In Silico Prediction and Prioritisation of Novel Selective Antimicrobial Drug Targets in Escherichia Coli

Frida Svanberg Frisinger
Department of Veterinary and Animal Sciences, Faculty of Health and Medical Sciences, University of Copenhagen, Frederiksberg

Bimal Jana
Department of Veterinary and Animal Sciences, Faculty of Health and Medical Sciences, University of Copenhagen, Frederiksberg

Stefano Donadio
Naicons Srl

Luca Guardabassi (lg@sund.ku.dk)
Department of Veterinary and Animal Sciences, Faculty of Health and Medical Sciences, University of Copenhagen, Frederiksberg

Research Article

Keywords: Escherichia coli, Treatment of infections, silico, Klebsiella

Posted Date: January 4th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-131560/v1

Abstract

Treatment of infections caused by *Escherichia coli* and other Enterobacteriaceae often requires broad-spectrum antimicrobials, which cause perturbations of the gut microbiota (dysbiosis). Novel antimicrobial drugs interfering with pathogen-specific targets would minimize the risk of such dysbiosis. Here, we employed an *in silico* approach to identify essential proteins in *E. coli*, including pathogenic ST131, that are either absent or have low homology to humans and beneficial taxa of the gut microbiota. We identified 37 potential new targets with little or no homology to the proteomes seven taxa representative of the healthy gut microbiota. The suitability of these proteins as drug targets was further analysed through essentiality and conservation in the closely related pathogen *Klebsiella pneumoniae*. None of them are targets of commercially used antibiotics. Eighteen proteins are involved in four functionally connected essential biological processes (replication, chromosome segregation, cell division, and outer membrane biogenesis). Our results indicate that it may be possible to selectively interfere with essential biological processes in Enterobacteriaceae that are absent or mediated by unrelated proteins in beneficial bacterial taxa residing in the gut. The identified targets can be used to discover antimicrobial drugs that are effective against these opportunistic pathogens with a decreased potential of causing dysbiosis.

Introduction

Due to the worldwide increase in resistance observed among certain bacterial pathogens, there is a pressing need for novel antimicrobials. Most of the antimicrobial drugs approved for human use since the end of the antibiotic golden age in the 1960s belong to known antimicrobial classes and are primarily active against Gram-positive bacteria[1]. Thus, there is an urgent need for truly new antimicrobial compounds targeting Gram-negative bacteria. In 2017 the World Health Organization released a list of global priority antimicrobial-resistant (AMR) pathogens in order to guide research, discovery and development of new antibiotics[2]. The highest category (‘Priority 1: Critical’) comprises *Acinetobacter baumannii, Pseudomonas aeruginosa* and several Enterobacteriaceae, including *Escherichia coli*. The latter enteric species is one of the most frequent causes of bacterial infections. Specific hyper virulent clones such as *E. coli* O25b sequence type (ST) 131 are responsible for a wide range of extra-intestinal diseases globally, ranging from urinary tract infection to septicaemia and meningitis[3]. Treatment of *E. coli* infections is often complicated by the high frequency of multidrug resistance. In particular, ST131 is a major contributor to the global spread of fluoroquinolone resistance and extended-spectrum b-lactamase-mediated resistance to β-lactams, and is responsible for millions of multidrug-resistant infections each year[4].

Oral antimicrobial therapy impacts the healthy gut microbiota by inducing a loss of beneficial microbes followed by expansion of opportunistic pathogenic bacteria, such as Enterobacteriaceae[5]. This phenomenon, generally referred to as dysbiosis, can in extreme cases lead to life-threatening secondary infections caused by *Clostridioides difficile*[6], which are becoming increasingly common and difficult to

treat. Certain antimicrobial-sensitive taxa residing in the healthy gut microbiota, *e.g. Bacteroidetes* and *Lachnospiraceae* have been shown to provide protection from *C. difficile* infections through colonisation resistance[5]Moreover, the gut-associated taxa *Lactobacillus* and *Bifidobacterium* both contain strains that are associated with beneficial effects on health and are used as probiotics in various food supplements, including *Lactobacillus planetarum, L. paracasei, L. acidophilus, Bifidobacterium infantis, B. longum,* and *B. breve*[7].

Pathogen-targeted antimicrobial drugs with limited effect on beneficial organisms could potentially decrease the risk of dysbiosis and antibiotic-induced secondary infections. One approach to discover such drugs is the employment of target-based assays to identify compounds that selectively interfere with the viability of *E. coli* and other Enterobacteriaceae without affecting the healthy gut microbiota. This requires identification of targets that are specific for this bacterial family. Here, we performed an *in silico* study to identify protein drug targets in *E. coli* that: i) are present in pathogenic *E. coli* ST131; ii) do not display significant homology to proteins in the host; iii) are absent or have low homology in selected members of the healthy gut microbiota. Furthermore, the selected targets were analysed for their conservation and essentiality in the closely related pathogen *K. pneumoniae*. Finally, the "druggability" of the selected targets was assessed based on subcellular localisation (SCL), availability of three-dimensional structures and presence of known inhibitors.

Methods

Homology searches

The GenBank record for *E. coli* BW25113 (GenBank: CP009273.1) was downloaded, and the protein sequences from the 358 genes found to be essential by Goodall *et al.*[8] were extracted. Five genes were removed due to being labelled as pseudogenes (*ttcC*, *yedN*, *ygeF*, *ygeN*) or putative protein (*yddL*).

To establish presence of proteins in *E. coli* O25b:H4-ST131 (NCBI:txid941322), NCBI+ BLASTp was used to BLAST the 353 protein sequences against this organism using the non-redundant BLAST database. Percent alignment was calculated by dividing the length of the hit by the length of the query protein, extracted from the NCBI record. A dual cut-off was used, where hits E-value $\leq 10E^{-10}$, or percent ID $\geq 70\%$ and percent alignment $\geq 75\%$ were excluded. These cut-offs were selected to be equal to, or more stringent than those used in previous *in silico* studies[52,53,56,57]. Any hits below the first or above the second cut-off were removed, and the remaining were taken on to the next step.

To find human analogues, the NCBI+ command line remote BLAST tool was used to BLAST the remaining protein sequences using first the Entrez queries ‘Homo sapiens [Organism]’ against the non-redundant database, and the output was sorted using the same cut-offs as described above.

Presence in beneficial gut taxa

To establish presence in the simplified gut microbiota, remote BLASTp command line applications were used to search in specific gut taxa using the Entrez queries 'Faecalibacterium [Organism]', 'Bacteroides [Organism], 'Ruminococcus [Organism]', 'Prevotella [Organism]', 'Lactobacillus [Organism]', 'Lachnospiraceae [Organism]' and 'Bifidobacterium [Organism]' using the non-redundant database. These filters provided 274,718 different entries as listed in Supplementary Materials 3-9. The results were downloaded, and analysed using the same methodology and cut-offs as described above.

All hits sorted as above cut-off in the homology search against the simplified microbiota were collected and the number of hits for each protein manually evaluated. A table with the number of hits, together with the information for the highest scoring hit for each protein was generated. The proteins with less than one hits were manually inspected, and proteins with high scoring similarity were removed.

Presence and essentiality in *K. pneumoniae*

To find conservation in *K. pneumoniae* command line applications for BLASTp was used to BLAST remotely against 'Klebsiella pneumoniae subsp. pneumoniae KPNIH1 [Organism]' using the non-redundant database. The supplementary dataset generated by Ramage *et al.*[17] was downloaded and used to search for the potential target genes using gene names.

Subcellular localisation and structural information

The SCL for each protein was manually checked by querying the UniProt/SwissProt database, and retrieving the information found under 'Subcellular Localisation'.

Information about 3D structure for proteins in *E. coli* K-12 was manually retrieved through the UniProt/SwissProt entries for each protein individually. The PDB accession number, molecules in complex and resolution were recorded.

Dataavailability

The raw data generated during the current study are available from the corresponding author on reasonable request.

Results

Essential genes in the target pathogen

Due to the lack of understanding of the essential genome in pathogenic strains such as ST131, the model strain BW25113 was used as a basis for our study. The predicted amino acid sequences of 353 out of the 358 essential genes identified by Goodall *et al.*[8] were retrieved (Figure 1). The remaining five genes (*ttcC*, *yddL*, *yedN*, *ygeF*, *ygeN*), all labelled 'pseudogenes' or 'putative protein' and were not found in the Keio collection[9], were excluded from further analysis. Their assignment as essential by Goodall *et al.*[8] may be thus an artefact of the methodology employed in the original study.

The sequences of the 353 proteins were compared to those found in *E. coli* O25b:H4-ST131 using a pre-defined cut-off (E-value ≤ $10E^{-10}$ or ≥ 70% sequence identity, and ≥ 75% alignment length, see Materials and Methods). All of the 353 essential BW25113 proteins were associated with at least one hit in ST131, apart from YqeL (Supplementary Materials 1). However, 15 proteins were excluded as they scored below the cut-off threshold. Inspection of these sequences revealed that many of them were prophage-related or uncharacterized proteins. The presence of phages in a bacterial genome is expected to vary with the specific strain history, and this may explain the observed difference between the laboratory strain BW25113 and the epidemic clonal lineage ST131. This left 337 essential and conserved *E. coli* proteins in the pipeline for further analyses.

Homology to proteins in mammalian hosts

The second step of the analysis aimed at removing *E. coli* targets homologous to the human proteome. A high degree of similarity between the pathogen's target and one or more proteins in the host proteome may result in off-target binding of a drug, leading to toxicity and unwanted side effects. The 337 selected essential proteins were therefore compared to the human proteome, leading to 186 proteins fulfilling the same stringent cut-offs as above (Figure 1, Supplementary Materials 2).

Homology to proteins in beneficial taxa of the gut microbiota

The next step in the selection pipeline aimed to exclude proteins with high similarity to those found in representatives of the beneficial gut microbiota. Given the complexity and variability of the gut microbiome, we decided to focus on seven taxa containing species have previously been shown to have beneficial and protective effects on the host[5,6,10–16]: *Faecalibacterium*, *Prevotella*, *Ruminococcus,Bacteroides, Lactobacillus, Lachnospiraceae* and *Bifidobacterium* (Supplementary Materials 3-9). The 186 proteins were blasted against the abovementioned taxa using the same cut-off values as before (Figure 1). As expected, this step was the most selective, leaving just 31 proteins to further analysis (Table 1) and removed all targets of commercially available antibiotics, including ParC (target of fluoroquinolones), FtsI, MrdA (targets of b-lactams), parts of the 30S and 50S ribosome (targets of macrolides, aminoglycosides tetracyclines) and RNA polymerase (target of rifamycins).

Among the identified 31 proteins, only PheM and TrpL, each encoding a leader peptide in the Phe tRNA synthetase and Trp biosynthetic operon, were found to be missing completely in all taxa. No hits for YobI (a protein of unknown function) were found in any of the taxa apart from *Faecalibacterium*, where one single hit was found (E-value 5.1, 61.9 % alignment and 69.2% id). SafA (part of the low pH stress response) was found to be missing in *Lachnospiraceae*, *Bifidobacterium* and *Faecalibacterium.* Furthermore, WzyE (probable ECA polymerase), MreD (rod shape determining protein), LolA and LolB (both part of the lipoprotein transport pathway) lacked hits in *Bifidobacterium*, FtsL (a cell division protein) and MukF (involved in chromosome partition) in *Bacteroides* and YciS (lipopolysaccharide assembly protein A) in *Faecalibacterium*. All other proteins were associated with hits below the cut-off in all taxa.

Due to the high stringency applied in the above step, potentially valuable targets may have been missed in the selection process. Thus, a second analysis of the microbiota BLAST results was undertaken to find proteins associated with only a few hits over cut-off. Six proteins (BamD, YfgZ, HolA, YrfF, LptD and ZipA) were each found to be associated with one hit only and had all been excluded based on E-value cut-off rather than sequence identity. Thus, these six proteins were included in further analyses, leading to 37 proteins as potential *E. coli*-selective targets (Table 1).

Target conservation in *K. pneumoniae*

We evaluated presence and essentiality of the selected targets in *K. pneumoniae* KPNIH1, another global priority pathogen closely related to *E. coli*. The essentiality of the 37 proteins was checked against the library generated by Ramage *et al.*[17], together with conservation of the amino acid sequence as above (Figure 1, Supplementary Materials 10). Eighteen were found to be essential in both organisms, all displaying high sequence conservation, apart from ZipA and HipB. However, the majority of the targets not reported to be essential in *K. pneumoniae* fulfilled the selection criteria, apart from HigA, IraM, SafA, YobI and TrpL (Table 1).

Biological function of selected targets

Of the 37 identified targets (Table 1), several were found to share or have similar biological functions (Figure 2). One of the largest groups comprise of the proteins involved in outer membrane (OM) biogenesis and maintenance (Figure 2). Here, BamD is directly associated with the OM, and is part of the b-barrel assembly machinery (BAM). LptA, LptD, LptE and LptF are all part of the lipopolysaccharide (LPS) transport (Lpt) machinery. LolA and LolB are found in the periplasmic space and the periplasmic side of the OM respectively, and belong to the lipoprotein transport machinery responsible for delivering OM lipoproteins to all three of the OM assembly machineries (LOL, BAM and LPT)[18]. SecE is part of the SecYEG protein translocation machinery responsible for transporting proteins into the periplasm[19], and PssA is involved in phospholipid biosynthesis[20]. Furthermore, the inner membrane-protein YciS (also known as LapA, lipopolysaccharide assembly protein A) is part of a machinery responsible for envelope stress-response and regulation of LPS production[21]. Finally, although not associated with OM maintenance, TonB is part of the machinery responsible for actively importing iron across the OM in the cell[22] (Figure 2).

Another two functional groups comprise of the proteins responsible for DNA replication (HolA, HolD, PriB, DnaT and YgfZ) and cell division (MukB, MukE, MukF, FtsB, FtsL, FtsQ and ZipA) (Figure 2). DNA replication is a tightly controlled mechanism and DNA Polymerase III holoenzyme is the major replication complex in *E. coli*, where both HolA (d subunit) and HolD (y subunit) make up parts of the clamp loading complex[23]. DNA damage can cause this machinery to be stalled and disassemble on the chromosome, leading to replication failure. To re-start replication the cell must make use of the replication restart primosome, where both the PriB helicase and DnaT primase are found[24]. YgfZ has been shown to be part of the system regulating chromosomal replication[25]. MukBEF are unique to the g-proteobacteria and are

involved in cell division, making up the only *E. coli* condensin for chromosome replication, segregation and organisation[26]. Further downstream in this process the transmembrane complex FtsBL is found[27], together with FtsQ[28] and ZipA. In a related process, MreD is involved in determining cell shape[29]

Among the proteins in the stress response category, CydX (Figure 2) is part of the CydAB cytochrome *bd* oxidase complex involved in aerobic respiration and maintaining the charge across the membrane used for synthesizing ATP[30]. IraM is a regulator of s^S, the stationary phase sigma factor responsible for controlling expression of a plethora of genes involved in stress response[31]. Although the exact function of YobI has not yet been established, it has been shown to accumulate upon heat shock[32].

Among the biosynthetic genes, WzyE has been implicated to be involved in assembly of the enterobacterial common antigen[33], TrpL is involved in controlling tryptophan biosynthesis[34] and HemD is a uroporphyrinogen III synthase[35] (Figure 2).

HipB and HigA together make up the category of anti-toxins of the Type II Toxin-Antitoxin system, and work to counteract the effect of their cognate toxins[36]. As the sole members of their functional groups PheM is a target of transcriptional regulation (Figure 2) and is responsible for attenuation of the phenylalanyl-tRNA synthetase[37], while SafA is a two-component system connector[38].

Finally, no information regarding biological function could be found for the three proteins YcaR, YrfF and YdhL (Figure 2).

Target localisation

An essential requirement for to develop an efficient antimicrobial drug is target access. This is especially important in Gram-negative bacteria, where the double membrane structure acts as a permeability barrier, efficiently blocking many compounds from accessing intracellular targets. Subcellular localisation (SCL) was therefore considered to evaluate protein's druggability. Swiss-Prot, the manually annotated section of UniProtKB, was used to find information on SCL for each of the 37 selected proteins (Figure 2, Table 1). The target proteins were found to be located in either the Inner Membrane (IM), Outer Membrane, Cytoplasm, Nucleoid or Periplasm (Figure 2). Notably PssA was annotated as located in both the IM and the cytoplasm. For eleven proteins (PriB, HemD, DnaT, HolA, HolD, YdhL, HigA, HipB, PheM, TrpL and YobI), no SCL had been experimentally determined. Here, the four OM associated proteins (LptD, LptE, LolB and BamD) are promising potential targets, especially LptD, which contains extracellular domains.

Existence of known inhibitors

Next, the literature was searched for previously reported inhibitors of the selected targets. As expected, none of the targets presented in Table 1 are inhibited by commercially available antibiotics. Through analysis of scientific literature we were able to identify inhibitors targeting a few of the listed targets but, to our knowledge, none has gone beyond laboratory studies: the ZipA/FtsZ interaction has been reported to be inhibited by certain antimicrobial compounds[39,40]; the insect peptide Thanatin blocks LptA[41];

compound IMB-881 blocks the interaction between LptA and LptC[42]; JB-95 inhibits b-barrel proteins including LptD[43]; MAC13243 inhibits LolA[44]; BamD is inhibited by an inhibitory peptide[45] while the compound IMB-H4 has been shown to block BamA-BamD interaction[46], and MukB is inhibited by the small molecules Michellamine B and NSC260594[47]. Finally, multiple inhibitory compounds targeting TonB have been identified[48–50]. Thanatin has been shown to possess antimicrobial activity against several Gram-negative bacteria beyond *E. coli*, including *K. pneumoniae*, *Salmonella typhimurium* and *Enterobacter cloacae*[41]. IMB-H4 was also able to inhibit growth in *K. pneumoniae*, *P. aeruginosa* and *A. baumannii*[46]. NSC176319 was found to be active against *S. aureus* and permeabilised *P. aeruginosa* and *A. baumannii*[47]. JB-95 was reported to have antimicrobial activity against *A. baumannii*, *P. aeruginosa* and *Staphylococcus aureus*[43], MAC13243 has been shown to also be active against *P. aeruginosa*[44] and TonB inhibition has been shown to affect *A. baumanii*[49]. With the information provided in this study, some of these inhibitors may represent starting scaffolds for development into pathogen-specific antibacterials. In addition, they might represent useful tools in validating future target-based assays.

Target structure

Structure-guided drug design is a powerful *in silico* approach that can rapidly screen millions of compounds for their ability to dock into a desired target, and identified hits can subsequently be tested *in vitro*. Thus, 3D structures at a high enough resolution represent an advantage for the targets identified in this study.

Information retrieved from the Protein Data Bank (PDB)[51] showed that 3D structures at a resolution of <3 Å existed for 18 proteins, >3 Å for 6 proteins and no structure could be found for 3 proteins, while YrfF was associated with a structure but no resolution information was reported in the database, and no structure had been reported for the remaining nine protein targets (Table 1).

Discussion

The originality of the present study lies in the identification of cellular targets that may lead to the discovery of innovative pathogen-selective antimicrobial drugs with limited effect on the healthy gut microbiota. Similar *in silico* studies have previously been conducted for verotoxigenic *E. coli* O157:H7, *K. pneumoniae, Yersinia pseudotuberculosis* and Enterobacteriaceae[52–56]. However, these studies were not designed to identify targets with low homology to the corresponding proteins in beneficial taxa residing in the intestinal tract or suffered from limitations related to the lack of a well-established essential genome for the target pathogen, or the comparisons to the human proteome.

We identified 37 potential drug targets selective for *E. coli* based on protein sequence homology. A large proportion of the identified proteins are functionally related between themselves. For example, MukBEF and FtsBL together with FtsQ and ZipA are all involved in regulation of the cell cycle; HolA, HolD and DnaT and PriB are part of two different complexes that are both part of the replication system in *E. coli;* PheM belongs to the separate, but related category of transcription regulation. OM maintenance was

another large category, which is unsurprising as this structure is unique to Gram-negative bacteria while the majority of the bacteria used in this study as representatives of the healthy gut microbiota are Gram-positive. Targets belonging to this category include proteins directly associated with the OM (LptD, LptE, BamD, LolB) as well as proteins found in various IM and periplasm related processes (LptA, LptF, LolA, SecE, YciS, TonB, MreD). The fact that multiple proteins from a single pathway were identified indicates that this is potentially a good cellular function to target. A number of proteins were found to be uniquely involved in various biosynthetic processes (HemD, PssA, TrpL and WzyE), stress response (CydX, IraM and YobI), or toxin-antitoxin systems (HigA and HipB). Proteins involved in two-competent system (SafA) or tRNA processing (YgfZ) were also found. Finally, four proteins without clear functions were identified (YcaR, YdhL, YgfZ and YrfF), indicating that there is more to discover regarding *E. coli* biology.

When searching for homologues in seven representative taxa of the healthy gut microbiota, only TrpL and PheM lacked hits in all these groups, indicating that these structures are unique to *E. coli* and possibly other opportunistic pathogenic bacteria residing in the intestinal tract. However, no info on SCL and 3D structure is available for either of these two proteins and only PheM is present and highly conserved in *K. pneumoniae*, although it is not proven to be essential in this species. Several proteins lacked hits in one or multiple taxa: YobI was only found in *Faecalibacterium*; SafA was missing *Lachnospiraceae*, *Bifidobacterium* and *Faecalibacterium*; no hits for WzyE, MreD, LolA, LolB and FtsL were found in *Bifidobacterium*; FtsL and MukF were missing in *Bacteroides*, and finally *Faecalibacterium* was lacking hits for YciS. These proteins may also be appropriate targets for developing pathogen-targeted antibiotics due to their absence in some gut beneficial bacteria. The remaining proteins were associated with hits below the cut-off, indicating that a similar protein may exist but that may be divergent enough that off-target activity can be avoided.

The potential to target an infecting pathogen without affecting the beneficial microbiota makes *E. coli* -selective antibiotics especially attractive. The clinical value of such antibiotics would be even higher if their spectrum covered other pathogenic bacterial species such. *K. pneumoniae* is a close relative to *E. coli*, and a major source of MDR infections in health care settings. The potential target proteins found in *E. coli* were therefore checked for sequence similarity and essentiality in *K. pneumoniae*. While the majority of the proteins were highly conserved in both species, only 18 out of the 37 targets were found to be essential in this pathogen. This could be due to disparity of information since the *E. coli* essential genome is well characterized, whereas the one in *K. pneumoniae* has been less studied. However, these 18 proteins are good candidates for developing pathogen-selective antibiotics that target both species.

An issue regarding essentiality in *K. pneumoniae* is its reliance on a relatively uncharacterised genome. The essentiality status is defined by a single study, which may be affected by different parameters than the one used to identify the *E. coli* 'essentialome'. The lack of a well-characterised essential genome in *K. pneumoniae* leads to limitations in this selection step, as misclassification in terms of essentiality can occur and due to the scarcity of information and verification through consensus from multiple studies was not possible here. However, sequence conservation indicates that majority of proteins are highly conserved across the two species. Due to the close relativeness between the two species, investigations

into the essentiality of the highly conserved but reported non-essential proteins may be worthwhile to further characterise their suitability as drug targets.

Inhibitors interfering with the proteins identified here were only found for seven of the selected targets: ZipA (Compound 10b, Compound 4)[39,40], LptA (Thanatin, IMB-881)[41,42], LptD (JB-95)[43], LolA (MAC13243)[44], BamD (peptide and IMB-H4)[45,46], MukB (Michellamine B and NSC176319)[47], and TonB (multiple small bacteriostatic molecules and inhibitory compounds)[48–50]. None of these are currently developed for clinical use, however the presence of inhibitors indicates that the targets are druggable. Furthermore, many of the identified inhibitors target functions related to OM related processes, further strengthening these functions as viable for development of antimicrobials.

Due to the double-membrane structure found in Gram-negative bacteria, the accessibility of a potential drug target is essential. Proteins found in the OM are therefore considered to be targets with extra potential. Here, we identified LptD, LptE, LolB and BamD as OM associated proteins. Out of these proteins, LptD also has extracellular domains, indicating that it may be possible to find an inhibitor that interferes with these portions and which would therefore not have to cross the OM. Interestingly, multiple of the OM proteins have already been targeted by various inhibitors, indicating that this is a viable strategy when developing novel antimicrobials. Notably, the four proteins listed above have an excellent 'druggability' potential since in addition to their optimal SCL, they are also essential in *K. pneumoniae* and have known 3D structures. Furthermore, all proteins were shown to be essential in *K. pneumoniae* and share a high degree of sequence similarity with this species, while they were found to only share low levels of sequence identity with the selected gut taxa used here. Here, LolB and LptE fell below the cut-offs used with alignment similarity $\leq$ 16%, and percent shared identity < 49%. Both LptD and BamD were recovered in the manual inspection of the proteins that were excluded due to falling above one of the cut-offs. However, both proteins were found to be excluded due to their E-value, and only be associated with one hit each: LptD in *Bacteroides* (12.5% alignment, 100% id and E-value 6,03E-62) and BamD in *Lactobacillus* (55.9% alignment, 24.8% id and E-value 6.06E-10). The selectivity of BamD and LptD can be further evaluated by future *in vitro* studies using known inhibitors that specifically interfere with these proteins.[43,45,46]

All *in silico* studies suffer from the drawback of cut-off criteria that may have little biological relevance. Stringent cut-off values potentially exclude valuable drug targets while loose criteria may result in an unmanageable list of targets. In the present study, while keeping stringent cut-offs throughout, we were able to identify 31 proteins present only in a selected number of beneficial gut microbes and worthy of further investigations. However, a manual analysis of the rejected sequences showed that six proteins could be recovered and that may be important targets for future drug development efforts.

Another issue related to *in silico* studies is that essentiality may differ between *in vitro* and *in vivo* conditions. The essential genome established by Goodall *et al.*[8] was characterized in rich media conditions, and therefore may miss genes essential for metabolism inside the host[8,57]. Certain biosynthetic pathways may be downregulated as the pathogen instead relies on the host to supply

nutrients such as amino acids, vitamins and nucleobases[57]. However, targeting biosynthetic pathways involved in maintenance of the cell is likely to represent a target relevant *in vitro* as well as *in vivo*[58].

In silico studies like this are a first but essential step towards the discovery of novel pathogen-targeted antimicrobials. The results of our study provide a starting point towards the identification and development of novel specific antimicrobials targeting *E. coli*. Future wet-lab studies are required to validate the presumptive selective targets identified by the study. High-throughput screens can be applied to find inhibitors interfering with the specific protein targets *e.g.* through knock-down strains with reduced expression of the target protein. The antimicrobial activity of the identified inhibitors could subsequently be evaluated on a comprehensive strain collection representative of the healthy gut microbiota or directly on faecal samples using a metagenomics approach in order to assess their selective toxicity towards *E. coli* and other pathogenic Enterobacteriaceae.

Declarations

Acknowledgements

We would like to thank Bastian V.H. Hornung and Defne Surjoun for their kind help with development of the Python scripts and pipelines. FSF is a recipient of a PhD fellowship from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie Grant agreement no. 765147.

Author contributions

FSF, LG and SD designed the study. FSF conducted the experiments. FSF, LG, SD and BJ analysed the results. All authors reviewed the manuscript.

Additional information

Competing interests

The author(s) declare no competing interests.

References

1. Lewis, K. Platforms for antibiotic discovery. *Nat. Rev. Drug Discov.* **12**, 371–387 (2013).
2. World health organization releases global priority list of antibiotic-resistant bacteria to guide research, discovery, and development of new antibiotics. **32**, 76 (2018).
3. Köhler, C. D. & Dobrindt, U. What defines extraintestinal pathogenic Escherichia coli? *Int. J. Med. Microbiol.* **301**, 642–647 (2011).
4. Pitout, J. D. D. & DeVinney, R. Escherichia coli ST131: A multidrug-resistant clone primed for global domination. *F1000Research* **6**, 1–7 (2017).

5. Buffie, C. G. & Pamer, E. G. Microbiota-mediated colonization resistance against intestinal pathogens. *Nat. Rev. Immunol.* **13**, 790–801 (2013).
6. Rodriguez, C., Taminiau, B., Van Broeck, J., Delmée, M. & Daube, G. Clostridium difficile infection and intestinal microbiota interactions. *Microb. Pathog.* **89**, 201–209 (2015).
7. Hills, R. D. *et al.* Gut microbiome: Profound implications for diet and disease. *Nutrients* **11**, 1–40 (2019).
8. Goodall, E. C. A. *et al.* The Essential Genome of Escherichia coli K-12. *MBio* **9**, 1–18 (2018).
9. Baba, T. *et al.* Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: The Keio collection. *Mol. Syst. Biol.* **2**, (2006).
10. Tamanai-Shacoori, Z. *et al.* Roseburia spp.: a marker of health? *Future Microbiol.* **12**, 157–170 (2017).
11. Naaber, P. *et al.* Inhibition of Clostridium difficile strains by intestinal Lactobacillus species. *J. Med. Microbiol.* **53**, 551–554 (2004).
12. Ferreira-Halder, C. V., Faria, A. V. de S. & Andrade, S. S. Action and function of Faecalibacterium prausnitzii in health and disease. *Best Pract. Res. Clin. Gastroenterol.* **31**, 643–648 (2017).
13. Precup, G. & Vodnar, D.-C. Gut Prevotella as a possible biomarker of diet and its eubiotic versus dysbiotic roles: a comprehensive literature review. *Br. J. Nutr.* **122**, 131–140 (2019).
14. Hsiao, A. *et al.* Members of the human gut microbiota involved in recovery from Vibrio cholerae infection. *Nature* **515**, 423–426 (2014).
15. Vacca, M. *et al.* The controversial role of human gut lachnospiraceae. *Microorganisms* **8**, 1–25 (2020).
16. Grimm, V., Westermann, C. & Riedel, C. U. Bifidobacteria-Host Interactions - An Update on Colonisation Factors. *Biomed Res. Int.* **2014**, (2014).
17. Ramage, B. *et al.* Comprehensive Arrayed Transposon Mutant Library of Klebsiella pneumoniae Outbreak Strain KPNIH1. *J. Bacteriol.* **199**, 1–9 (2017).
18. Grabowicz, M. Lipoprotein Transport: Greasing the Machines of Outer Membrane Biogenesis. *BioEssays* **40**, 1700187 (2018).
19. Lycklama A Nijeholt, J. A., De Keyzer, J., Prabudiansyah, I. & Driessen, A. J. M. Characterization of the supporting role of SecE in protein translocation. *FEBS Lett.* **587**, 3083–3088 (2013).
20. Matsumoto, K. Phosphatidylserine synthase from bacteria. *Biochim. Biophys. Acta - Lipids Lipid Metab.* **1348**, 214–227 (1997).
21. Klein, G., Kobylak, N., Lindner, B., Stupak, A. & Raina, S. Assembly of Lipopolysaccharide in Escherichia coli Requires the Essential LapB Heat Shock Protein. *J. Biol. Chem.* **289**, 14829–14853 (2014).
22. Postle, K. & Larsen, R. A. TonB-dependent energy transduction between outer and cytoplasmic membranes. *BioMetals* **20**, 453–465 (2007).
23. Xu, Z. Q. & Dixon, N. E. Bacterial replisomes. *Curr. Opin. Struct. Biol.* **53**, 159–168 (2018).

24. Huang, Y. H. & Huang, C. Y. Structural insight into the DNA-binding mode of the primosomal proteins PriA, PriB, and AnaT. *Biomed Res. Int.* **2014**, (2014).
25. Ote, T. *et al.* Involvement of the Escherichia coli folate-binding protein YgfZ in RNA modification and regulation of chromosomal replication initiation. *Mol. Microbiol.* **59**, 265–275 (2006).
26. Rybenkov, V. V., Herrera, V., Petrushenko, Z. M. & Zhao, H. MukBEF, a Chromosomal Organizer. *J. Mol. Microbiol. Biotechnol.* **24**, 371–383 (2014).
27. Condon, S. G. F. *et al.* The FtsLB subcomplex of the bacterial divisome is a tetramer with an uninterrupted FtsL helix linking the transmembrane and periplasmic regions. *J. Biol. Chem.* **293**, 1623–1641 (2018).
28. Buddelmeijer, N. & Beckwith, J. A complex of the Escherichia coli cell division proteins FtsL, FtsB and FtsQ forms independently of its localization to the septal region. *Mol. Microbiol.* **52**, 1315–1327 (2004).
29. Kruse, T., Bork-Jensen, J. & Gerdes, K. The morphogenetic MreBCD proteins of Escherichia coli form an essential membrane-bound complex. *Mol. Microbiol.* **55**, 78–89 (2005).
30. Vanorsdel, C. E. *et al.* The escherichia coli CydX protein is a member of the CydAB cytochrome bd oxidase complex and is required for cytochrome bd oxidase activity. *J. Bacteriol.* **195**, 3640–3650 (2013).
31. Bougdour, A., Cunning, C., Baptiste, P. J., Elliott, T. & Gottesman, S. Multiple pathways for regulation of σS (RpoS) stability in Escherichia coli via the action of multiple anti-adaptors. *Mol. Microbiol.* **68**, 298–313 (2008).
32. Hemm, M. R. *et al.* Small stress response proteins in Escherichia coli: Proteins missed by classical proteomic studies. *J. Bacteriol.* **192**, 46–58 (2010).
33. Kajimura, J., Rahman, A. & Rick, P. D. Assembly of cyclic enterobacterial common antigen in Escherichia coli K-12. *J. Bacteriol.* **187**, 6917–6927 (2005).
34. Zurawski, G., Elseviers, D., Stauffer, G. V. & Yanofsky, C. Translational control of transcription termination at the attenuator of the Escherichia coli tryptophan operon. *Proc. Natl. Acad. Sci. U. S. A.* **75**, 5988–5992 (1978).
35. Jordan, P. M., Mgbeje, B. I., Thomas, S. D. & Alwan, A. F. Nucleotide sequence for the hemD gene of Escherichia coli encoding uroporphyrinogen III synthase and initial evidence for a hem operon. *Biochem. J.* **249**, 613–616 (1988).
36. Fraikin, N., Goormaghtigh, F. & Van Melderen, L. Type II Toxin-Antitoxin Systems: Evolution and Revolutions. *J. Bacteriol.* **202**, 1–14 (2020).
37. Springer, M. *et al.* Attenuation control of the Escherichia coli phenylalanyl-tRNA synthetase operon. *J. Mol. Biol.* **181**, 467–478 (1985).
38. Ishii, E., Eguchi, Y. & Utsumi, R. Mechanism of activation of PhoQ/PhoP two-component signal transduction by SafA, an auxiliary protein of PhoQ histidine kinase in escherichia coli. *Biosci. Biotechnol. Biochem.* **77**, 814–819 (2013).

39. Tsao, D. H. H. *et al.* Discovery of novel inhibitors of the ZipA/FtsZ complex by NMR fragment screening coupled with structure-based design. *Bioorganic Med. Chem.* **14**, 7953–7961 (2006).
40. Jennings, L. D. *et al.* Design and synthesis of indolo[2,3-a]quinolizin-7-one inhibitors of the ZipA-FtsZ interaction. *Bioorganic Med. Chem. Lett.* **14**, 1427–1431 (2004).
41. Vetterli, S. U. *et al.* Thanatin targets the intermembrane protein complex required for lipopolysaccharide transport in Escherichia coli. *Sci. Adv.* **4**, 1–9 (2018).
42. Zhang, X. *et al.* Identification of an anti-Gram-negative bacteria agent disrupting the interaction between lipopolysaccharide transporters LptA and LptC. *Int. J. Antimicrob. Agents* **53**, 442–448 (2019).
43. Urfer, M. *et al.* A peptidomimetic antibiotic targets outer membrane proteins and disrupts selectively the outer membrane in Escherichia coli. *J. Biol. Chem.* **291**, 1921–1932 (2016).
44. Pathania, R. *et al.* Chemical genomics in Escherichia coli identifies an inhibitor of bacterial lipoprotein targeting. *Nat. Chem. Biol.* **5**, 849–856 (2009).
45. Hagan, C. L., Wzorek, J. S. & Kahne, D. Inhibition of the β-barrel assembly machine by a peptide that binds BamD. *Proc. Natl. Acad. Sci.* **112**, 2011–2016 (2015).
46. Li, Y. *et al.* Identification of a Compound That Inhibits the Growth of Gram-Negative Bacteria by Blocking BamA–BamD Interaction. *Front. Microbiol.* **11**, 1–13 (2020).
47. Zhao, H. *et al.* Small Molecule Condensin Inhibitors. *ACS Infect. Dis.* **4**, 1737–1745 (2018).
48. Yep, A., McQuade, T., Kirchhoff, P., Larsen, M. & Mobley, H. L. T. Inhibitors of TonB Function Identified by a High-Throughput Screen for Inhibitors of Iron Acquisition in Uropathogenic Escherichia coli CFT073. *MBio* **5**, 1–10 (2014).
49. Nairn, B. L. *et al.* Fluorescence High-Throughput Screening for Inhibitors of TonB Action. *J. Bacteriol.* **199**, 1–20 (2017).
50. Hanson, M., Jordan, L. D., Shipelskiy, Y., Newton, S. M. & Klebba, P. E. High-Throughput Screening Assay for Inhibitors of TonB-Dependent Iron Transport. *J. Biomol. Screen.* **21**, 316–322 (2016).
51. Berman, H. M. *et al.* The protein data bank. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **58**, 899–907 (2002).
52. Ramos, P. I. P. *et al.* An integrative, multi-omics approach towards the prioritization of Klebsiella pneumoniae drug targets. *Sci. Rep.* **8**, 10755 (2018).
53. Hadizadeh, M. *et al.* Genome-Wide Identification of Potential Drug Target in Enterobacteriaceae Family: A Homology-Based Method. *Microb. Drug Resist.* **24**, 8–17 (2018).
54. Georrge, J. J. & Umrania, V. In silico identification of putative drug targets in Klebsiella pneumonia MGH78578. *Indian J. Biotechnol.* **10**, 432–439 (2011).
55. Duffield, M. *et al.* Predicting conserved essential genes in bacteria: in silico identification of putative drug targets. *Mol. Biosyst.* **6**, 2482 (2010).
56. Mondal, S. I. *et al.* Identification of potential drug targets by subtractive genome analysis of Escherichia coli O157:H7: an in silico approach. *Adv. Appl. Bioinforma. Chem.* **8**, 49 (2015).

57. Mobegi, F. M. *et al.* From microbial gene essentiality to novel antimicrobial drug targets. *BMC Genomics* **15**, 958 (2014).
58. Bakheet, T. M. & Doig, A. J. Properties and identification of antibiotic drug targets. *BMC Bioinformatics* **11**, (2010).

Tables

Table 1 The identified potential protein targets with gene name, PDB accession number, RCSB PDB structural information, Swissprot SCL, available inhibitors, sequence conservation and essential status in *K. pneumoniae*. N/A = Not available, N/D = Not determined.

Gene	Protein role	PDB accession no	RCSB PDB	Swissprot SCL	Inhibitor	Conservation in *K. pneumoniae* %alignment, %id *E-value*
bamD	Outer membrane protein assembly factor BamD	P0AC02	5D0O 2.90 Å	Cell outer membrane; Lipid anchor	Inhibitory peptide[45] IMB-H4[46]	100%, 92% *1.06E-172* Essential
cydX	Cytochrome bd-I ubiquinol oxidase subunit X	P56100	6RKO 2.68 Å	Cell inner membrane; Single-pass membrane protein	N/A	97%, 83% *3.77E-18*
dnaT	Primosomal protein 1	P0A8J2	4OU6 1.96 Å	N/D	N/A	100%, 75% *2.03E-99*
ftsB	Cell division protein FtsB	P0A6S5	4IFF 2.30 Å	Cell inner membrane	N/A	102%, 91% *2.32E-58* Essential
ftsL	Cell division protein FtsL	P0AEN4	ND	Cell inner membrane	N/A	100%, 96% *4.39E-84* Essential
ftsQ	Cell division protein FtsQ	P06136	2VH1 2.70 Å	Cell inner membrane; Single-pass type II membrane protein	N/A	96%, 89% *2.82E-176* Essential
hemD	Uroporphyrinogen-III synthase	P09126	ND	N/D	N/A	100%, 80% *2.38E-142* Essential
higA	Antitoxin HigA	P67701	6JQ4 2 Å	N/D	N/A	99%, 46% *4.25E-41*
hipB	Antitoxin HipB	P23873	2WIU 2.35 Å	N/D	N/A	100%, 60% *3.40E-33* Essential
holA	DNA polymerase III subunit delta	P28630	1JR3 2.70 Å	N/D	N/A	100%, 85% *0* Essential
holD	DNA polymerase III subunit psi	P28632	3SXU 1.85 Å	N/D	N/A	99%, 71% *1.05E-70*
iraM	Anti-adapter protein IraM	P75987	ND	Cytoplasm	N/A	100%, 33% *1.10E-16*
lolA	Outer-membrane lipoprotein carrier protein	P61316	1IWL 1.56 Å	Periplasm	MAC13243[44]	100%, 94% *3.34E-144* Essential
lolB	Outer-membrane lipoprotein LolB	P61320	1IWM 1.90 Å	Cell outer membrane; Lipid anchor	N/A	97%, 82% *2.23E-127* Essential
lptA	Lipopolysaccharide export system protein LptA	P0ADV1	2R19 2.16 Å	Periplasm	Thanatin[4]; Compound IMB-8814[2]	94%, 84% *5.15E-99* Essential
lptD	LPS-assembly protein LptD	P31554	4RHB 3.35 Å	Cell outer membrane	Inhibitory peptide JB-95[43]	100%, 82% *0* Essential
lptE	LPS-assembly lipoprotein LptE	P0ADC1	4RHB 3.35 Å	Cell outer membrane	N/A	102%, 70% *8.46E-102* Essential
lptF	Lipopolysaccharide export system permease protein LptF	P0AF98	6MHU 4.10 Å	Cell inner membrane ; Multi-pass membrane protein	N/A	100%, 86% *0*
mreD	Rod shape-determining protein MreD	P0ABH4	ND	Cell inner membrane ; Multi-pass membrane protein	N/A	100%, 93% *2.14E-105*
mukB	Chromosome partition protein MukB	P22523	1QHL 2.20 Å	Nucleoid	NSC176319, Michellamine B[47]	100%, 92% *0* Essential
mukE	Chromosome partition protein MukE	P22524	3EUH 2.90 Å	Nucleoid	N/A	100%, 95% *2.13E-167* Essential
mukF	Chromosome partition protein MukF	P60293	3EUH 2.90 Å	Nucleoid	N/A	100%, 93% *0* Essential
pheM	Phenylalanine--tRNA ligase operon leader peptide	P0AD74	ND	N/D	N/A	100%, 100% 0.000373
priB	Primosomal replication protein N	P07013	5WQV 1.97 Å	N/D	N/A	100%, 92% *9.89E-71*
pssA	CDP-diacylglycerol--serine O-phosphatidyltransferase	P23830	ND	Cytoplasm; Cell inner membrane; Peripheral membrane protein; Cytoplasmic side	N/A	96%, 87% *0* Essential
safA	Two-component-system connector protein SafA	P76136	ND	Cell inner membrane; Single-pass type II membrane protein	N/A	69%, 40% *0.23*
secE	Protein translocase subunit SecE	P0AG96	5GAE 3.33 Å	Cell inner membrane; Multi-pass membrane protein	N/A	100%, 94% *6.30E-82* Essential
tonB	Protein TonB	P02929	2GRX 3.30 Å	Cell inner membrane; Single-pass membrane protein; Periplasmic side	Bacteriostatic compounds[48]; inhibitory compounds[49,50]	101%, 72% *3.52E-95*
trpL	trp operon leader peptide	P0AD92	ND	N/D	N/A	N/A
ycaR	UPF0434 protein YcaR	P0AAZ7	ND	Cytoplasm	N/A	100%, 85% *4.78E-18*
yciS	Lipopolysaccharide assembly protein A	P0ACV4	ND	Cell inner membrane	N/A	100%, 86% *2.14E-59*
ydhL	Uncharacterized protein YdhL	P64474	ND	N/D	N/A	100%, 81% *4.14E-38*
ygfZ	tRNA-modifying protein YgfZ	P0ADE8	1NRK 2.80 Å	Cytoplasm	N/A	100%, 82% *0*
yobI	Uncharacterized protein YobI	C1P604	ND	N/D	N/A	67%, 43% *4.1*
yrfF	Putative membrane protein IgaA homolog	P45800	4UZM Resolution N/A	Cell inner membrane; Multi-pass membrane protein	N/A	100%, 73% *0*
wzyE	Probable ECA polymerase	P27835	ND	Cell inner membrane	N/A	100%, 87% *0*
zipA	Cell division protein ZipA	P77173	1F46 1.50 Å	Cell inner membrane; Single-pass type I membrane protein	Antimicrobial compounds[39,40]	30%, 83% *1.53E-34* Essential

Figures

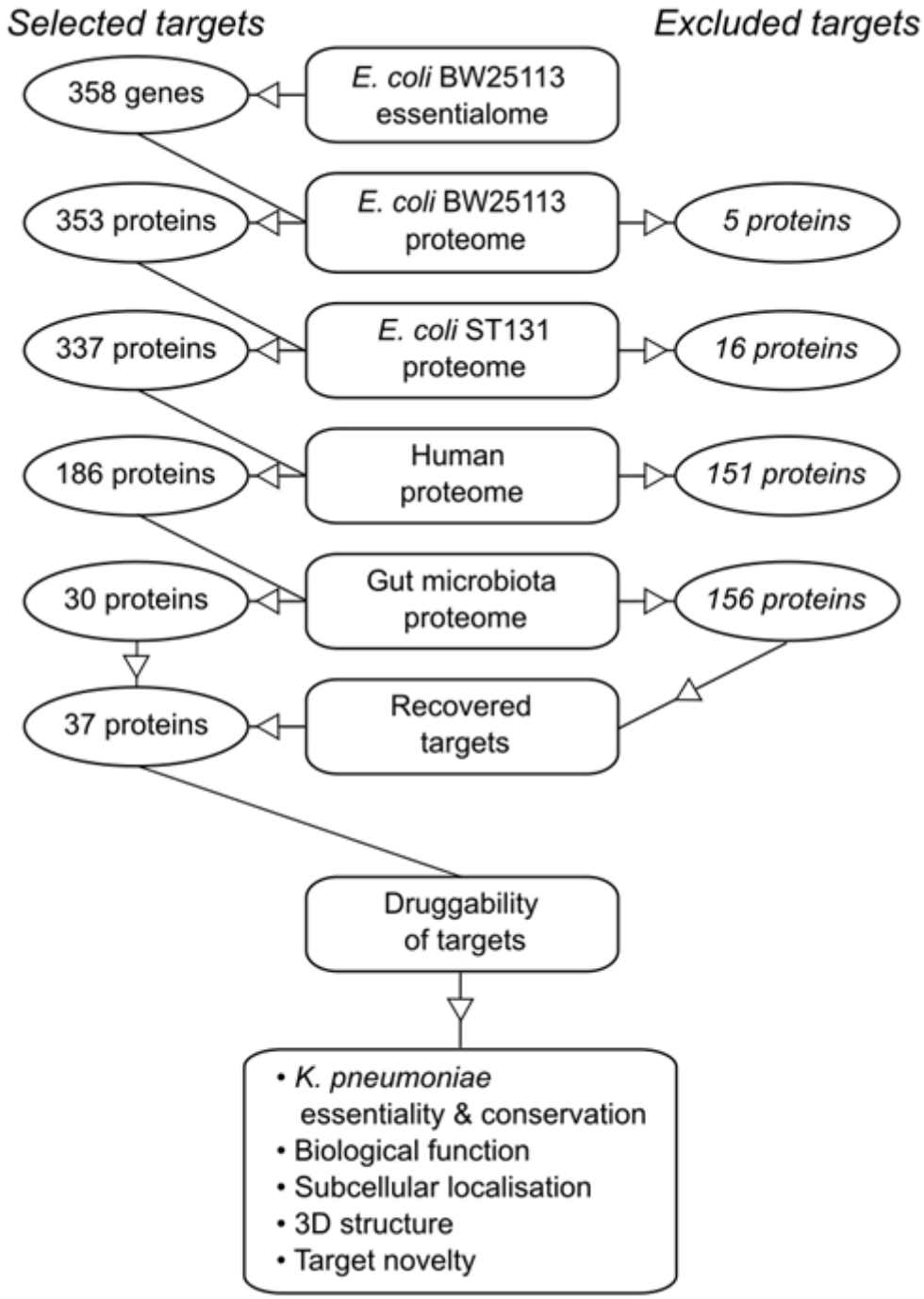

Figure 1

Flow diagram showing the different steps of the selection procedure used to identify novel selective antimicrobial drug targets against E. coli.

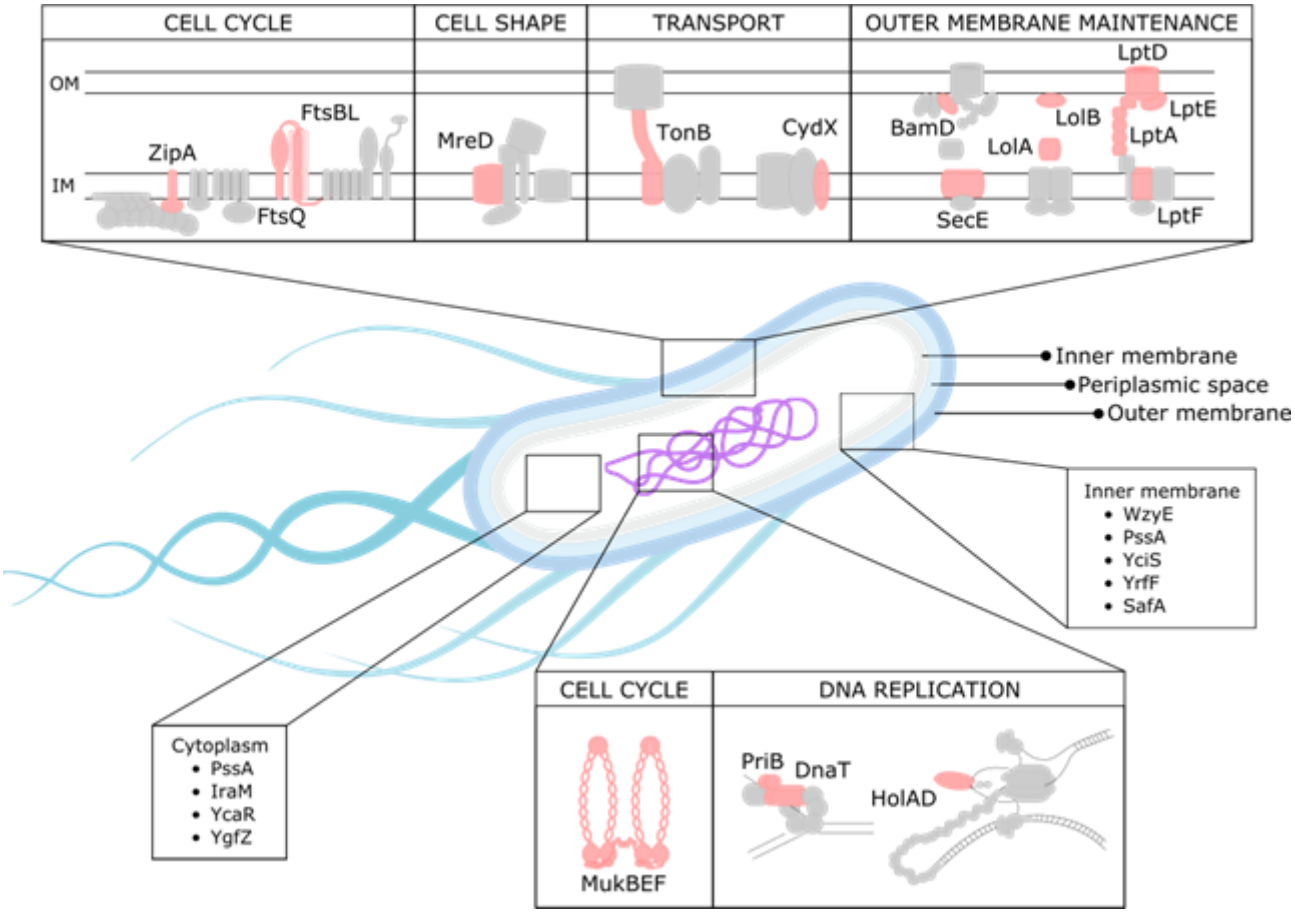

Figure 2

The identified protein targets' functions and subcellular localisation (SCL). The identified targets are drawn in pink, with surrounding proteins in grey. The proteins drawn are part of various cellular processes, including cell cycle (ZipA, FtsBL, FtsQ, MukBEF), cell shape (MreD), transport (TonB, CydX), outer membrane processes (BamD, SecE, LolA, LptA, LptD, LptE and LptF) and DNA replication (PriB, DnaT, HolA, HolD). Other IM proteins are WzyE, PssA, YciS, YrfF, and Cytoplasmic proteins are PssA (note – listed as both IM and Cytoplasmic), IraM, YcaR, YgfZ, Rnt, and FabA. Proteins without a determined SCL are not shown: HemD, TrpL, HigA, HipB, YobI, PheM, and YdhL.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryMaterials1ST131.xlsx
- SupplementaryMaterials2Human.xlsx
- SupplementaryMaterials3Faecalibacterium.xlsx
- SupplementaryMaterials4Prevotella.xlsx
- SupplementaryMaterials5Ruminococcus.xlsx
- SupplementaryMaterials6Bacteroides.xlsx

- SupplementaryMaterials7Lachnospiraceae.xlsx
- SupplementaryMaterials8Lactobacillus.xlsx
- SupplementaryMaterials9Bifidobacterium.xlsx
- SupplementaryMaterials10KlebsiellapneumoniaeKPNIH1.xlsx
- SupplementaryMaterials.docx