (Supplementary Table S2). Videos of the MD simulations showing the movement of the ligands are available in Supplementary Video S1 and Video S2.

Kinetic analysis of pure proteins showed that Ddl7 is more active than Ddl6. The integron-encoded Ddl6 and Ddl7 were characterised kinetically and the kinetic parameters were compared with DdlTd (the Ddl encoded by housekeeping *ddl* of *T. denticola*) and DdlAec as positive control. The $K_{m,D-ala2}$ of Ddl6 and Ddl7 was determined as 0.39 and 0.54 mM, respectively (Table 2), thus, the values of $K_{m,D-ala2}$ could not explain the alteration of D-cycloserine phenotype between these two variants. However, the turnover number (K_{cat}) of Ddl7 (3.85 s⁻¹) was approximately three-fold higher than Ddl6 (1.20 s⁻¹), which nearly matches the fourfold increase of MIC of D-cycloserine for Ddl7 as compared to Ddl6.

Molecular docking analysis showed that flavonoids could be responsible for selection of the *ddl6/ddl7* within the first position of the integron. We also tried to identify the likely selection pressure that could drive the evolution of *dcls* within integrons and maintain them as the first GC. As the use of D-cycloserine is restricted only for the MDR and XDR *Mycobacterium tuberculosis*, it is very unlikely to be responsible for this. We assumed that this selective pressure is common for both countries from which *ddl* containing strains were isolated (Bangladesh and the UK), as the *dcls* were ubiquitous within samples from both geographic locations. Candidates for providing this selective pressure are two plant flavonoids, quercetin and apigenin, which are abundantly present in many teas, fruits and vegetables and inhibit Ddl²⁹.

Molecular docking was used to see if these flavonoids can bind to Ddl within the predicted D-alanine and/or ATP binding sites of Ddl6 and Ddl7. Salvicine (a terpenoid isolated from a Chinese medicinal plant, *Salvia prionitis*) was used as a positive control for the docking because this terpenoid was previously identified as a potent inhibitor of Ddl of *T. pallidum* using molecular docking³⁰. GOLD molecular docking showed that quercetin and apigenin can dock to both D-alanine (Fig. 5; Supplementary Table S3) and ATP binding sites (Supplementary Fig. S6; Supplementary Table S3) of Ddl6 and its mutants therefore, the integron-encoded Ddls may play a role in both the sequestration of these inhibitors, lowering the intracellular concentration and allowing the house-keeping Ddl to perform its physiological functions and also take on the role of the house-keeping Ddl during cell wall development.

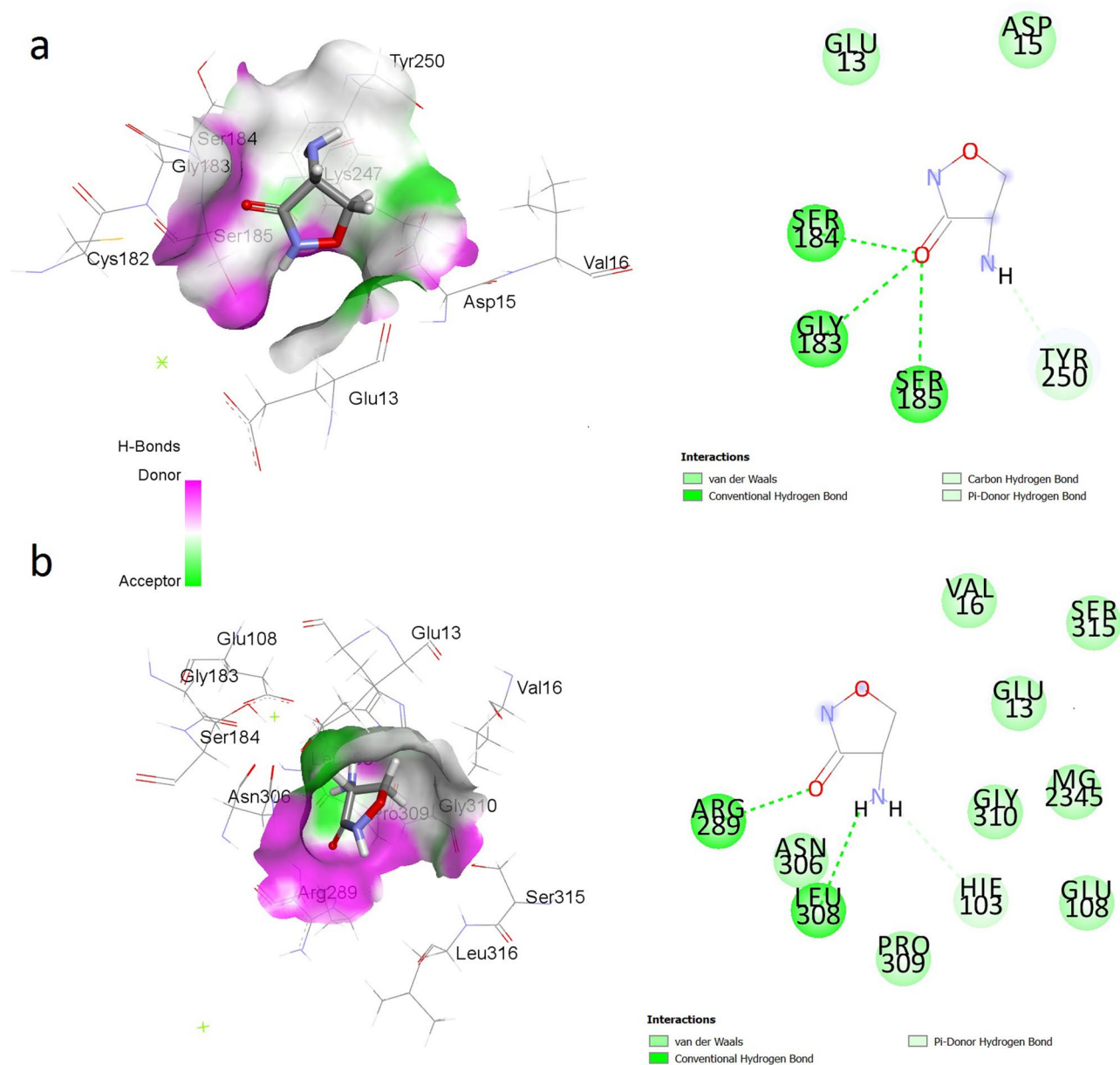


Figure 4. 2D (right panels) and 3D (left panels) structures of the binding sites of the W259C mutant of Ddl6 (a) and wild-type Ddl6 (b) in interaction with D-cycloserine in average structures extracted from MD trajectories. After placing in the same location at the beginning of the MD simulations, both ligands were found in the adjacent grooves in the same binding site of their corresponding average structures.

Proteins	$K_{m,ATP}$ (μ M)	K_{cat} (s^{-1})	$K_{m,D\text{-}ala2}$ (mM)	$K_{cat}/K_{m,D\text{-}ala2}$ ($M^{-1} s^{-1}$)
Ddl6	54.96	1.19	0.39	3008.85
Ddl7	43.28	3.85	0.55	7010.32
DdlTd	19.08	3.74	0.45	8294.55
DdlAEc	48.99	6.43	0.52	12,517.21
DdlAEc*	38*	NR	0.55*	NR

Table 2. Steady-state kinetic parameters of Ddl6, Ddl7, DdlTd and DdlEc. *The values of DdlAEc reported by Zawadzke, et al.²⁸. NR not reported.

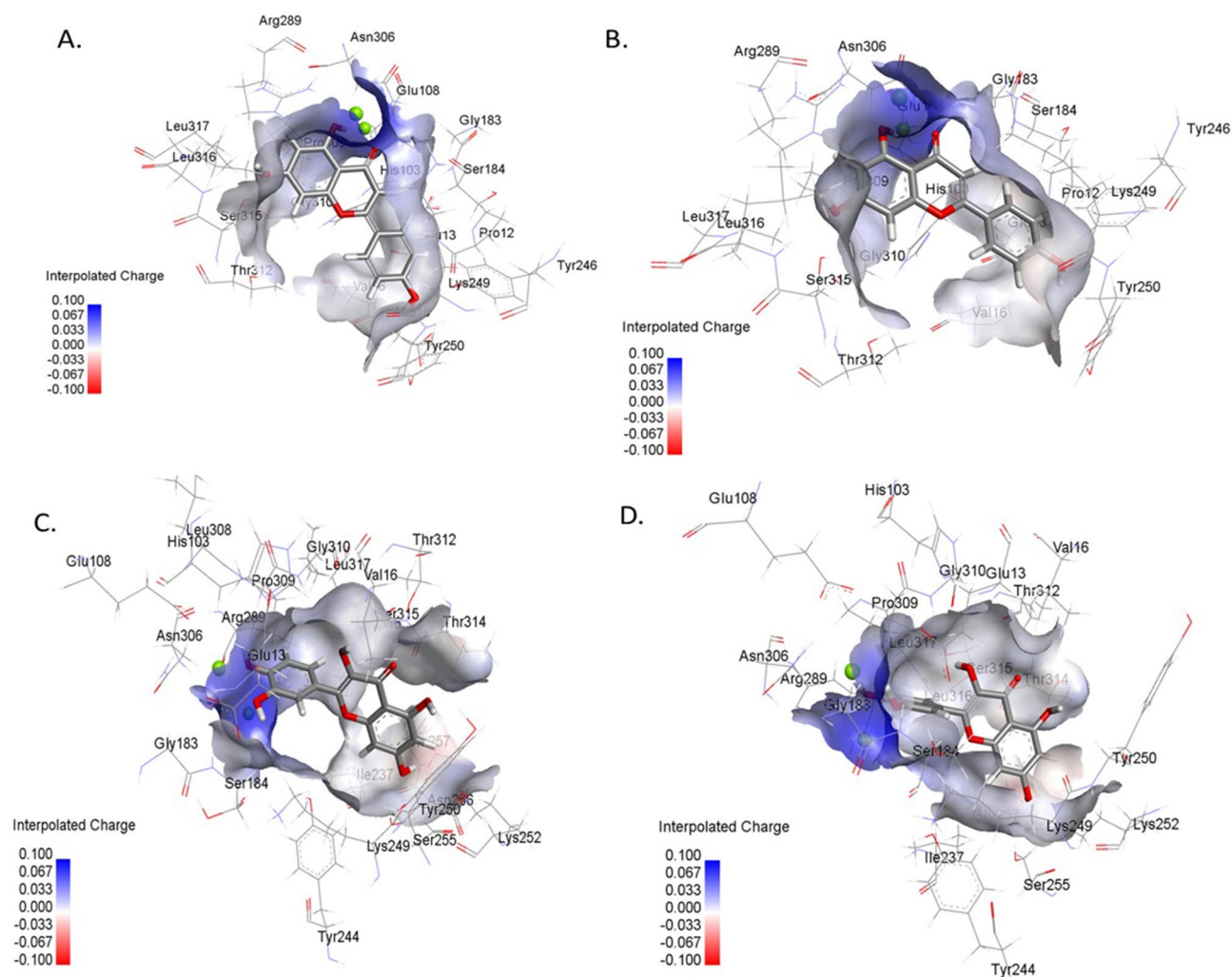


Figure 5. Mode of interaction of apigenin and quercetin with the D-alanine binding sites of wild-type Ddl6 and its W259C mutant. (A) Apigenin with wild-type Ddl6, (B) Apigenin with W259C mutant, (C) Quercetin with wild-type Ddl6 and (D) Quercetin with W259C mutant.

Discussion

This study reports for the first time the discovery of integron GCs encoding two novel isoforms of Ddl (Ddl6 and Ddl7) derived from the oral cavity microbiota of healthy humans in the UK and Bangladesh. The ORFs encoding Ddl were located immediately downstream of *intI* on reverse integrons. All of the genetic characteristics of an integron and its associated GCs were found within the amplicon, including the presence of a putative *Pc*, an *attI* region harbouring putative integrase binding sites (S1 and S2), a putative *attI/attC* recombination site, a putative RBS located 6-bp upstream of the start codon of the *ddIs* as well *attC* core sites (R" and L")^{6,7,10,31}.

The closest homologue of *ddl6* and *ddl7* with approximately 98.0% nucleotide identity was located on a whole genome sequence (WGS) contig of *T. pedis* B683 and is one of two copies of *ddl* within this genome and is not present in the genome of other *T. pedis* strains, thus it is likely part of the accessory genome. *T. pedis* is an animal pathogen that causes digital dermatitis in cattle and necrotic ulcers in pigs³². B683 (GenBank (NZ_AOTN00000000.1) was isolated from a shoulder ulcer of a pig. The evidence showed that there is a likely genetic exchange between *T. pedis* and the host of *ddl6/ddl7*, likely to be *T. denticola*, which may involve one or more transfer intermediates.

Analysis of the upstream sequence of the integron carrying *ddl6/ddl7* suggests that the likely host of the integron is a strain of *T. denticola* which carries two copies of *ddl* in its genome: one as an integron GC and another as house-keeping gene (this house-keeping *ddl* of *T. denticola* ATCC35404 showed only 46.0% nucleotide identity to *ddl6* and *ddl7*). *T. denticola* is one of the most comprehensively characterised species among the 49 species of *Treponema* in the human oral cavity³³. It is one of the members of the 'red complex' associated with clinical progression of periodontitis³⁴ and is capable of transferring genes to different genera of bacteria in dental plaque³⁵. It is the only bacterial species in the oral cavity that has been found to carry integrons¹¹ although we have recently demonstrated other bacterial species are also likely to contain integrons³⁶. From the in silico analysis of the metagenomic datasets of the NIH Human microbiome project, a total of 826 gene cassettes related to *Treponema*

sp. were detected in the oral cavity³⁷. The closely related strains of *T. denticola* shared a high nucleotide identity (>80%) of their integron gene cassettes³⁸.

Our results provide novel insights into the evolutionary mechanisms that have facilitated the acquisition of a second copy of *ddl* on an integron in the original host of *ddl6* and *ddl7*. Zawadzke et al.²⁸ proposed multiple explanations for the duplication of *ddl* seen in *E. coli* however these are likely not applicable here because, unlike the duplicated *ddlA* and *ddlB* genes of *E. coli*, the *ddl* cassettes described here are likely to have been acquired by HGT and thus, they may be related to gaining a new function (resistance to antimicrobial compounds) in addition to, or rather than, providing increased amount of the gene product^{28,39}. The kinetic analysis of the purified proteins (Ddl6, Ddl7 and DdlTd) also indicates that the *ddls* are interchangeable with respect to their core function of cell wall assembly because, the K_{cat} (s^{-1}) of DdlTd was very close to Ddl7 thus, the acquisition of a second copy of *ddl* within integrons provide resistance by titrating out intracellular antimicrobials plus complementation of the original Ddl function.

We found that the ectopic expression of *ddl6* and *ddl7* confers resistance to D-cycloserine. As Ddl is a target of D-cycloserine, overproduction of Ddl is likely to facilitate resistance by mechanisms similar to D-cycloserine resistance in *Mycobacterium* which is mediated by high expression of both Ddl and Alr^{21,40}. The other known mechanisms of D-cycloserine resistance in *Mycobacterium* were the non-synonymous mutations in Alr including S22L, K133E, L89R, M319T, Y364D and R373G^{41,42} as well as loss of function mutation in *alr*⁴². In *E. coli* the mutation in *cycA*⁴³, *dadA* and *pnp*⁴⁴ confer resistance to D-cycloserine. The mutations in a penicillin binding protein (PBP4) homologue in *M. smegmatis* confer resistance to D-cycloserine and vancomycin⁴⁵. Although several mutations in *alr* have been implicated in D-cycloserine resistance, so far no mutations in *ddl* were found to be associated with alteration of D-cycloserine resistance by the host. Moreover, several functional metagenomic analyses reported that ectopic expression of various *ddl* also conferred resistance to D-cycloserine^{46–49}.

While the expression of both integron encoded *ddls* conferred resistance to D-cycloserine, the expression of *ddl7* had a four-fold higher MIC compared to *ddl6*. Site-directed mutagenesis confirmed that a single substitution at c.777, which alters Trp259 of Ddl6 to Cys259, was responsible for this. Using MD simulations, we have shown how this mutation is predicted to alter the binding affinity of the ligand D-cycloserine to the active sites of the proteins. The wild-type Ddl6 with a tryptophan residue at its 259-position had more stable binding to D-cycloserine, therefore less D-cycloserine would be required to occupy active sites which in turn will decrease the MIC. When the aromatic, non-polar, hydrophobic tryptophan residue was changed to polar, thiol reactive⁵⁰ and disulphide bridge forming⁵¹ cysteine (the residue at 259 for Ddl7), binding between the mutant Ddl6 and D-cycloserine complex was comparatively less stable and thus, more antibiotic will be needed to occupy the active sites which causes an increase in the MIC.

Four of the closest structural homologues of Ddl6/Ddl7 were related to vancomycin resistance including VanA of *E. faecium*, and VanG of *E. faecalis*. It is known that the function of a protein is more related to the structure rather than the sequence⁵². The inclination of the predicted 3D structures of Ddl6 and Ddl7 towards vancomycin resistant determinants suggests that if the corresponding residues of VanG or VanA, which are responsible for binding and positioning D-ser (VanG-type) or D-lac (VanA-type) instead of D-ala2, are substituted to the corresponding residue of Ddl6 and Ddl7 (Tyr250), they could gain D-ala-D-ser dipeptide or D-ala-D-lac depsipeptide ligase activity. The *E. coli* DdlB gained a weak D-ala-D-lac depsipeptide activity following Tyr216 and Ser150 substitutions with phenylalanine and alanine of LmDdl2⁵³, respectively supporting our hypothesis.

We also suggest a hypothesis for evolution of *ddl* within an integron cassette at the first position. D-Cycloserine is currently only used in the treatment of MDR and XDR-tuberculosis⁵⁴ and there is no history of environmental use of D-cycloserine in agriculture. Additionally, none of the volunteers received this drug (or any other antibiotic), at least in the three months prior to saliva collection. Therefore, it is very unlikely that D-cycloserine represents the driving force for this event. We hypothesize that dietary flavonoids or other inhibitory small molecules present in our diet that can bind Ddl drive the acquisition and selection of *ddl* within integrons. Some flavonoids including apigenin, quercetin and catechin are known to have antibacterial activity^{55,56} and are found in fruits and vegetables as well as widely consumed beverages such as tea. Apigenin and quercetin has been shown to inhibit the activity of Ddl of *H. pylori* by competitively inhibiting the binding of ATP to ATP-binding sites^{29,57}. MD analysis performed in this study supported the findings and demonstrated that these flavonoids are likely to bind to both D-ala and ATP binding sites of Ddl6 and Ddl7. Thus, the excess Ddl produced by the efficient expression under the control of *Pc* could confer a selective advantage on the host titrating out inhibitory flavonoids (Fig. 6). This hypothesis requires testing experimentally which is now under way.

The results show that *ddl*, which is generally considered a housekeeping gene has moved onto an integron and can be found in the first position, with the most likely selection pressure being dietary compounds with antimicrobial activity. Because these genes also confer resistance to D-cycloserine and are potentially one mutation away from vancomycin resistance, the study also sheds light on potential environmental drivers of clinically relevant antimicrobial resistance.

Materials and methods

Source of chemicals, reagents and antibiotics. Chemicals, solvents, culture media and antibiotics were purchased from Sigma-Aldrich Ltd (Dorset, UK), BDH (UK) and Life Technologies (USA) unless stated otherwise. Plasmid preparation kits were obtained from Qiagen (UK). Restriction enzymes were obtained either from New England Biolabs (UK) Ltd or from Promega (UK). The EnzChek Phosphate assay kit and Pierce BCA Protein assay kits were purchased from Life Technologies (USA). *E. coli* α -Select competent cells were obtained from Bioline Reagents Ltd (UK). The primers were obtained from Sigma-Aldrich (UK).

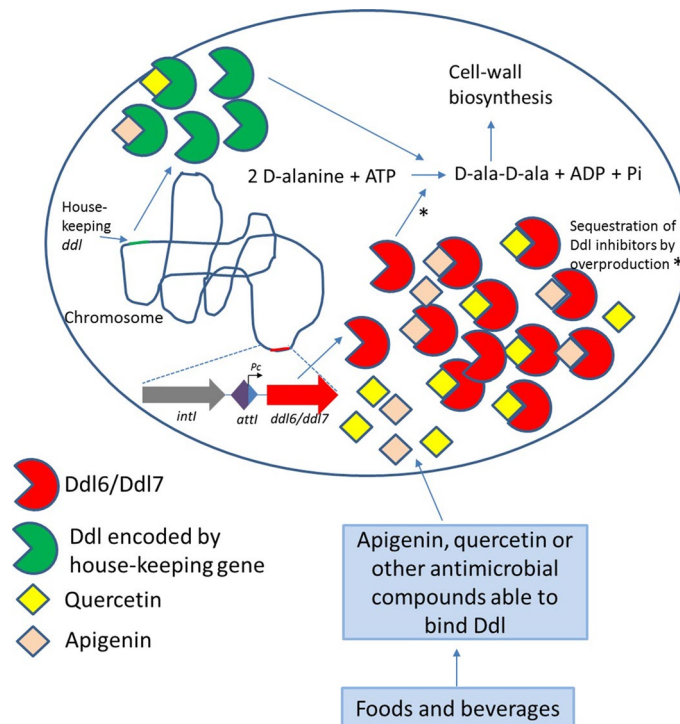


Figure 6. Proposed model for plant flavonoids promoting the selection of *ddls* and their maintenance within the first GC of the integron of *T. denticola*. The integron encoded Ddl6/Ddl7 could play a dual role (marked with asterisk): sequestration of flavonoids or other molecules that are able to bind Ddl and catalysing the formation of D-alanyl-D-alanine dipeptide in presence of inhibited Ddl encoded by the housekeeping genes.

Bacterial strains and plasmids. The pGEM-TEasy vector (Promega) was used for constructing the metagenomic library of PCR amplicons of integrons and gene cassettes. *E. coli* α -select silver efficiency (Bioline) was used to manipulate the pGEM-TEasy library of gene cassettes. pET28a (Novagen) and *E. coli* BL21 (DE3) (New England Biolabs) were used for cloning and expression of *ddls* and purification of corresponding His-tagged proteins, respectively. *Bacillus subtilis* 168 and pHCMC05 (Bacillus Genetic Stock Centre) were used for cloning and expression of *ddl6* and *ddl7* to determine the susceptibility of D-cycloserine and vancomycin in Gram-positive bacteria. *E. coli* strains were cultured in LB medium at 37 °C. *Bacillus subtilis* 168 was cultured in Brain–heart infusion (BHI) medium unless stated otherwise. The bacterial strains were preserved at – 80 °C in 30% (v/v) sterile glycerol.

Detection of *ddl* gene cassettes in the metagenomic DNA obtained from saliva. The collection of saliva samples, preparation of metagenomic DNA from pooled saliva from UK and Bangladesh and detection of integrons and gene cassettes in the library PCR amplicons have been described previously³⁶. Primers for reverse integrons were utilized as PCR primers to detect integrons failed to yield amplicons from metagenomic DNA³⁶. Ethical approval was obtained from University College London (UCL) Ethics Committee (project number 5017/001) and the Institutional Animal, Medical Ethics, Biosafety and Biosecurity Committee (IAMEBBC) for Experimentations on Animal, Human, Microbes and Living Natural Sources, University of Rajshahi (project number 54/320/IAMEBBC/IBSC). All volunteers provided informed consent and all procedures were carried out according to relevant guidelines and regulations. Briefly, the PCR products produced using the forward integrase primer (TDIF) and reverse *attC*-based primer (MARS-2) were cloned into pGEM-TEasy and transformed into *E. coli* alpha select cells. Transformants were randomly picked, plasmids were prepared from 5 mL overnight culture using Miniprep kits (Qiagen) and the plasmids were sequenced using M13F-M13R primers (Supplementary Table S4).

Determination of the upstream region of integron carrying *ddl6* and *ddl7*. A total of four different forward primers (Upint_5100F, Upint_3685F, Upint_721F, Upint_122F) (Supplementary Table S4) were designed randomly at different locations upstream of the *intI* gene of the chromosome of *T. denticola* ATCC35405 (NC_002967.9). PCR was performed by combining these primers with the reverse primer, ddLR designed from the 3'-end of *ddl6/ddl7*. It was assumed that if the integrons carrying *ddl* GCs are located on the chromosome of *T. denticola*, the PCR products will yield a different length of upstream region of *intI*. Q5 High-Fidelity 2X Master Mix (NEB, UK) was used for PCR using the pooled metagenomic DNA of UK saliva samples as template. The reaction condition for PCR was as follows: an initial denaturing temperature of 98 °C for 30 s followed by 30 cycles at 98 °C for 10 s (denaturing), 50–58 °C for 30 s (annealing), and 72 °C for 200 s (extension). The PCR run

ended with a final 72 °C elongation step for 10 min before holding the reaction at 4 °C. The expected size PCR products were purified by gel extraction using QIAquick Gel Extraction Kit (Qiagen, Germany) and A-tailed following the protocol for cloning blunt-ended PCR products described in pGEM-T Easy Vector Systems Technical Manual (Promega, UK). A 30 µL reaction mixture for A-tailing contained about 30 ng purified PCR product, 3 µL of 10× standard *Taq* reaction buffer, 1.5 µL 100 mM dATP, 2 µL *Taq* DNA polymerase (NEB, UK) and molecular grade water up to 30 µL. The reaction mixture was then incubated for 30 min at 72 °C in a block of PCR machine. An aliquot of A-tailed PCR products was cloned into pGEM-T Easy and then transformed into *E. coli* alpha select cells. The transformants were selected on LB agar plates supplemented with ampicillin, IPTG and X-gal. Plasmids were prepared and the inserts were sequenced.

Site-directed mutagenesis. The Phusion Site-Directed Mutagenesis Kit (Thermo Scientific, USA) was used to construct two different mutants of *ddl6*: *ddl6* c.490 C>T and *ddl6* c.777 G>T using the pGEM-TEasy::*int1-att1-ddl6* plasmid as template. The phosphorylated mutagenic primers used for constructing the mutants are listed in the Supplementary Table S4. After confirming the success of the mutagenesis reactions, the mutant plasmids were transformed to *E. coli* α-select cells and MIC of the transformants were determined using the agar dilution method.

Sub-cloning of *ddl6* and *ddl7* into pHCMC05 and transformation into *B. subtilis*. The genes for *Ddl6* and *Ddl7* were amplified by PCR using the pGEM-TEasy vectors with the 2024-bp insert containing *ddl6* and *ddl7* as template (see Supplementary Table S4 for primers used). The genes were cloned into *Bam*HI and *Xba*I sites of pHCMC05 (*E. coli*-*B. subtilis* shuttle vector) using the primers *TddlF* and *TddlR* (Supplementary Table S4). Chemically competent *B. subtilis* 168 cells were prepared and transformed with the recombinant pHCMC05 vectors using the method described by Hardy⁵⁸. Transformants were spread onto BHI plates supplemented with 10 µg/mL of chloramphenicol and incubated at 37 °C for 16–18 h.

Cloning and expression of *ddl6* and *ddl7*. The coding sequence of the genes for *Ddl6* and *Ddl7* (1032-bp) were amplified and cloned into the *Bam*HI and *Xho*I sites of pET28a vector (Novagen) using the primers *TddlF* and *Tddl28aR* (Supplementary Table S4). The recombinant plasmids were transformed into *E. coli* BL21 (DE3) cells (New England Biolabs) and transformants were selected on LB agar supplemented with kanamycin (30 µg/mL). 1L LB was inoculated with 10 ml overnight culture, and expression was induced in exponential growth phase by addition of 0.5 mM isopropyl β-D-1-thiogalactopyranoside for 3 h at 37 °C. Cells were harvested by centrifugation at 15,000×g at 4 °C and stored at –20 °C until protein purification.

Purification of N- and C-terminal His-tagged *Ddl6* and *Ddl7* proteins by FPLC. Preserved cells were thawed on ice for 30 min and resuspended in buffer A (50 mM NaH₂PO₄, 300 mM NaCl, 10 mM imidazole, pH 7.4) containing cOmplete EDTA-free protein inhibitor cocktail (Roche, 1 tablet/10 mL), lysozyme (1.0 mg/mL) and DNase (1 U/µL). The suspension was incubated on ice for 30 min and sonicated (6 pulse, each 10 s). Cellular debris was removed by centrifugation at 15,000×g for 30 min at 4 °C. Recombinant, N- and C-terminally His-tagged proteins were purified using a Ni-NTA Superflow cartridge (1 mL, Qiagen) and the Fast Protein Liquid Chromatography (FPLC) system (BioRad). The proteins were eluted with filter sterilized buffer B (50 mM NaH₂PO₄, 300 mM NaCl, 250 mM imidazole, pH 7.4). Protein concentration was determined using Pierce BCA Protein Assay Kit (ThermoFisher Scientific) and purity was checked by sodium-dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE). The gels were stained with Coomassie Blue (BioRad) and destained with destaining reagent (10% acetic acid and 20% methanol).

Enzymatic activity assay. The D-ala-D-ala ligase activity of purified Ddl enzymes was assayed as described previously with some modifications^{59,60}: (1) *Formation of d-ala-d-ala dipeptide*: The reaction mixture containing the following components were incubated for 4 h at 37 °C: 50 mM Tris-HCl, 10 mM MgCl₂, 10 mM KCl, 20 mM D-alanine, 2 mM ATP, 2.5 mM Glutathione and 20 µg/mL of purified Ddl. To stop the reaction, the mixtures were boiled for 5 min at 95 °C. To determine the substrate specificity, an equimolar amount of D-Ser or D-Lac was substituted with D-Ala. Next, 4 µL of the reactions were loaded onto cellulose chromatography paper (Whatman) and ascending chromatography was developed for 3 h in butanol-acetic acid-water (12:3:5 vol/vol/vol) containing 1.0% ninhydrin. The paper was dried for 5 min at 95 °C. D-ala-D-ala dipeptide formation in the reaction catalysed by Ddl was identified by comparison with commercially available D-ala-D-ala (Sigma). (2) *Measurement of inorganic phosphate in the Ddl-catalysed reaction*: The enzymatic activity of Ddl was also assayed by measuring the release of inorganic phosphate in the reaction (similar to the D-ala-D-ala dipeptide formation assay) with EnzChek phosphate assay kit (Life Technologies). 10 µL of the reaction product was added to 790 µL of standard reaction mixture containing 230 µL dH₂O, 50 µL 20X reaction buffer, 200 µL 2-amino-6-mercapto-7-methylpurine riboside (MESG) substrate solution and 10 µL purine nucleoside phosphorylase and incubated at 22 °C for 30 min. After incubation absorbance of the reactions was measured at 360 nm with a spectrophotometer (Ultrospec 2000, Pharmacia Biotech). The amount of inorganic phosphate released into the reactions was calculated from the standard curve of inorganic phosphate.

Kinetic analysis of *Ddl6* and *Ddl7*, *DdlAec* and *DdlTd*. The V_{max} and $K_{m, D-ala2}$ of *Ddl6*, *Ddl7*, *DdlTd* and *DdlAec* were determined by using pyruvate kinase (PK)/lactate dehydrogenase (LDH)-coupled spectrophotometric system^{60,61}.

All assays were conducted at 37 °C in 96-well plate system using CLARIOstar (BMG Labtech, Germany) in 200 µL reaction buffer. The assay mixture (200µL) contained 100 mM HEPES (pH 7.5), 10 mM MgCl₂, 10 mM KCl, 0.2 mM NADH, 6–10 U/mL PK, 9–14 U/mL LDH, 2 mM phosphoenolpyruvate (PEP) and variable concentrations of ATP and D-alanine (saturating concentrations were 500 µM and 100 mM for ATP and D-alanine, respectively). The reactions were initiated by adding Ddl enzymes at a final concentration of 0.0025 µg/µL. The initial velocity (v_0) of the reactions was determined by measuring the changes in the absorbance at 340 nm for 500 s (8.3 min). The slope of the reaction progression curves (time (s) vs change in absorbance at 340 nm) is equivalent to the v_0 ($v_0 = \text{slope} = \Delta Y/\Delta X$)⁶². The slope was calculated by the MARS data analysis software (BMG Labtech, Germany). After the v_0 values for each substrate concentration (ATP and D-alanine) were obtained, GraphPad Prism version 7 (GraphPad Software, Inc., USA) was used to determine the V_{max} and K_m values by fitting the curves for Michaelis–Menten model.

The values for $K_{m,\text{ATP}}$ of different Ddls were determined at a fixed saturating concentration of D-alanine (100 mM) and varying concentrations of ATP starting from 5 µM to 320 µM. The K_m values for second D-alanine binding ($K_{m,\text{D-ala2}}$) was determined by using a fixed saturating concentration of ATP (500 µM; approximately 10 times of $K_{m,\text{ATP}}$). The values of K_{cat} (s⁻¹) for all enzymes were determined by dividing the V_{max} (µM/s) with the concentration of enzymes, [E] (µM).

Phylogenetic analysis. The phylogenetic tree was constructed using MEGA X software¹⁹. The sequences of Ddl6 and Ddl7 and their homologs were aligned with Clustal Omega. The default settings for Neighbour-Joining method in MEGA X were applied to construct the phylogenetic tree.

Homology modelling of Ddl6 and Ddl7. The homology models of Ddl6 and Ddl7 were constructed using I-Tasser server (<https://zhanglab.cmb.med.umich.edu/I-TASSER>)²². The amino acid sequences were submitted to the server and among the top five models built by the server, model 1 was selected based on the highest C-score and TM-score. A higher value of C-score signifies a model with a high confidence. A TM-score > 0.5 indicates a model of correct topology and a TM-score < 0.17 means a random similarity⁶³. Models were visualised with PyMOL (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC). The homology models of Ddl6 and Ddl7 were matched with each other with the TM-align structural alignment program (<https://zhanglab.cmb.med.umich.edu/TM-align>) and the overlapped structures were evaluated with PyMOL.

Determination of minimum inhibitory concentrations (MIC):. The MIC of D-cycloserine, vancomycin and beta-lactams were determined by agar dilution method following the guidelines of Clinical and Laboratory Standards Institute (CLSI) published in 2012. Three individual colonies of *E. coli* harbouring pET-*ddl6* or pET-*ddl7* and *B. subtilis* 168 strains harbouring pHCMC05-*ddl6* or pHCMC05-*ddl7* were inoculated in Mueller–Hinton (MH) broth supplemented with appropriate antibiotics and cultured at 37 °C for 4–5 h and the OD₆₀₀ was adjusted to 0.08 to 0.1 using sterile saline. MH agar plates supplemented with various concentration of antibiotics starting from 0.125 to 256 µg/mL were prepared and inoculated with the adjusted bacterial suspensions with a multipoint inoculator. The spots were allowed to dry for 30 min and the plates were then incubated aerobically at 37 °C for 16–20 h. The MIC value was defined as the lowest concentration of the inhibitory compounds giving rise to no visible growth. MIC determinations were performed at least three times in triplicates. As a control, *E. coli* or *B. subtilis* 168 strains transformed with empty vectors were used.

Molecular docking. The predicted 3D-structures of Ddl6 and its two single mutants, L164F and W259C, as well as Ddl encoded by house-keeping gene of *T. denticola* (DdlTd) were used as the targets for molecular docking analyses. The 3D structure of Ddl6 was applied in PyMol program to generate L164F and W259C mutants. These protein targets were used for molecular docking of six ligands including D-alanine, D-cycloserine, ATP, apigenin, quercetin and salvicine, after minimisation and equilibration using AMBER program. AutoDock SMINA⁶⁴ was used to determine the best binding pocket by exploring all probable binding cavities in the proteins. Then GOLD (Genetic Optimization for Ligand Docking)^{65,66} was used for docking the ligands into the SMINA-located binding site to perform flexible molecular docking analyses.

Genetic algorithm (GA) was used in GOLD ligand docking to thoroughly examine the ligand conformational flexibility along with partial flexibility of the protein⁶⁷. The maximum number of runs for the ligand was set to 20 and in each run a population size of 100 with 100,000 operations was employed. The number of islands was 5, and the niche size of 2 was considered. The default cut-off values of hydrogen bonds was set to 2.5 Å (dH-X), and for the van-der-Waals distance it was 4.0 Å. The GA docking was terminated when the top solutions attained the root mean square deviation (RMSD) values within 1.5 Å.

Molecular dynamics (MD) simulations. After molecular docking, the best poses of D-cycloserine in the D-Ala binding site of wild-type and W259C mutant of Ddl6 were selected as the starting structures to run MD simulations for 50 ns. The MD simulations were carried out using the AMBER 16.0 software package⁶⁸. The force fields parameters for D-cycloserine were generated using the ANTECHAMBER module of the AMBER programme. Each system was solvated using an octahedral box of TIP3P water molecules. Periodic boundary conditions and particle-mesh Ewald (PME) method were employed in all the simulations⁶⁹. Particle-mesh Ewald method enabled us to calculate the 'infinite' electrostatics without truncating the parameters. During each simulation, all bonds in which the hydrogen atom was present were considered fixed, and all other bonds were constrained to their equilibrium values by applying the SHAKE algorithm⁷⁰. A cutoff radius of 12 Å was used for the systems. Minimization was performed in two phases, and each phase was performed in two stages. In the first phase, ions and all water molecules were minimized for 1500 cycles of steepest descent followed by 1500

cycles of conjugate gradient minimization. Afterwards, the systems were minimized for a total of 5000 cycles without restraint wherein 2500 cycles of steepest descent were followed by 2500 cycles of conjugate gradient minimization. After minimizations, the systems were heated for 1000 ps while the temperature was raised from 0 to 300 K, and then equilibration was performed without a restraint for 1000 ps while the temperature was kept at 300 K. Sampling of reasonable configurations was conducted by running a 50 ns simulation with a 2 fs time step at 300 K and 1 atm pressure. A constant temperature was maintained by applying the Langevin algorithm while the pressure was controlled by the isotropic position scaling protocol used in AMBER 16 package program⁶⁸.

Statistical analysis. Statistical analysis was performed by using GraphPad Prism (GraphPad Software, Inc., USA; version 7.0). Statistical significance was calculated using one-way ANOVA. *P* values of ≤ 0.05 were considered statistically significant.

GenBank accession numbers. The sequences of the 2024 bp PCR amplicons containing *ddl6* and *ddl7* were submitted separately to GenBank with the accession number of KU886208 and KU886209, respectively. The 4421 bp sequence of PCR product containing the upstream sequence of *intI* along with the downstream *ddl7* was submitted to GenBank with the accession number KY039278.

Received: 28 May 2020; Accepted: 10 November 2020

Published online: 26 November 2020

References

- Bush, K. *et al.* Tackling antibiotic resistance. *Nat. Rev. Micro.* **9**, 894–896 (2011).
- D’Costa, V. M., McGrann, K. M., Hughes, D. W. & Wright, G. D. Sampling the antibiotic resistome. *Science* **311**, 374–377. <https://doi.org/10.1126/science.1120800> (2006).
- Holmes, A. H. *et al.* Understanding the mechanisms and drivers of antimicrobial resistance. *Lancet* **387**, 176–187. [https://doi.org/10.1016/S0140-6736\(15\)00473-0](https://doi.org/10.1016/S0140-6736(15)00473-0) (2016).
- Labbate, M. *et al.* Use of chromosomal integron arrays as a phylogenetic typing system for *Vibrio cholerae* pandemic strains. *Microbiology* **153**, 1488–1498. <https://doi.org/10.1099/mic.0.2006/001065-0> (2007).
- Gillings, M. R. Integrons: past, present, and future. *Microbiol. Mol. Biol. Rev.* **78**, 257–277. <https://doi.org/10.1128/MMBR.00056-13> (2014).
- Collis, C. M. & Hall, R. M. Expression of antibiotic resistance genes in the integrated cassettes of integrons. *Antimicrob. Agents Chemother.* **39**, 155–162 (1995).
- Collis, C. M., Grammaticopoulos, G., Briton, J., Stokes, H. W. & Hall, R. M. Site-specific insertion of gene cassettes into integrons. *Mol. Microbiol.* **9**, 41–52 (1993).
- Recchia, G. D. & Hall, R. M. Gene cassettes: a new class of mobile element. *Microbiology* **141**(Pt 12), 3015–3027. <https://doi.org/10.1099/13500872-141-12-3015> (1995).
- Cambay, G., Guerout, A. M. & Mazel, D. Integrons. *Annu. Rev. Genet.* **44**, 141–166. <https://doi.org/10.1146/annurev-genet-102209-163504> (2010).
- Partridge, S. R., Tsafnat, G., Coiera, E. & Iredell, J. R. Gene cassettes and cassette arrays in mobile resistance integrons. *FEMS Microbiol. Rev.* **33**, 757–784. <https://doi.org/10.1111/j.1574-6976.2009.00175.x> (2009).
- Coleman, N., Tetu, S., Wilson, N. & Holmes, A. An unusual integron in *Treponema denticola*. *Microbiology* **150**, 3524–3526. <https://doi.org/10.1099/mic.0.27569-0> (2004).
- Neuhaus, F. C. The enzymatic synthesis of D-alanyl-D-alanine. *Biochem. Biophys. Res. Commun.* **3**, 401–405. [https://doi.org/10.1016/0006-291X\(60\)90053-X](https://doi.org/10.1016/0006-291X(60)90053-X) (1960).
- Škedelj, V. *et al.* 6-Arylpyrido[2,3-d]pyrimidines as novel ATP-competitive inhibitors of bacterial D-alanine:D-alanine ligase. *PLoS ONE* **7**, e39922. <https://doi.org/10.1371/journal.pone.0039922> (2012).
- Halouska, S. *et al.* Metabolomics analysis identifies d-alanine-d-alanine ligase as the primary lethal target of d-cycloserine in mycobacteria. *J. Proteome Res.* **13**, 1065–1076. <https://doi.org/10.1021/pr4010579> (2014).
- Tytgat, I. *et al.* DD-ligases as a potential target for antibiotics: past, present and future. *Curr. Med. Chem.* **16**, 2566–2580 (2009).
- Caminero, J. A., Sotgiu, G., Zumla, A. & Migliori, G. B. Best drug treatment for multidrug-resistant and extensively drug-resistant tuberculosis. *Lancet. Infect. Dis.* **10**, 621–629. [https://doi.org/10.1016/S1473-3099\(10\)70139-0](https://doi.org/10.1016/S1473-3099(10)70139-0) (2010).
- Anderson, R. J., Groundwater, P. W., Todd, A. & Worsley, A. J. *Antibacterial Agents* 319–326 (Wiley, Hoboken, 2012).
- Saitou, N. & Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425 (1987).
- Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549. <https://doi.org/10.1093/molbev/msy096> (2018).
- Guerin, É. *et al.* The SOS response controls integron recombination. *Science* **324**, 1034–1034. <https://doi.org/10.1126/science.1172914> (2009).
- Feng, Z. & Barletta, R. G. Roles of *Mycobacterium smegmatis* D-alanine:D-alanine ligase and D-alanine racemase in the mechanisms of action of and resistance to the peptidoglycan inhibitor D-cycloserine. *Antimicrob. Agents Chemother.* **47**, 283–291 (2003).
- Yang, J. & Zhang, Y. I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res.* **43**, W174–181. <https://doi.org/10.1093/nar/gkv342> (2015).
- Zhang, Y. & Skolnick, J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* **33**, 2302–2309. <https://doi.org/10.1093/nar/gki524> (2005).
- Meziane-Cherif, D., Saul, F. A., Haouz, A. & Courvalin, P. Structural and functional characterization of VanG D-Ala-D-Ser ligase associated with vancomycin resistance in *Enterococcus faecalis*. *J. Biol. Chem.* **287**, 37583–37592. <https://doi.org/10.1074/jbc.M112.405522> (2012).
- Fan, C., Moews, P. C., Walsh, C. T. & Knox, J. R. Vancomycin resistance: structure of D-alanine:D-alanine ligase at 2.3 Å resolution. *Science* **266**, 439–443 (1994).
- Roper, D. I., Huyton, T., Vagin, A. & Dodson, G. The molecular basis of vancomycin resistance in clinically relevant Enterococci: crystal structure of d-alanyl-d-lactate ligase (VanA). *Proc. Natl. Acad. Sci. USA.* **97**, 8921–8925 (2000).
- Liu, S. *et al.* Allosteric inhibition of *Staphylococcus aureus* d-alanine:d-alanine ligase revealed by crystallographic studies. *Proc. Natl. Acad. Sci. USA.* **103**, 15178–15183. <https://doi.org/10.1073/pnas.0604905103> (2006).

28. Zawadzke, L. E., Bugg, T. D. & Walsh, C. T. Existence of two D-alanine:D-alanine ligases in *Escherichia coli*: cloning and sequencing of the *ddlA* gene and purification and characterization of the *DdlA* and *DdlB* enzymes. *Biochemistry* **30**, 1673–1682 (1991).
29. Wu, D. *et al.* D-Alanine:D-alanine ligase as a new target for the flavonoids quercetin and apigenin. *Int. J. Antimicrob. Agents* **32**, 421–426. <https://doi.org/10.1016/j.ijantimicag.2008.06.010> (2008).
30. Dwivedi, U. N. *et al.* *Treponema pallidum* putative novel drug target identification and validation: rethinking syphilis therapeutics with plant-derived terpenoids. *OMICS* **19**, 104–114. <https://doi.org/10.1089/omi.2014.0154> (2015).
31. Partridge, S. R. *et al.* Definition of the attI1 site of class 1 integrons. *Microbiology* **146**(Pt 11), 2855–2864 (2000).
32. Svartstrom, O., Mushtaq, M., Pringle, M. & Segerman, B. Genome-wide relatedness of *Treponema pedis*, from gingiva and necrotic skin lesions of pigs, with the human oral pathogen *Treponema denticola*. *PLoS ONE* **8**, e71281. <https://doi.org/10.1371/journal.pone.0071281> (2013).
33. Dashper, S. G., Seers, C. A., Tan, K. H. & Reynolds, E. C. Virulence factors of the oral spirochete *Treponema denticola*. *J. Dent. Res.* **90**, 691–703. <https://doi.org/10.1177/0022034510385242> (2011).
34. Holt, S. C. & Ebersole, J. L. Porphyromonas gingivalis, *Treponema denticola*, and *Tannerella forsythia*: the ‘red complex’, a prototype polybacterial pathogenic consortium in periodontitis. *Periodontology* **2000**(38), 72–122. <https://doi.org/10.1111/1/j.1600-0757.2005.00113.x> (2005).
35. Wang, B. Y., Chi, B. & Kuramitsu, H. K. Genetic exchange between *Treponema denticola* and *Streptococcus gordonii* in biofilms. *Oral Microbiol. Immunol.* **17**, 108–112. <https://doi.org/10.1046/j.0902-0055.2001.00001.x> (2002).
36. Tansirichaiya, S., Rahman, M. A., Antepowicz, A., Mullany, P. & Roberts, A. P. Detection of novel integrons in the metagenome of human saliva. *PLoS ONE* **11**, e0157605. <https://doi.org/10.1371/journal.pone.0157605> (2016).
37. Wu, Y. W., Rho, M., Doak, T. G. & Ye, Y. Oral spirochetes implicated in dental diseases are widespread in normal human subjects and carry extremely diverse integron gene cassettes. *Appl. Environ. Microbiol.* **78**, 5288–5296. <https://doi.org/10.1128/AEM.00564-12> (2012).
38. Wu, Y. W., Doak, T. G. & Ye, Y. The gain and loss of chromosomal integron systems in the *Treponema* species. *BMC Evol. Biol.* **13**, 16. <https://doi.org/10.1186/1471-2148-13-16> (2013).
39. Treangen, T. J. & Rocha, E. P. C. Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes. *PLoS Genet.* **7**, e1001284. <https://doi.org/10.1371/journal.pgen.1001284> (2011).
40. Caceres, N. E. *et al.* Overexpression of the D-alanine racemase gene confers resistance to D-cycloserine in *Mycobacterium smegmatis*. *J. Bacteriol.* **179**, 5046–5055 (1997).
41. Merker, M. *et al.* Whole genome sequencing reveals complex evolution patterns of multidrug-resistant *Mycobacterium tuberculosis* Beijing strains in patients. *PLoS ONE* **8**, e82551. <https://doi.org/10.1371/journal.pone.0082551> (2013).
42. Desjardins, C. A. *et al.* Genomic and functional analyses of *Mycobacterium tuberculosis* strains implicate *ald* in D-cycloserine resistance. *Nat Genet* **48**, 544–551. <https://doi.org/10.1038/ng.3548> (2016).
43. Curtiss, R., Charamella, L. J., Berg, C. M. & Harris, P. E. Kinetic and genetic analyses of d-cycloserine inhibition and resistance in *Escherichia coli*. *J. Bacteriol.* **90**, 1238–1250 (1965).
44. Baisa, G., Stabo, N. J. & Welch, R. A. Characterization of *Escherichia coli* d-Cycloserine transport and resistant mutants. *J. Bacteriol.* **195**, 1389–1399. <https://doi.org/10.1128/jb.01598-12> (2013).
45. Karakousis, P. C. *Antimicrobial Drug Resistance* 271–283 (Humana Press, New York, 2009).
46. Forsberg, K. J. *et al.* The shared antibiotic resistome of soil bacteria and human pathogens. *Science* **337**, 1107–1111. <https://doi.org/10.1126/science.1220761> (2012).
47. Sommer, M. O., Dantas, G. & Church, G. M. Functional characterization of the antibiotic resistance reservoir in the human microflora. *Science* **325**, 1128–1131. <https://doi.org/10.1126/science.1176950> (2009).
48. Forsberg, K. J. *et al.* Bacterial phylogeny structures soil resistomes across habitats. *Nature* **509**, 612–616. <https://doi.org/10.1038/nature13377> (2014).
49. Moore, A. M. *et al.* Pediatric fecal microbiota harbor diverse and novel antibiotic resistance genes. *PLoS ONE* **8**, e78822. <https://doi.org/10.1371/journal.pone.0078822> (2013).
50. Poole, L. B. The basics of thiols and cysteines in redox biology and chemistry. *Free Radic. Biol. Med.* **80**, 148–157. <https://doi.org/10.1016/j.freeradbiomed.2014.11.013> (2015).
51. Sevier, C. S. & Kaiser, C. A. Formation and transfer of disulphide bonds in living cells. *Nat. Rev. Mol. Cell Biol.* **3**, 836–847 (2002).
52. Sanchez, R. *et al.* Protein structure modeling for structural genomics. *Nat. Struct. Biol.* **7**(Suppl), 986–990. <https://doi.org/10.1038/80776> (2000).
53. Park, I. S. & Walsh, C. T. D-Alanyl-D-lactate and D-alanyl-D-alanine synthesis by D-alanyl-D-alanine ligase from vancomycin-resistant *Leuconostoc mesenteroides*. Effects of a phenylalanine 261 to tyrosine mutation. *J. Biol. Chem.* **272**, 9210–9214 (1997).
54. Evangelopoulos, D. *et al.* Comparative fitness analysis of D-cycloserine resistant mutants reveals both fitness-neutral and high-fitness cost genotypes. *Nat. Commun.* **10**, 4177. <https://doi.org/10.1038/s41467-019-12074-z> (2019).
55. Cowan, M. M. Plant products as antimicrobial agents. *Clin. Microbiol. Rev.* **12**, 564–582 (1999).
56. Osonga, F. J. *et al.* Antimicrobial activity of a new class of phosphorylated and modified flavonoids. *ACS Omega* **4**, 12865–12871. <https://doi.org/10.1021/acsomega.9b00077> (2019).
57. Salam Pradeep, S. *et al.* Competitive inhibition of quercetin and apigenin at the ATP binding site of D-Alanine-D-Alanine Ligase of *Helicobacter pylori*: a molecular modeling approach. *Curr. Biotechnol.* **7**, 340–348. <https://doi.org/10.2174/2211550107666180612100441> (2018).
58. Hardy, K. G. *Bacillus Cloning Methods In DNA Cloning* 1–11 (IRL Press, New York, 1985).
59. Bruning, J. B., Murillo, A. C., Chacon, O., Barletta, R. G. & Sacchetti, J. C. Structure of the *Mycobacterium tuberculosis* D-alanine:D-alanine ligase, a target of the antituberculosis drug D-cycloserine. *Antimicrob. Agents Chemother.* **55**, 291–301. <https://doi.org/10.1128/AAC.00558-10> (2011).
60. Daub, E., Zawadzke, L. E., Botstein, D. & Walsh, C. T. Isolation, cloning, and sequencing of the *Salmonella typhimurium* *ddlA* gene with purification and characterization of its product, D-alanine:D-alanine ligase (ADP forming). *Biochemistry* **27**, 3701–3708 (1988).
61. Neuhaus, F. C. The enzymatic synthesis of D-alanyl-D-alanine. I. Purification and properties of D-alanyl-D-alanine synthetase. *J. Biol. Chem.* **237**, 778–786 (1962).
62. Brooks, H. B. *et al.* in *Assay Guidance Manual* (eds G. S. Sittampalam *et al.*) (Eli Lilly & Company and the National Center for Advancing Translational Sciences, 2004).
63. Zhang, Y. & Skolnick, J. Scoring function for automated assessment of protein structure template quality. *Proteins* **57**, 702–710. <https://doi.org/10.1002/prot.20264> (2004).
64. Koes, D. R., Baumgartner, M. P. & Camacho, C. J. Lessons learned in empirical scoring with smina from the CSAR 2011 benchmarking exercise. *J. Chem. Inf. Model.* **53**, 1893–1904. <https://doi.org/10.1021/ci300604z> (2013).
65. Jones, G., Willett, P., Glen, R. C., Leach, A. R. & Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **267**, 727–748. <https://doi.org/10.1006/jmbi.1996.0897> (1997).
66. Jones, G., Willett, P. & Glen, R. C. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J. Mol. Biol.* **245**, 43–53. [https://doi.org/10.1016/S0022-2836\(95\)80037-9](https://doi.org/10.1016/S0022-2836(95)80037-9) (1995).
67. Nissink, J. W. *et al.* A new test set for validating predictions of protein-ligand interaction. *Proteins* **49**, 457–471. <https://doi.org/10.1002/prot.10232> (2002).

68. Case, D. A. *et al.* The Amber biomolecular simulation programs. *J. Comput. Chem.* **26**, 1668–1688. <https://doi.org/10.1002/jcc.20290> (2005).
69. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: an N-log(N) method for Ewald sums in large systems. *J. Chem. Phys.* <https://doi.org/10.1063/1.464397> (1998).
70. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* [https://doi.org/10.1016/0021-9991\(77\)90098-5](https://doi.org/10.1016/0021-9991(77)90098-5) (1977).

Acknowledgements

We gratefully acknowledge the Commonwealth Scholarship Commission in the UK for supporting MAR (BDCA 2013-4), Public Health England for supporting SJ through a research grant to KMR (award code JGALASR) and the Erasmus programme for supporting MFM. We are thankful to Dr Haitham Hussain (UCL) for providing us with the pHCMC05 vector.

Author contributions

M.A.R.; Conceptualization, experimental, analysis, manuscript writing and editing. F.K.; experimental, analysis, manuscript editing. S.J.; experimental, analysis, manuscript editing. M.F.M.; experimental, analysis, manuscript editing. K.M.R.; experimental, analysis, manuscript editing. P.M.; analysis, manuscript editing. A.P.R.; conceptualization, analysis, manuscript writing and editing.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-77377-4>.

Correspondence and requests for materials should be addressed to A.P.R.

Reprints and permissions information is available at www.nature.com/reprints.

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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020