An Azobenzene G-quadruplex Ligand Exhibits Promising Antibacterial Activity against *Escherichia coli*

Javier Ramos-Soriano[1#], Yuiko Takebayashi[2 #], Jenny Samphire[1,2], Michael P. O'Hagan[1], Catherine Gurr[2], Kate J. Heesom[3], Philip A. Lewis[3], James Spencer[2]*, M. Carmen Galan[1]*

Email: jim.spencer@bristol.ac.uk and m.c.galan@bristol.ac.uk

[1]School of Chemistry, Cantock's Close, BS8 1TS, University of Bristol; [2]School of Cellular & Molecular Medicine, Biomedical Sciences Building, BS8 1TD, University of Bristol; [3]Proteomics Facility, Biomedical Sciences Building, BS8 1TD, University of Bristol.

[#]Equal contribution

Abstract

With the rise in bacterial antimicrobial resistance (AMR) and decline in antibiotic discovery, global healthcare is at a point of jeopardy. There is great need for the development of novel antimicrobials that target bacteria that have become resistant to our existing antibiotics, in particular Gram-negative species such as *Escherichia coli* which is responsible for opportunistic infections of already compromised patients. Here we demonstrate that a novel pyridinium-functionalised azobenzene scaffold **L20**, identified as a candidate ligand able target G-quadruplex (G4) structures in bacterial genomes/transcriptomes, shows promising antibacterial activity (MIC values ≤ 4 μg/ml) against multi-drug resistant *E. coli*. Tandem Mass Tag (TMT) proteomics applied to cultures of the *E. coli* type strain ATCC 25922 treated with sub-lethal concentrations of **L20**, identified seven G4-containing sequences as potential targets for **L20**. FRET (fluorescence resonance energy transfer) stabilisation assays indicate **L20** binds these selected sequences with variable and moderate affinity, in contrast to two comparator G4 ligands (stiff-stilbene **L5** and **pyridostatin** (**PDS**)) that better stabilise G4 structures but exhibit a lower antimicrobial activity. However, proteomic experiments also reveal that, alongside its superior antibacterial activity, **L20** treatment influences expression levels of more G4-associated proteins than either **L5** or PDS, and upregulates multiple essential proteins involved in translation. These findings identify strategies discovering potential G4 ligands as approaches that can lead to antibacterial candidates active against priority targets such as multi-drug resistant *E. coli*, and that targeting G4 sequences, and ligands such as **L20**, warrant further exploration as potential novel therapeutics with G4-mediated modes of action.

Introduction

Antimicrobial resistance (AMR) in pathogenic bacteria has become one of the most critical healthcare issues in the 21st century. With the continuing weakness of the antibacterial pipeline and decades of empirical antibiotic overuse[1], infections once readily treatable are generating increasing morbidity and mortality, conferring great urgency on the search for effective alternatives to currently available agents. Amongst bacterial pathogens, the World Health Organisation (WHO) has identified carbapenem-resistant, extended spectrum beta-lactamase (ESBL)-producing *Enterobacterales* (Gram-negative bacteria including *Escherichia* and *Klebsiella* species) as one of a select group of critical, priority level 1 pathogens on which research and development (R&D) of new antibiotics should be focused.[2]

G-quadruplexes (G4s) are thermodynamically stable secondary structures that form in guanine-rich regions of DNA and RNA sequences in both eukaryotes and prokaryotes.[3,4] These structures form transiently under physiological conditions, with four guanines interacting via Hoogsteen hydrogen bonds to form G-tetrads that stack in groups or two or more, and are further stabilised by monovalent cations such as Na^+ and K^+.[5,6] G4s have been observed to form within genes as well as in regulatory regions of DNA and mRNA, including telomeres, promoter sites, transcriptional start sites (TSS) and 5' and 3' untranslated regions (UTRs), and consequently influence gene regulation and expression.[7,8] The prevalence of these motifs, along with

growing realisation of their importance in human disease progression (in conditions including cancer, diabetes and neurogenerative diseases), and involvement in cellular processes in a wider range of organisms including plants, parasites, fungi, protozoa, bacteria and viruses; has excited considerable interest in their potential as therapeutic targets.[9] A database of G4 ligands, first compiled in 2013 and recently updated in 2021, now curates chemical and biological information for > 3200 small molecules, testifying to the continuing expansion of interest in R&D of G4-based drugs.[10,11].

Compared to the human system, however, the study of G4s in bacteria remains in its infancy. A 2006 *in silico* study of 18 prokaryotes, including *Escherichia coli, Pseudomonas aeruginosa* and *Mycobacterium bovis*, observed enrichment of G4s in regulatory regions within 200 bp upstream of genes involved in secondary metabolite biosynthesis, transcription and signal transduction.[12] Amongst the motifs highlighted, 40 were found to be conserved across multiple species, further supporting their likely importance in gene regulation and function. More recent whole genome sequencing (WGS) studies have identified hundreds of G4 sites in the DNA and RNA of *E. coli*.[13,14] Other studies have utilised a combination of techniques ranging from biophysical to biological assays (e.g. NMR, circular dichroism (CD) spectroscopy, isothermal titration calorimetry (ITC), thermal stability measurements, IC_{50} and qRT-PCR) to understand the role of G4s in the expression of essential genes involved in bacterial survival and pathogenesis in clinically relevant microorganisms. In *Mycobacterium tuberculosis* (Mtb), G4 sites were identified in the coding regions of essential genes responsible for virulence in host cells (*espK, espB*) and survival by membrane fluidity maintenance (*cyp51*).[15] The well characterised G4 ligand TMPyP4 ((5,10,15,20-Tetrakis-(N-methyl-4-pyridyl)porphine), an inhibitor of telomerase activity in cancer cell lines[16]) was shown to stabilise these secondary structures and to downregulate expression of these three genes, leading to inhibition of Mtb growth with an IC_{50} of 6.25 µM. In *Streptococcus pneumoniae*, essential genes involved in DNA repair (*recD*), drug efflux (*pmrA*) and virulence phase regulation (*hsdS*) all harboured intragenic G4 sites which were stabilised by TMPyP4, leading to inhibition of expression of a fluorescent reporter protein in the human HEK 293 cell line.[17] In multi-drug resistant *Klebsiella pneumoniae*, G4 sites in the promoter regions of five essential genes were identified as being stabilised by another well-known anti-viral G4 ligand BRACO-19, with decreased expression of their respective RNAs at a concentration of 5.38 µM and an IC_{50} for growth inhibition of 10.77 µM.[18]. These studies, although to date limited to a few bacterial species, highlight the potential of G4 motifs as effective targets for candidate antibacterial agents able to disrupt cellular function and viability.

In this work, a small G4 ligand library was screened against a selection of Gram-positive and -negative bacteria to investigate their possible antibacterial activity, and one lead antibacterial candidate **L20** was identified. LC-MS/MS analysis of *E. coli* cultures treated with this 4-methylpyridium azobenzene ligand **L20** was used to select a panel of candidate G4 target sequences for further biophysical characterisation. Fluorescence resonance energy transfer (FRET) analysis was undertaken on the selected sequences to establish the extent and stability of G4 formation in both DNA and RNA. **L20** and two comparator G4 ligands, stiff-benzene ligand **L5** known for its anticancer and antiparasitic activity[19] (which also showed antibacterial activity albeit more moderate) and the benchmark ligand **pyridostatin** (**PDS**), a well-known G4 inducer and stabiliser, were also tested for their ability to stabilise these sequences. Finally, the possible mechanism of **L20** antibacterial activity was explored by investigating the effect of exposure to **L20**, **L5** and **PDS** on the proteomes of *E. coli*. The results showed that **L20** does not stabilise the selected G4 target sequences as well as the comparators; however it does influence the expression of more G4-associated proteins and essential proteins involved in translation.

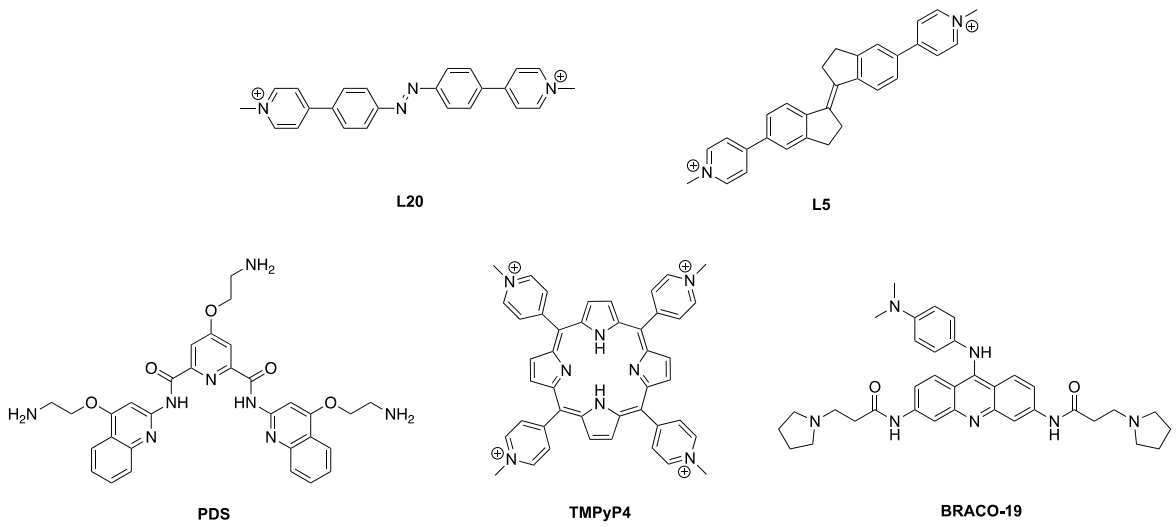
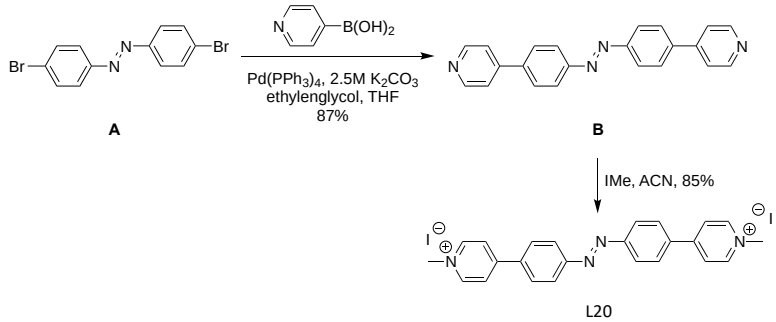


Figure 1. Structures of azobenzene **L20**, stiff-stilbene **L5**, pyridostatin (**PDS**), (5,10,15,20-Tetrakis-(N-methyl-4-pyridyl)porphine) (**TMPyP4**) and **BRACO-19** ligands.

Results

Synthesis of L20 G4 ligand

In brief, azobenzene derivative **L20** was prepared as the bis-iodo salt through a straightforward 2-step procedure (**Scheme 1**). Suzuki coupling of 4-pyridinylboronic acid with 1,2-bis(4-bromophenyl)-diazene **A**[20] afforded compound **B** in 87% yield. Finally, alkylation with iodomethane provided the target compound **L20** in excellent yield. All compounds were completely characterized by using standard spectroscopic and analytical techniques.



Scheme 1. Synthesis of azobenzene **L20**.

Antimicrobial activity against *E. coli* and *S. aureus*

A small library of G4 ligand candidates, along with the benchmark G4 ligand TMPyP4,[21] BRACO-19[22] and pyridostatin (PDS) were initially screened for antibacterial activity, using the agar disc susceptibility method specified by the Clinical and Laboratory Standard Institute (CLSI), against a range of reference and clinical *E. coli* (Gram-negative) and *S. aureus* (Gram-positive) strains with varying antimicrobial susceptibility profiles as summarised in **Table 5**. Of the tested ligands, three exhibited activity against *E. coli* strains (**L5, L6, L20**) and 11 against *S. aureus* strains (**L3, L5-8, L10, L14-15, L18-20**) (**Figure S1**). No growth inhibition activity was exhibited by any of the three comparator ligands against any of the strains tested.

Minimal inhibitory concentrations (MICs) were then determined for the *S. aureus* Newman and *E. coli* ATCC 25922, UTI 808 and IR60 strains by broth microdilution following CLSI guidelines. MIC values ranging from

0.25 - 8 µg/ml for the 11 ligands active against *S. aureus*, and 2 - 32 µg/ml for the three ligands active against *E. coli* were obtained, indicating antibacterial activity of these candidate G4 ligands (**Table 1**). Of the tested ligands, **L20** was selected for further study due to its promising activity against the multidrug resistant clinical *E. coli* isolates UTI 808 and IR60. **L5** was also selected as a less potent comparator, and was chosen over **L6** due to the similarity of its more planar structure to that of **L20**.

Table 1 Minimal inhibitory concentration (MIC) values for selected candidate G4 ligands against *S. aureus* and *E. coli*.

	MIC (µg/ml)			
	S. aureus	***E. coli***		
Ligand	**Newman**	**ATCC 25922**	**UTI 808**	**IR60**
L3	8	-	-	-
L5	0.25	32	-	32
L6	1	16	16	16
L7	2	-	-	-
L8	4	-	-	-
L10	4	-	-	-
L14	4	-	-	-
L15	4	-	-	-
L18	8	-	-	-
L19	8	-	-	-
L20	4	4	4	2
PDS	-	32	-	-

'-' indicates MIC values were not tested.

Proteomic analysis of *E. coli* treated with L20 to select for G4 sequences of interest

Next-generation sequencing-based techniques have recently been applied to both DNA and RNA of *E. coli* to experimentally identify G4 motifs.[13,14] These represent an advance, over previous *in silico* efforts, in mapping these functionally important "kinks" in the genome. Marsico *et al*[13] identified 47 observed G4 sequences (OQs; 22 in the sense (minus file) and 25 in the antisense (plus file) DNA strands) that were stabilised by the addition of K^+; and 560 G4 sequences (OQs; 287 on the sense (minus file) and 273 on the antisense (plus file) DNA strands) that were stabilised by the addition of K^+ + PDS; while Shao *et al*[14] identified 168 mRNA sequences that were stabilised by K^+. Given this large number of candidate binding sites for G4 ligands in the *E. coli* genome and transcriptome, a proteomic approach was therefore explored in an effort to identify candidate G4 sequences, and potential targets for **L5** and **L20**, for more detailed biophysical exploration (i.e FRET analysis). It was hypothesised that **L20**-induced stabilisation of G4 motifs in the *E. coli* genome would result in changes in protein expression from genes containing these sequences (either in upstream regulatory or coding regions) in comparison to the non-treated control proteome. *E. coli* ATCC 25922 cultures (in biological triplicate) were individually treated with 1 µg/ml of **L20** for 7 h, and their total cell proteomes analysed by TMT-labelled LC-MS/MS as described in Methods. For comparison, the total proteome of untreated control cultures, grown in biological triplicate under otherwise identical conditions, was also analysed at the 7 h timepoint. The harvest timepoint was selected based on observation of retardation of bacterial growth curves in the presence of ligand at the chosen concentration (**Figure S2**).

Analysis of LC-MS/MS data highlighted 106 proteins with changes in abundance levels that corresponded to the genes with OQs observed by Marsico *et al* (i.e. 21 genes from the K^+ list and 94 genes from the K^+ + PDS list with 9 genes overlapping between the two).[13] All corresponding 21 OQs from the K^+ list were screened using G4Hunter to identify the likelihood of G4 structures forming *in vitro*.[23,24] For the corresponding 94 OQs from the K^+ + PDS list, a threshold of ≥ ±1 Log2 fold change (FC) was first applied to focus on proteins with at least 2-fold changes in levels, and the resulting 27 OQs were screened using G4Hunter. A total of seven G4

sequences were chosen for FRET analysis, based on their high propensity to form G4 structures and/or the significant LogFC in the abundance of the corresponding protein compared to the non-treated control (**Table 2**).

Proteomic analysis of *E. coli* treated with L5 comparators, L20 and PDS

To allow comparisons with the **L20**-treated proteomics and FRET results and to provide further insight, cultures in biological triplicate were grown in the presence or absence of 64 µg/ml of **L5** and **PDS** for 7 h (**Figure S2**) and total cell lysates collected and processed for TMT-labelled LC-MS/MS analysis. The 7 h timepoint was chosen to align with the **L20** experiment. The LogFC in levels of proteins corresponding to the seven selected G4 sequences were calculated and are summarised in **Table 2**.

FRET analysis of G4-associated genes selected by proteomics

FRET melting assays were carried out on the seven G4 sequences (**Table 6**), selected from LC-MS/MS analysis as described above, each in both DNA and RNA format in the presence of 100 µM K^+ and 5 µM ligand (**PDS, L20** or **L5)**. The results are summarised in **Table 2**. Although these sequences were selected based upon analysis of the proteomic profile of *E. coli* treated with **L20**, FRET analysis revealed this ligand to be capable of no more than moderate stabilisation of any of these sequences, compared to **L5** or **PDS**, which stabilised all of the DNA and RNA sequences (except *pdxA*) to a similar degree. In contrast, **L20** appears more selective, stabilising four out of the 7 sequences tested (i.e. *glnD, pdxA, thrA* and *yjbD*). Interestingly, the ability to stabilise a wider range of G4 sequences does not appear to correlate with antibacterial activity; while **L20** exhibited 8-fold better antibacterial activity against *E. coli* (MIC of 4 µg/ml) compared to **L5** and **PDS** (MIC 32 µg/ml in each case, **Table 1**), with only two exceptions (RNAs containing G4 sequences associated with the *glnD* and *yjbD* genes) it did not stabilise the selected G4 sequences as strongly as the latter two ligands. Furthermore, G4 binding/stabilisation does not always correlate with changes in protein expression levels, as highlighted in **Table 2** (see colour scheme). Proteomic data showed that four of the G4 sequences (i.e. *clpB, dppA, hemL* and *yjbD*) bound by **PDS** *in vitro* showed significant expression changes (LogFC ≥ ±1), for **L20** the corresponding figure is three (i.e. *pdxA, thrA* and *yjbD*) and for **L5** only one (*dppA*). In the case of **L20**, one of the G4 sequences showing the lowest level of stabilisation (*dppA*) exhibited the largest decrease in protein expression (LogFC of -2.945). These data suggest two possibilities – either the antibacterial mode of action of **L20** is not dependent upon binding to G4-containing sequences, or the target G4-containing sequences important to antibacterial activity were not among the seven selected for FRET analysis.

Table 2 LC-MS/MS and FRET melting assay results of the selected seven G4 sequences

				PDS		L20		L5	
Gene	**Strand**	**Protein Description**	**G4Hunter Score**	**LogFC**	**°C Stabilised**	**LogFC**	**°C Stabilised**	**LogFC**	**°C Stabilised**
clpB	Antisense	Chaperone protein	1.04	-1.028	24 (RNA) 19 (DNA)	-1.128	3 (RNA) / 2 (DNA)	0.058*	24 (RNA) 17 (DNA)
dppA	Antisense	Periplasmic dipeptide transport protein	1.3	-1.925	28 (RNA) 28 (DNA)	-2.945	5 (RNA) / 3 (DNA)	-1.23	25 (RNA) 28 (DNA)
glnD	Sense	Bifunctional uridylyltransferase	1.08	0.250*	11 (RNA) 32 (DNA)	0.195	16 (RNA) / 11 (DNA)	0.140*	14 (RNA) 26 (DNA)
hemL	Sense	Glutamate-1-semialdehyde 2,1-aminomutase	NA	1.744	17 (RNA) 20 (DNA)	0.443	1 (RNA) / 2 (DNA)	0.283	16 (RNA) 21 (DNA)
pdxA	Sense	4-hydroxythreonine-4-phosphate dehydrogenase	1.38	-0.593	28 (DNA)	-1.979	10 (DNA)	ND	16 (DNA)
thrA	Antisense	Bifunctional aspartokinase/homoserine dehydrogenase	1.2	0.026*	20 (RNA) 30 (DNA)	1.364	11 (RNA) / 12 (DNA)	-0.355*	18 (RNA) 25 (DNA)
yjbD	Sense	Uncharacterised protein	1.44	-1.142	11 (RNA) 38 (DNA)	-1.317	14 (RNA) / 15 (DNA)	ND	10 (RNA) 25 (DNA)

Strand indicates location of G4 motif. Asterisk highlights non-significant LogFC with p>0.05 (n=3, T-test). G4Hunter score indicates the propensity of the sequence to form a G4 (score > 1.0). NA; not available. ND; not detected. Green font indicates where proteomic and FRET results correlate. Red font indicates no correlation.

L20, L5 and PDS influence the expression of G4-associated proteins

To further investigate whether the antibacterial activity of **L20** is driven by effects upon G4 formation, regulation or stability, the total cell proteomes of *E. coli* cultures individually treated with **PDS, L20** and **L5** were analysed in more depth. From the total cell lysates of the respective cultures, 1613, 2168 and 1518 proteins were identified. The total numbers of proteins with changes in abundance compared to the untreated controls (LogFC ≥ 1.0 and LogFC ≤ -1.0) were similar across the three conditions (**PDS, L20** and **L5**-treated); 16-20% of proteins were upregulated, and 23-28% were downregulated by G4-ligand treatment (p

< 0.05) (**Table S2**). To elucidate the possible mechanism of antibacterial activity of **L20**, differentially expressed proteins were mapped in STRING to compare protein-protein interactions (PPI) and enrichments.[25,26] Despite the difference between their antibacterial activities, **L20** and **PDS** both upregulated similar sets of proteins (e.g. proteins found in the small and large ribosomal subunits) that clustered in the biological process of translation, and showed significant enrichment in the interactions ($p < 1.0 \times 10^{-16}$ for both). **L5**, on the other hand, upregulated proteins involved in anaerobic/cellular respiration and response to stimulus, with no significant enrichment in PPI ($p = 0.119$).

Of the 404 G4-containing genes identified by Marsico and Shao *et al*[13,14], 19.6%, 24.3% and 15.6% of the corresponding proteins were observed to be upregulated, and 19.6%, 26.2% and 23.0% downregulated in the **PDS, L20 and L5**-treated cultures respectively ($p < 0.05$) (**Table 3A**). This suggests that, compared to the other two ligands, **L20** influences the expression of more G4-associated proteins, consistent with the possibility of a G4-mediated mechanism, although further experimental work is required to confirm this. Similar observations were made on proteins with ≥ 1.0 and ≤ -1.0 LogFC, although the effect of **L20** on downregulation of G4-associated proteins was similar to that of **L5** (11.4% vs 10.6%).

Next, the overall effect of **PDS, L20 and L5** exposure on expression of essential genes was examined, as this would be expected to influence the viability of *E. coli* cells. Previous groups have also observed the effect of G4 ligands on expression of essential genes in various species such as *M. tb*, *S. pneumoniae* and *K. pneumoniae* involved in processes such as DNA repair, drug efflux etc.[15,17,18] The protein abundances of 302 genes listed as essential for cell growth on the Profiling of *E. coli* Chromosome (PEC) database[27] (as of 24 Feb 2021) were identified from the proteomic datasets. (The PEC database is curated and maintained based on research reports and deletion mutation studies of *E. coli* MG1655). The results are summarised in **Table 3B**. A larger number of essential genes were upregulated in the total proteome of the **L20**-treated cultures, compared to those treated with **PDS** and **L5** (190 compared to 155 and 79 respectively). However, twice as many essential genes were downregulated in the **L5**-treated cultures, compared to those exposed to **L20** (88 vs 44). Closer inspection of the list of candidate proteins identified multiple essential genes involved in the translational machinery to be upregulated in the **PDS-** and **L20-,** and downregulated in the **L5**-treated cultures. This is consistent with the behaviour observed in the bacterial growth curves (**Figure S2**), where growth of **PDS**- and **L20-**treated cells recovered after an extended lag phase, but that of **L5 - treated** cells did not.

Forty-three of the G4-containing genes identified by Marsico and Shao *et al*[13,14] are essential genes listed in PEC (**Table 4**). Protein expression from 31 of these essential G4-associated genes was upregulated in the **L20**-treated cultures, compared to 26 and 11 in the **PDS** and **L5**-treated datasets respectively. However, protein levels for only 3 genes (*groL, murD, efp*) were downregulated by **L20** treatment, compared to 4 and 13 for **PDS** and **L5**, respectively, and this was not to the cut-off level of LogFC ≥ 1.0 where these changes would be deemed significant. This suggests that **PDS** and **L20** may not exhibit antibacterial effect against *E. coli* in a G4-dependent manner while **L5** is more likely to, although this will need to be further explored.

Table 3 Summary of abundance changes in (A) G4-associated and (B) essential PEC-associated proteins from total cell proteome of *E. coli* cultures treated with PDS, L20 or L5.

A	**PDS**		**L20**		**L5**	
	G4 proteins	**% G4**	**G4 proteins**	**% G4**	**G4 proteins**	**% G4**
	p<0.05					
Upregulated	79	19.6%	98	24.3%	63	15.6%
Downregulated	79	19.6%	106	26.2%	93	23.0%
	p<0.05, LogFC ≥ ±1.0					
Upregulated	18	4.5%	32	7.9%	25	6.2%
Downregulated	34	8.4%	46	11.4%	43	10.6%

B	PDS		L20		L5	
	Protein	%	Protein	%	Protein	%
↑(All)	155	51%	190	63%	79	26%
↑(All) G4	26	9%	31	10%	11	4%
↑(LogFC ≥ ± 1.0)	57	19%	69	23%	16	5%
↑(LogFC ≥ ± 1.0) G4	5	2%	11	4%	1	0%
↓(All)	30	10%	44	15%	88	29%
↓(All) G4	4	1%	3	1%	13	4%
↓ (LogFC ≥ ± 1.0)	4	1%	2	1%	31	10%
↓ (LogFC ≥ ± 1.0) G4	0	0%	0	0%	5	2%
Total PEC	302					
Total G4 seq in PEC	43					

Table 4 Summary of 43 essential PEC genes associated with G4 motifs.

		LogFC (p<0.05)		
PEC	Protein Description	PDS	L20	L5
nadD	nicotinic acid mononucleotide adenylyltransferase, NAD(P)-dependent	1.575		
ftsW	lipid II flippase; integral membrane protein involved in stabilizing FstZ ring during cell division	0.937		
ftsZ	GTP-binding tubulin-like cell division protein	-0.276		
mukE	Chromosome condensin MukBEF, MukE localization factor	0.232	1.582	
gyrA	DNA gyrase (type II topoisomerase), subunit A	0.690	1.063	
hemB	5-aminolevulinate dehydratase (porphobilinogen synthase)	0.494	1.038	
pth	peptidyl-tRNA hydrolase	0.561	1.038	
glyS	glycine tRNA synthetase, beta subunit	0.474	0.877	
topA	DNA topoisomerase I, omega subunit	0.312	0.863	
mreB	cell wall structural complex MreBCD, actin-like component MreB	1.673	0.641	
zipA	FtsZ stabilizer	0.745	0.466	
mrdB	cell wall shape-determining protein		0.255	
glmU	fused N-acetyl glucosamine-1-phosphateuridyltransferase/glucosamine-1-phosphate acetyl transferase	0.531	0.195	
ligA	DNA ligase, NAD(+)-dependent		0.143	
efp	polyproline-specific translation elongation factor EF-P	1.037	-0.152	
prfB	peptide chain release factor RF-2	1.115	1.200	1.037
thrS	threonyl-tRNA synthetase	0.628	0.265	0.962
rho	transcription termination factor	0.899	0.362	0.899
nusA	transcription termination/antitermination L factor		0.681	0.727
lnt	apolipoprotein N-acyltransferase		0.712	0.696
mnmA	tRNA(Gln,Lys,Glu) U34 2-thiouridylase, first step in mnm(5)-s(2)U34-tRNA synthesis	0.719	1.798	0.638
suhB	inositol monophosphatase	0.910	0.989	0.564
yidC	membrane protein insertase	0.696	1.551	0.342
murC	UDP-N-acetylmuramate:L-alanine ligase	0.386	0.419	0.299
hemL	glutamate-1-semialdehyde aminotransferase (aminomutase)	1.744	0.443	0.283
rpoB	RNA polymerase, beta subunit	0.727	0.380	0.277
valS	valyl-tRNA synthetase	0.687	0.113	-0.171
asnS	asparaginyl tRNA synthetase		1.622	-0.188
leuS	leucyl-tRNA synthetase	0.593	1.033	-0.226
groL	Cpn60 chaperonin GroEL, large subunit of GroESL	-0.631	-0.705	-0.325
bamA	outer membrane protein assembly factor, forms pores; required for OM biogenesis; in BamABCDE OM protein co	-0.395		-0.429
gyrB	DNA gyrase, subunit B	0.563	1.015	-0.477
pyrG	CTP synthetase	0.892	1.089	-0.685
murD	UDP-N-acetylmuramoyl-L-alanine:D-glutamate ligase		-0.223	-0.937
tyrS	tyrosyl-tRNA synthetase		0.900	-1.021
accC	acetyl-CoA carboxylase, biotin carboxylase subunit	0.558	0.835	-1.202
prmC	N5-glutamine methyltransferase, modifies release factors RF-1 and RF-2			-1.242
ispH	4-hydroxy-3-methylbut-2-enyl diphosphate reductase, 4Fe-4S protein	-0.179	0.557	-1.429
folC	bifunctional folylpolyglutamate synthase/ dihydrofolate synthase		0.586	-1.508
alaS	alanyl-tRNA synthetase			
dnaA	chromosomal replication initiator protein DnaA, DNA-binding transcriptional dual regulator			
murF	UDP-N-acetylmuramoyl-tripeptide:D-alanyl-D- alanine ligase			
rnpB	RNase P, M1 RNA component			

Heatmap shading: red indicate increases in protein abundance, blue indicate decreases compared to non-treated control, white indicate no proteins detected.

Conclusions

In summary, we have shown that the candidate G4 ligand **L20** has promising antibacterial activity against reference and multi-drug resistant clinical *E. coli* strains. This compares with the findings for the comparator G4 ligands **PDS** and **L5**, for which antibacterial activity was 8-fold weaker. To date, this is the first azobenzene G4 ligand to be reported to have activity against resistant *E. coli* strains. However, FRET analyses of G4 formation in the presence of the three ligands established that **L20** does not stabilise any of the seven sequences tested as well as either **PDS** or **L5**, suggesting either that the mechanism of antibacterial activity of **L20** is not G4-mediated, or that the **L20**-specific G4-sequences responsible remain to be identified and were not among those selected for FRET analysis. In support of the latter conclusion, proteomic data revealed that **L20** influences the expression of more G4-associated proteins than either **PDS** or **L5**, hinting at the possibility that antibacterial activity is indeed associated with a G4-related mechanism. Furthermore, **L20** affects the expression levels of more essential genes than either **PDS** or **L5**, in particular upregulating multiple essential proteins involved in the cellular translational machinery. In summary, our study identifies G4 ligands to be interesting antibacterial candidates, with the scope to treat multi-drug resistant *E. coli* and warrants further exploration as potential G4-mediated novel therapeutics.

Materials and Methods

Bacterial strains, growth conditions

Bacterial strains used in this study are summarised in **Table 5**. Unless otherwise specified, all microorganisms were grown on Nutrient or Luria-Bertani agar and broth (Sigma) for 18-20 h at 37°C static and 180 rpm shaking respectively.

Table 5 Bacterial strains used in this study.

Strain Name	Description	Source
E. coli ATCC 25922	CLSI recommended susceptible control strain	American Type Culture Collection
E. coli UTI 808	Multi-drug resistant clinical UTI isolate	Gift from Dr Jacqueline Findlay
E. coli IR60	Multi-drug resistant clinical isolate	Gift from Prof Tim Walsh (Oxford University)
S. aureus strain Newman	Clinical isolate	Gift for Dr Angela Nobbs
S. aureus ATCC 29213	CLSI recommended susceptible control strain	American Type Culture Collection
S. aureus 21773	hVISA	Gift from Prof Allistair MacGowan (Southmead Hospital, Bristol)
S. aureus USA 300 LAC	Community acquired MRSA	Gift from Dr Angela Nobbs
S. aureus 17608	VSSA, MRSA	Gift from Prof Allistair MacGowan (Southmead Hospital, Bristol)

hVISA, heterogenous vancomycin intermediate *S. aureus*; MRSA, methicillin resistant *S. aureus*; VSSA, vancomycin susceptible *S. aureus*

Antimicrobial susceptibility assays

Bacterial strains were tested against G4 ligands by disc susceptibility and/or broth microdilution methods as outlined by the CLSI with Mueller-Hinton agar and cation-adjusted Mueller-Hinton broth (Sigma) respectively.[28] For the disc susceptibility assays, 25 µg of each ligand was impregnated onto sterile 5 mm filter discs before placing on inoculated bacterial plates.

Growth curves for determining proteomic culture conditions

10 ml overnight *E. coli* ATCC 25922 cultures were diluted to a final OD_{600} of 0.08-0.1 in Muller Hinton Broth 2 (Sigma-Aldrich) with doubling concentrations of PDS, L20 and L5 in flat-bottom 96-well plates (Corning) and

placed in a plate reader (BMG Omega) for absorbance readings at 600 nm every 10 min for 25 h at 37°C to select for the appropriate ligand concentration. The growth curves assays were then upscaled to 10 ml volumes in 50 ml conical flasks with 0, 32 and 64 µg/ml PDS and L5 and 0, 1, 2 µg/ml L20 and grown at 37°C, 180rpm. Absorbance readings were taken every 1 h for 6-12 h and then left overnight. The last absorbance reading was taken at +24 h.

Culture preparation, 6-plex TMT-labelling and LC-MS/MS

Whole cell lysates were obtained from *E. coli* ATCC 25922 cultures grown in Mueller Hinton Broth 2 in triplicate for 7 h, with or without 1, 64 and 64 µg/ml L20, PDS and L5 respectively. Aliquots of 50 µg of each sample were digested with trypsin (2.5 µg trypsin per 100 µg protein; 37°C, overnight), labelled with Tandem Mass Tag (TMT) 6-plex reagents according to the manufacturer's protocol (Thermo Fisher Scientific) and the labelled samples pooled.

The pooled sample was evaporated to dryness, resuspended in 5% formic acid and then desalted using a SepPak cartridge according to the manufacturer's instructions (Waters, Milford, Massachusetts, USA). Eluate from the SepPak cartridge was again evaporated to dryness and resuspended in buffer A (20 mM ammonium hydroxide, pH 10) prior to fractionation by high pH reversed-phase chromatography using an Ultimate 3000 liquid chromatography system (Thermo Scientific). In brief, the sample was loaded onto an XBridge BEH C18 Column (130Å, 3.5 µm, 2.1 mm X 150 mm, Waters, UK) in buffer A and peptides eluted with an increasing gradient of buffer B (20 mM Ammonium Hydroxide in acetonitrile, pH 10) from 0-95% over 60 minutes. The resulting fractions were evaporated to dryness and resuspended in 1% formic acid prior to analysis by nano-LC MSMS using an Orbitrap Fusion Lumos mass spectrometer (Thermo Scientific).

Nano-LC Mass Spectrometry

High pH RP fractions were further fractionated using an Ultimate 3000 nano-LC system in line with an Orbitrap Fusion Lumos mass spectrometer (Thermo Scientific). In brief, peptides in 1% (vol/vol) formic acid were injected onto an Acclaim PepMap C18 nano-trap column (Thermo Scientific). After washing with 0.5% (vol/vol) acetonitrile 0.1% (vol/vol) formic acid peptides were resolved on a 250 mm × 75 µm Acclaim PepMap C18 reverse phase analytical column (Thermo Scientific) over a 150 min acetonitrile gradient, divided into 7 gradient segments (1-6% solvent B over 1 min., 6 - 15% B over 58 min., 15 - 32% B over 58 min., 32 - 40% B over 5 min., 40 - 90% B over 1 min., held at 90% B for 6 min and then reduced to 1% B over 1 min.) with a flow rate of 300 nl min^{-1}. Solvent A was 0.1% formic acid and Solvent B was aqueous 80% acetonitrile in 0.1% formic acid. Peptides were ionized by nano-electrospray ionization at 2.0kV using a stainless steel emitter with an internal diameter of 30 µm (Thermo Scientific) and a capillary temperature of 275°C.

All spectra were acquired using an Orbitrap Fusion Lumos mass spectrometer controlled by Xcalibur 3.0 software (Thermo Scientific) and operated in data-dependent acquisition mode using an SPS-MS3 workflow. FTMS1 spectra were collected at a resolution of 120 000, with an automatic gain control (AGC) target of 200 000 and a max injection time of 50ms. Precursors were filtered with an intensity threshold of 5000, according to charge state (to include charge states 2-7) and with monoisotopic peak determination set to Peptide. Previously interrogated precursors were excluded using a dynamic window (60 s +/- 10 ppm). The MS2 precursors were isolated with a quadrupole isolation window of 0.7m/z. ITMS2 spectra were collected with an AGC target of 10 000, max injection time of 70ms and CID collision energy of 35%.

For FTMS3 analysis, the Orbitrap was operated at 50 000 resolution with an AGC target of 50 000 and a max injection time of 105 ms. Precursors were fragmented by high energy collision dissociation (HCD) at a normalised collision energy of 60% to ensure maximal TMT reporter ion yield. Synchronous Precursor Selection (SPS) was enabled to include up to 5 MS2 fragment ions in the FTMS3 scan.

Proteomic analysis of G4-associated and essential proteins for cell growth

The files summarised in **Table S1** from the Marsico *et al* study were used to identify the genes carrying observed G4 sequences (OQs) in the sense and antisense DNA strands of *E. coli* ATCC 25922.[13] An R script was written to extract the OQ sequences and associated genes from NCBI GenBank. G4 sites that were > 3,000 bp upstream any ORFs were classified as non-consequential to gene expression and removed from further analysis. A list of the genes associated with RNA G4 sequences identified by Shao *et al* was taken from their Table S1A.[14] All G4-associated genes (hereby referred to as G4 genes) identified by these two studies were collated and duplicates removed. The final list consisted of 404 G4 genes for further analysis. Analysis on essential proteins for cell growth was conducted based on the essential genes listed in the Profiling of *E. coli* Chromosome (PEC) database (as of 24 Feb 2021) (https://shigen.nig.ac.jp/ecoli/pec/index.jsp).

Bioinformatic analysis

STRING Version 11.5 (https://string-db.org/) was used to study the protein-protein interaction networks of the proteins with changes in abundance due to G4 ligand treatment.[25,26] The highest confidence score of 0.900 was applied to the analysis. G4Hunter (https://bioinformatics.cruk.cam.ac.uk/G4Hunter/) was used to score the propensity of G4-structures to form from the sequences of interest.[23,24]

FRET melting assays

Fluorescence resonance energy transfer (FRET) melting assays were performed according to the procedure reported by De Cian and co-workers[29] on Roche LightCycler 480 qPCR instrument. In these assays, the oligonucleotides of interest were obtained labelled at the 5' and 3' ends with FAM (a fluorescence donor) and TAMRA (a fluorescence quencher) respectively. In the folded state, proximity of the donor and quencher mean that FAM fluorescence is not observed since energy is transferred non-radiatively to TAMRA by FRET. As the temperature is raised and the secondary structure denatures, the fluorophores move further apart and hence the fluorescence signal increases. From the resulting curve, the characteristic melting temperature (T1/2) is defined as that at which the normalised fluorescence signal equals 0.5. The change in melting temperature (ΔTm) induced by a small molecule ligand compared to that of the oligonucleotide in the absence of ligand provides an indication of the ligand's ability to stabilise the G4 structure.

All oligonucleotides used (**Table 6**) were purchased from Eurogentec (Belgium), purified by HPLC and delivered dry. Oligonucleotide concentrations were determined by UV-absorbance using a NanoDrop 2000 Spectrophotometer from Thermo Scientific.

All sequences were annealed before use by heating for 2 minutes at 90°C and then placed immediately into ice. The final concentration of oligonucleotide was 200 nM in all cases. Ligand concentrations were either 1 µM, 2 µM, 5 µM or 10 µM. Each sample was tested in duplicate on the same plate, and each plate was repeated in at least duplicate to assess the reproducibility of all results.

Appropriate control experiments were also carried out for each sample set. Data processing was carried out using Origin 9, with ΔT1/2 used to represent ΔTm.

Table 6 G4 sequences selected by LC-MS/MS and analysed by FRET melting assay

Oligo name	Mod5'	Sequence	Mod3'	Bases
E.coli_dppA_DNA	6-FAM	GGG-CTG-GAC-TGG-CGA-TAA-CGG-GG	TAMRA	23
E.coli_thrA_DNA	6-FAM	GGG-CGA-TGG-GGG-TAA-TGG-TGC-GGG-GG	TAMRA	26
E.coli_yjbD_DNA	6-FAM	GGG-AAA-AGG-GTT-AGG-GTG-AGG-G	TAMRA	22
E.coli_pdxA_DNA	6-FAM	GGG-GAG-TTG-GGG-GAA-TAA-GGG-CGG-AGG-G	TAMRA	28
E.coli_clpB_DNA	6-FAM	GGC-GCG-TTG-GAC-GGG-G	TAMRA	16
E.coli_glnD_DNA	6-FAM	GCG-GTT-GAC-CGG-GCA-GGG-TGG-G	TAMRA	22
E.coli_hemL_DNA	6-FAM	GGT-CCG-GTC-TAT-CAG-GCG-GGT	TAMRA	21
E.coli_dppA_RNA	6-FAM	GGG-CUG-GAC-UGG-CGA-UAA-CGG-GG	TAMRA	23
E.coli_thrA_RNA	6-FAM	GGG-CGA-UGG-GGG-UAA-UGG-UGC-GGG-GG	TAMRA	26
E.coli_yjbD_RNA	6-FAM	GGG-AAA-AGG-GUU-AGG-GUG-AGG-G	TAMRA	22
E.coli_pdxA_RNA	6-FAM	GGG-GAG-UUG-GGG-GAA-UAA-GGG-CGG-AGG-G	TAMRA	28
E.coli_clpB_RNA	6-FAM	GGC-GCG-UUG-GAC-GGG-G	TAMRA	16
E.coli_glnD_RNA	6-FAM	GCG-GUU-GAC-CGG-GCA-GGG-UGG-G	TAMRA	22
E.coli_hemL_RNA	6-FAM	GGU-CCG-GUC-UAU-CAG-GCG-GGU	TAMRA	21

Acknowledgements

This research was supported by ERC-COG: 648239 (MCG), EPSRC EP/G036764/1 (JSamphire) and EPSRC EP/S026215/1 (MCG, JS). EP/R043361/1 (TS/R014329/1) (MCG, JS) and GCRF EP/T020288/1 (MCG, JS). EPSRC EP/L015366/1 (MPO) and the European Research Council (ERC-COG: 648239) and J. R.-S. thanks MSCA fellowship (project 843720-BioNanoProbes)

References

1. Hutchings, M. I., Truman, A. W. & Wilkinson, B. Antibiotics: past, present and future. *Curr. Opin. Microbiol.* **51**, 72–80 (2019).
2. WHO publishes list of bacteria for which new antibiotics are urgently needed. *World Health Organization* http://www.who.int/news-room/detail/27-02-2017-who-publishes-list-of-bacteria-for-which-new-antibiotics-are-urgently-needed.
3. Gellert, M., Lipsett, M. N. & Davies, D. R. HELIX FORMATION BY GUANYLIC ACID. *Proc. Natl. Acad. Sci. U. S. A.* **48**, 2013–2018 (1962).
4. Saranathan, N. & Vivekanandan, P. G-Quadruplexes: More Than Just a Kink in Microbial Genomes. *Trends Microbiol.* **27**, 148–163 (2019).

5. Henderson, E., Hardin, C. C., Walk, S. K., Tinoco, I. & Blackburn, E. H. Telomeric DNA oligonucleotides form novel intramolecular structures containing guanine·guanine base pairs. *Cell* **51**, 899–908 (1987).
6. Williamson, J. R., Raghuraman, M. K. & Cech, T. R. Monovalent cation-induced structure of telomeric DNA: The G-quartet model. *Cell* **59**, 871–880 (1989).
7. Alberti, P. & Mergny, J.-L. DNA duplex–quadruplex exchange as the basis for a nanomolecular machine. *Proc. Natl. Acad. Sci.* **100**, 1569–1573 (2003).
8. Alberti, P., Bourdoncle, A., Saccà, B., Lacroix, L. & Mergny, J.-L. DNA nanomachines and nanostructures involving quadruplexes. *Org. Biomol. Chem.* **4**, 3383–3391 (2006).
9. Street, S. T. G. *et al.* Imide Condensation as a Strategy for the Synthesis of Core-Diversified G-Quadruplex Ligands with Anticancer and Antiparasitic Activity**. *Chem. – Eur. J.* **27**, 7712–7721 (2021).
10. Li, Q. *et al.* G4LDB: a database for discovering and studying G-quadruplex ligands. *Nucleic Acids Res.* **41**, D1115–D1123 (2013).
11. Wang, Y.-H. *et al.* G4LDB 2.2: a database for discovering and studying G-quadruplex and i-Motif ligands. *Nucleic Acids Res.* gkab952 (2021) doi:10.1093/nar/gkab952.
12. Rawal, P. *et al.* Genome-wide prediction of G4 DNA as regulatory motifs: Role in Escherichia coli global regulation. *Genome Res.* **16**, 644–655 (2006).
13. Marsico, G. *et al.* Whole genome experimental maps of DNA G-quadruplexes in multiple species. *Nucleic Acids Res.* **47**, 3862–3874 (2019).
14. Shao, X. *et al.* RNA G-Quadruplex Structures Mediate Gene Regulation in Bacteria. *mBio* **11**, (2020).
15. Mishra, S. K. *et al.* Characterization of G-Quadruplex Motifs in espB, espK, and cyp51 Genes of Mycobacterium tuberculosis as Potential Drug Targets. *Mol. Ther. - Nucleic Acids* **16**, 698–706 (2019).
16. Wheelhouse, R. T., Sun, D., Han, H., Han, F. X. & Hurley, L. H. Cationic Porphyrins as Telomerase Inhibitors: the Interaction of Tetra-(N-methyl-4-pyridyl)porphine with Quadruplex DNA. *J. Am. Chem. Soc.* **120**, 3261–3262 (1998).
17. Mishra, S. K. *et al.* Characterization of highly conserved G-quadruplex motifs as potential drug targets in Streptococcus pneumoniae. *Sci. Rep.* **9**, 1791 (2019).

18. Shankar, U., Jain, N., Mishra, S. K., Sharma, T. K. & Kumar, A. Conserved G-Quadruplex Motifs in Gene Promoter Region Reveals a Novel Therapeutic Approach to Target Multi-Drug Resistance Klebsiella pneumoniae. *Front. Microbiol.* **11**, 1269 (2020).
19. O'Hagan, M. P., Peñalver, P., Gibson, R. S. L., Morales, J. C. & Galan, M. C. Stiff-Stilbene Ligands Target G-Quadruplex DNA and Exhibit Selective Anticancer and Antiparasitic Activity. *Chem. – Eur. J.* **26**, 6224–6233 (2020).
20. Kato, T., Gon, M., Tanaka, K. & Chujo, Y. Modulation of stimuli-responsiveness toward acid vapor between real-time and write-erase responses based on conjugated polymers containing azobenzene and Schiff base moieties. *J. Polym. Sci.* **59**, 1596–1602 (2021).
21. Fujiwara, N., Mazzola, M., Cai, E., Wang, M. & Cave, J. W. TMPyP4, a Stabilizer of Nucleic Acid Secondary Structure, Is a Novel Acetylcholinesterase Inhibitor. *PLOS ONE* **10**, e0139167 (2015).
22. Burger, A. M. *et al.* The G-quadruplex-interactive molecule BRACO-19 inhibits tumor growth, consistent with telomere targeting and interference with telomerase function. *Cancer Res.* **65**, 1489–1496 (2005).
23. Bedrat, A., Lacroix, L. & Mergny, J.-L. Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic Acids Res.* **44**, 1746–1759 (2016).
24. Brázda, V. *et al.* G4Hunter web application: a web server for G-quadruplex prediction. *Bioinforma. Oxf. Engl.* **35**, 3493–3495 (2019).
25. Szklarczyk, D. *et al.* STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019).
26. Szklarczyk, D. *et al.* The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* **49**, D605–D612 (2021).
27. Yamazaki, Y., Niki, H. & Kato, J. Profiling of Escherichia coli Chromosome database. *Methods Mol. Biol. Clifton NJ* **416**, 385–389 (2008).

28. Patel, J. *et al.* M07Ed11 | Methods for Dilution Antimicrobial Susceptibility Tests for Bacteria That Grow Aerobically, 11th Edition. *Clinical & Laboratory Standards Institute* https://clsi.org/standards/products/microbiology/documents/m07/.
29. De Cian, A. *et al.* Fluorescence-based melting assays for studying quadruplex ligands. *Methods San Diego Calif* **42**, 183–195 (2007).