

Sequence Diversity of FliC Protein from *Enterobacteriaceae* Family to Introducing a Promising Epitope-Delivery Platform

Sepideh Fereshteh¹, Farzad Badmasti^{*1,2}

¹Department of Bacteriology, Pasteur Institute of Iran, Tehran, Iran, ²Microbiology Research Center (MRC), Pasteur Institute of Iran, Tehran, Iran

ARTICLE INFO

Research Article

VacRes, 2020

Vol. 7, No.2, 85- 91

Received: April 26, 2021

Accepted: May 29, 2021

Pasteur Institute of Iran

*Corresponding Author:

Farzad Badmasti, Department of Bacteriology, Pasteur Institute of Iran, Tehran, Iran.

Email: fbadmasti2008@gmail.com

Tel/Fax: (+98) 2166405535

KEYWORDS: FliC, *Enterobacteriaceae* family, Phylogenetic tree, TLR-5

ABSTRACT

Introduction: Flagellin (FliC) is an essential and universal subunit of flagella in motile bacteria. Analysis of the sequence diversity of the FliC protein in the *Enterobacteriaceae* family may provide insight into the pathogenic strategy of these bacteria in the context of interaction with pathogen-associated molecular patterns (PAMPs). In addition, the efficacy of FliC as an adjuvant or a component of the multi-epitope vaccine has been demonstrated. **Methods:** We analyzed 392 full-length and non-redundant FliC proteins from the *Enterobacteriaceae* family. In the first step, isoelectric pH, molecular weight (kDa), and amino acid composition of FliC protein sequences were calculated using the ProtParam program. Next, the tertiary structure of the FliC proteins and its interaction with zebrafish TLR-5 was performed. Sequence alignments were performed using ClustalW software and the curricular phylogenetic tree was depicted using MEGA-7 software and iTOL web server. Finally, we evaluated the interaction of FliC proteins of *Enterobacteriaceae* family with TLR-5 to find the strongest docking. **Results:** Physicochemical properties and multiple sequence alignments revealed that FliC has a unique characterization in each genus. However, D1 as the binding domain site associated with TLR-5 exhibited high sequence conservation and the FliC protein of *S. enterica* subsp. *enterica* had the strongest interaction with TLR-5. **Conclusions:** FliC protein of *S. enterica* subsp. *enterica* can be consider as a promising epitope-delivery platform. In addition, phylogenetic analysis revealed that FliC may be acceptable marker for distinguishing genera in the family *Enterobacteriaceae*.

Citation:

Fereshteh S, Badmasti F. Sequence Diversity of FliC Protein from Enterobacteriaceae Family to Introducing a Promising Epitope-Delivery Platform. vacres. 2020; 7 (2) :85-91
URL: <http://vacres.pasteur.ac.ir/article-1-245-en.html>

INTRODUCTION

Flagella are filamentous organelles with essential roles in bacterial physiology, such as motility and chemotaxis [1]. Intensive studies have shown that more than 50 genes are complicated in the assembly and function of flagella [2, 3]. Flagellin (FliC) is an essential and universal subunit of flagella in motile bacteria. The flagellin subunit, encoded by the *fliC* gene, has a tubular structure at the distal end of flagella [4]. The atomic model of flagellin as a component of the 11 proto-filaments was constructed from electron micrographs [5]. The 3D structure of the FliC protein showed that this protein typically has four main domains, consisting of D0, D1, D2 and D3. The D0 and D1 domains are required to mediate flagellar polymerization. They have conserved segments in the N and C terminals of FliC and are considered to be the filament core. The inner domain (D0) consists of ~30 N-terminal and ~30 C-terminal residues forming an α -helical coiled-coil structure. The outer domain (D1) consists of three α -helices, designated

ND1a, ND1b and CD1, two β -turns, a β -hairpin and a long chain extending along the three α -helices [6]. Crystal structure of the N-terminal fragment of zebrafish TLR-5 in complex with *Salmonella* flagellin revealed that the D1 domain interacts with TLR-5 as a pathogen-associated molecular pattern (PAMPs) [7].

The D2 and D3 domains are highly variable segments consisting mainly of β -strands projecting from the central core of this protein. These domains would be exposed to the solution and could play an essential role in immunogenicity. Studies have shown that motility-mediated flagella contribute as virulence factors to infectious bacteria [8]. Mutational approaches showed that EHEC and EPEC *fliC* deletion mutants significantly impaired adherence and micro-colony formation on bovine intestinal tissues compared with parental wild-type strains [9]. An increasing number of investigations have

demonstrated the efficacy of FliC as an adjuvant for vaccine development [10, 11].

Moreover, its ability to induce pro-inflammatory cytokines has confirmed that flagellin has good potential for multi-epitope vaccine design [12]. Based on this background, this study aimed to investigate flagellin sequence conservation and diversity in light of recently determined FliC structures among *Enterobacteriaceae* family and introducing a potent self-adjuvanting protein carrier. We focused on the family *Enterobacteriaceae*, which are predominantly motile, except for the genera *Shigella* and *Klebsiella* spp.

MATERIALS AND METHODS

Sequence Retrieval

The sequences of FliC proteins were obtained from the UniProt database (<https://www.uniprot.org/>) [13] using the following keywords "bacterial flagellin" or "FliC" in *Enterobacteriaceae* family and was saved in FASTA format for further investigations.

Evaluation of Physicochemical Properties of FliC

ProtParam (<https://web.expasy.org/protparam/>) is a program that calculates physicochemical properties for a protein contained in Swiss-Prot or TrEMBL, as well as for a user-entered protein sequence [14]. In this step, isoelectric pH, molecular weight (kDa), and amino acid composition were obtained from this database.

Multiple Sequence Analysis and Phylogenetic Tree Drawn

All of the alignments were done using ClustalW software [16]. The phylogenetic tree was created with MEGA-7 software [17] using the neighbor-joining method for any species (BLOSUM64 matrix, Bootstrap= 500). The curricular tree was depicted by the iTOL web server (<https://itol.embl.de/>) [18].

Docking of FliC Proteins with TLR-5

To measure the interaction of FliC proteins of *Enterobacteriaceae* family with TLR5 the pyDockWEB (<https://life.bsc.es/pid/pydockweb>) was used. The PyDockWEB returns the best rigid-body docking orientations generated by FTDock and evaluated by pyDock scoring function, which includes electrostatics, desolvation energy and limited van der Waals contribution [19]. It should be mentioned that before this experiment the 3D structure prediction of FliC proteins was conducted using Robetta (<https://rosetta.bakerlab.org/>).

RESULTS

Sequence Retrieval

As a result, a total number of 1023 FliC FASTA sequences were retrieved. First, truncated or unrelated sequences were excluded. Finally, 329 non-redundant full-length sequences were selected. These sequences included ten groups of species, including *S. enterica* subsp. *enterica* serovars, *Escherichia coli*, *Yersinia enterocolitica*, *Proteus mirabilis*, *Citrobacter freundii*, *Edwardsiella tarda*, *Morganella morganii*, *Providencia* spp., *Serratia* spp., and *Enterobacter cloacae*. The distribution of sequences in each group has been presented in Table 1.

Evaluation of Physicochemical Properties of FliC Proteins

The data obtained from the ProtParam program showed that the isoelectric pH of FliC proteins from the *Enterobacteriaceae* family ranged from 3.96 to 8.32 and their molecular weight ranged from 29 to 68 kDa (Table 2). The amino acid composition of FliC protein is shown in Fig. 1. The distribution of amino acid sequences of all FliC proteins from the *Enterobacteriaceae* family showed that Alanine (Ala), Threonine (Thr) and Serine (Ser) are common amino acids while there is no Cysteine (Cys) residue in these proteins. These results indicated that these proteins are free of disulfide bonds.

Table 1. Retrieving of FliC protein in *Enterobacteriaceae* family extracted from UniProt database.

Bacteria	Total entry	Full-length entry	Non-redundant entry
<i>Salmonella enterica</i> subsp. <i>enterica</i> serovars	511	331	236
<i>Escherichia coli</i>	449	376	107
<i>Yersinia enterocolitica</i>	30	22	18
<i>Proteus mirabilis</i>	8	8	6
<i>Citrobacter freundii</i>	2	2	2
<i>Edwardsiella tarda</i>	4	4	4
<i>Morganella morganii</i>	5	5	5
<i>Providencia</i> spp.	3	3	3
<i>Serratia</i> spp.	9	9	9
<i>Enterobacter cloacae</i>	2	2	2
Total	1023	762	392

Table 2. Data on isoelectric pH and molecular weight (kDa) of FliC proteins in the family *Enterobacteriaceae*.

Bacteria	Isoelectric pH	Molecular weight (kDa)
<i>S. enterica</i> subsp. <i>enterica</i> serovars	4.47 - 4.85	34 - 61
<i>E. coli</i>	4.24 - 4.85	36 - 68
<i>Y. enterocolitica</i>	4.70 - 5.11	38 - 39
<i>P. mirabilis</i>	4.78 - 8.32	39
<i>C. freundii</i>	4.75 - 4.83	47
<i>E. tarda</i>	4.76 - 5.05	37 - 44
<i>M. morganii</i>	4.68 - 5.39	38 - 41
<i>Providencia</i> spp.	4.68 - 4.95	39 - 41
<i>Serratia</i> spp.	4.63 - 4.82	37 - 44
<i>E. cloacae</i>	3.96 - 5.50	29

The Interaction of FliC Protein with TLR-5

The interaction of FliC protein of with TLR-5 of *Danio rerio* (Zebrafish) have shown in Fig. 2. It was confirmed that TLR5 is known to specifically recognize flagellin [20].

It seems that Asn277 of TLR-5 has a central role in this interaction. On the other hand, Arg90, Gln97 and Asn440 are key residues in FliC. See Fig. 2. In addition, the multiple sequence alignments and amino acids of the D1 domain involved in this interaction from *S. enterica* subsp. *enterica* serovars have been shown in Fig. 3.

The FliC protein consists of different parts, such as the D1, the D2, C-terminal and the N-terminal domains. According to the obtained results, the amino acids of the D1 domain in *S. enterica* subsp. *enterica* serovars were involved in TLR-5 binding. In Fig. 3, the red amino acids interact electrostatically with TLR-5, while the yellow amino acids participate in Van

der Waals bonding. The positions of Arg90, Gln97 (R and Q amino acids, left hand), and Asn440 (N amino acid, right hand)

are shown in red. Q97 and N440 are conserved, while there is an amino acid substitution in R90 (R90S).

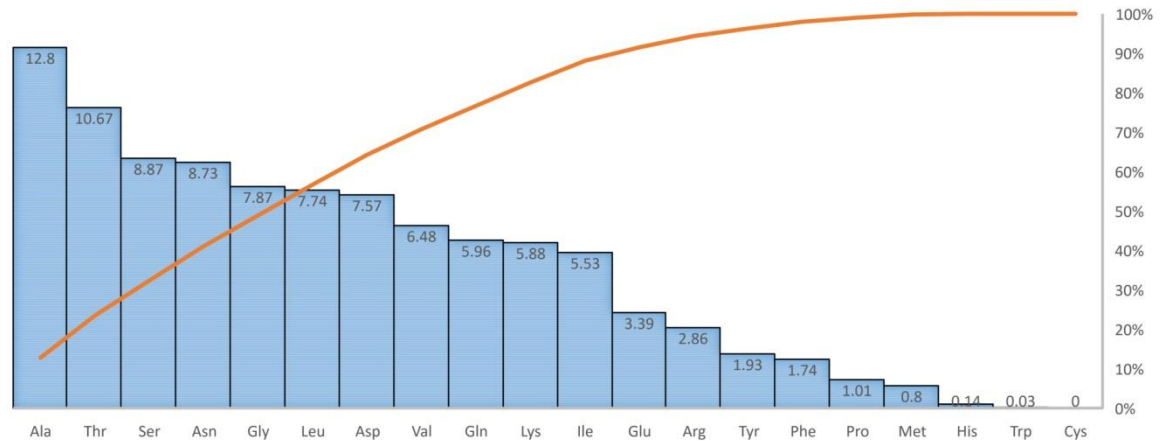


Fig. 1. Amino acid sequence distribution among 392 FliC proteins extracted from the *Enterobacteriaceae* family using the ProtParam program. According to obtained results Ala, Thr, and Ser were the most common amino acids. The orange curve shows the cumulative frequency of amino acid composition in FliC protein.

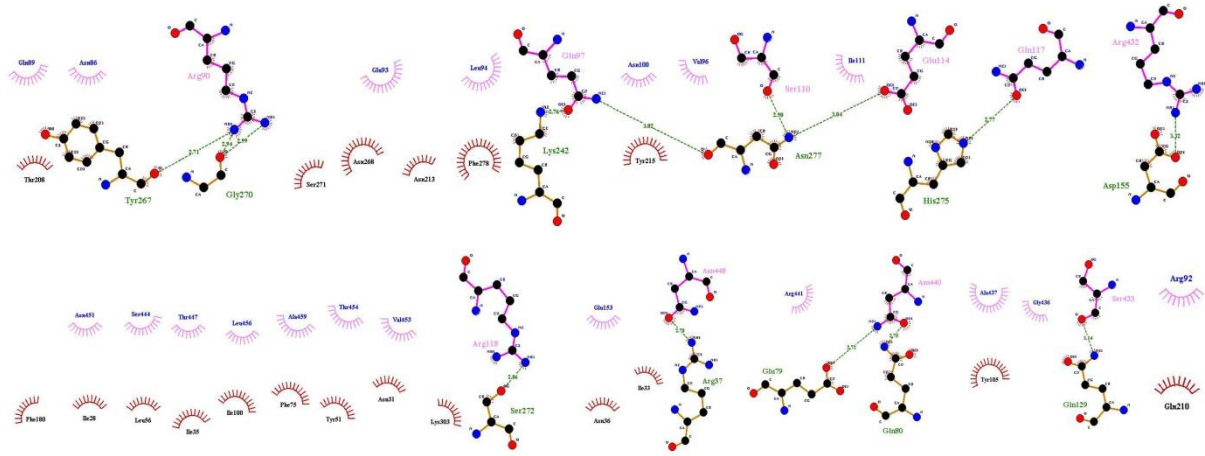


Fig. 2. The LIGPLOT representation of FliC- TLR-5 interactions. The pink amino acids of FliC interact with the green amino acids of TLR-5 in terms of electrostatic interaction. In addition, several amino acids (pink and red fluffy amino acids) are involved in Van der Waals binding. It seems that Asn277 of TLR-5 has a central role in this interaction. On the other hand, Arg90, Gln97 and Asn440 are key residues in FliC.

N-terminal of D1				C-terminal of D1			
3v47.CSRNANDGISIAQTTEGALNEINNLQVRELSVATNGTNSDSDIK	ODIDRLNEIDRVSGOTQFNGVKVLAODNTL	TIQVANDGETTIDILQ	NPLASIDSALS KVDVRS LGLA IQRFDSATTLNGVTNLSARSRIE			
11o1.ANIKGLTQASRNDGISIAQTTEGALNEINNLQVRELSVATNGTNSDSDIK	QAEITRLNEIDRVSGOTQFNGVKVLAODNTL	TIQVANDGETTIDILK	NPLKIDDAALAOVDLRLSLGAVORRFNSATTLNGVTNLSARSRIE			
1UCU.A	DDAAGQAIANRFANIKGLTQASRNDGISIAQTTEGALNEINNLQVRELSVATNGTNSDSDIK	QAEITRLNEIDRVSGOTQFNGVKVLAODNTL	TIQVANDGETTIDILK	NPLKIDDAALAOVDLRLSLGAVORRFNSATTLNGVTNLSARSRIE			
34S.KA	DDAAGQAIANRFANIKGLTQASRNDGISIAQTTEGALNEINNLQVRELSVATNGTNSDSDIK	QAEITRLNEIDRVSGOTQFNGVKVLAODNTL	TIQVANDGETTIDILK	NPLKIDDAALAOVDLRLSLGAVORRFNSATTLNGVTNLSARSRIE			
spjP061.78	DDAAGQAIANRFANIKGLTQASRNDGISIAQTTEGALNEINNLQVRELSVATNGTNSDSDIK	QAEITRLNEIDRVSGOTQFNGVKVLAODNTL	TIQVANDGETTIDILK	NPLKIDDAALAOVDLRLSLGAVORRFNSATTLNGVTNLSARSRIE			
trjD3JX55	DDAAGQAIANRFANIKGLTQASRNDGISIAQTTEGALNEINNLQVRELSVATNGTNSDSDIK	QAEITRLNEIDRVSGOTQFNGVKVLAODNTL	TIQVANDGETTIDILK	NPLKIDDAALAOVDLRLSLGAVORRFNSATTLNGVTNLSARSRIE			
trjD3JXT1	DDAAGQAIANRFANIKGLTQASRNDGISIAQTTEGALNEINNLQVRELSVATNGTNSDSDIK	QAEITRLNEIDRVSGOTQFNGVKVLAODNTL	TIQVANDGETTIDILK	NPLKIDDAALAOVDLRLSLGAVORRFNSATTLNGVTNLSARSRIE			
trjD3JKT2	DDAAGQAIANRFANIKGLTQASRNDGISIAQTTEGALNEINNLQVRELSVATNGTNSDSDIK	QAEITRLNEIDRVSGOTQFNGVKVLAODNTL	TIQVANDGETTIDILK	NPLKIDDAALAOVDLRLSLGAVORRFNSATTLNGVTNLSARSRIE			
trjE7C146	DDAAGQAIANRFANIKGLTQASRNDGISIAQTTEGALNEINNLQVRELSVATNGTNSDSDIK	QAEITRLNEIDRVSGOTQFNGVKVLAODNTL	TIQVANDGETTIDILK	NPLKIDDAALAOVDLRLSLGAVORRFNSATTLNGVTNLSARSRIE			
trjE7CZ12	DDAAGQAIANRFANIKGLTQASRNDGISIAQTTEGALNEINNLQVRELSVATNGTNSDSDIK	QAEITRLNEIDRVSGOTQFNGVKVLAODNTL	TIQVANDGETTIDILK	NPLKIDDAALAOVDLRLSLGAVORRFNSATTLNGVTNLSARSRIE			
trjO53991	DDAAGQAIANRFANIKGLTQASRNDGISIAQTTEGALNEINNLQVRELSVATNGTNSDSDIK	QAEITRLNEIDRVSGOTQFNGVKVLAODNTL	TIQVANDGETTIDILK	NPLKIDDAALAOVDLRLSLGAVORRFNSATTLNGVTNLSARSRIE			
trjO53993	DDAAGQAIANRFANIKGLTQASRNDGISIAQTTEGALNEINNLQVRELSVATNGTNSDSDIK	QAEITRLNEIDRVSGOTQFNGVKVLAODNTL	TIQVANDGETTIDILK	NPLKIDDAALAOVDLRLSLGAVORRFNSATTLNGVTNLSARSRIE			
trjO54863	DDAAGQAIANRFANIKGLTQASRNDGISIAQTTEGALNEINNLQVRELSVATNGTNSDSDIK	QAEITRLNEIDRVSGOTQFNGVKVLAODNTL	TIQVANDGETTIDILK	NPLKIDDAALAOVDLRLSLGAVORRFNSATTLNGVTNLSARSRIE			
trjO54864	DDAAGQAIANRFANIKGLTQASRNDGISIAQTTEGALNEINNLQVRELSVATNGTNSDSDIK	QAEITRLNEIDRVSGOTQFNGVKVLAODNTL	TIQVANDGETTIDILK	NPLKIDDAALAOVDLRLSLGAVORRFNSATTLNGVTNLSARSRIE			
trjO56999	DDAAGQAIANRFANIKGLTQASRNDGISIAQTTEGALNEINNLQVRELSVATNGTNSDSDIK	QAEITRLNEIDRVSGOTQFNGVKVLAODNTL	TIQVANDGETTIDILK	NPLKIDDAALAOVDLRLSLGAVORRFNSATTLNGVTNLSARSRIE			
trjO66PR1	DDAAGQAIANRFANIKGLTQASRNDGISIAQTTEGALNEINNLQVRELSVATNGTNSDSDIK	QAEITRLNEIDRVSGOTQFNGVKVLAODNTL	TIQVANDGETTIDILK	NPLKIDDAALAOVDLRLSLGAVORRFNSATTLNGVTNLSARSRIE			
trjO66P55	DDAAGQAIANRFANIKGLTQASRNDGISIAQTTEGALNEINNLQVRELSVATNGTNSDSDIK	QAEITRLNEIDRVSGOTQFNGVKVLAODNTL	TIQVANDGETTIDILK	NPLKIDDAALAOVDLRLSLGAVORRFNSATTLNGVTNLSARSRIE			
trjO6V265	DDAAGQAIANRFANIKGLTQASRNDGISIAQTTEGALNEINNLQVRELSVATNGTNSDSDIK	QAEITRLNEIDRVSGOTQFNGVKVLAODNTL	TIQVANDGETTIDILK	NPLKIDDAALAOVDLRLSLGAVORRFNSATTLNGVTNLSARSRIE			
trjO6V268	DDAAGQAIANRFANIKGLTQASRNDGISIAQTTEGALNEINNLQVRELSVATNGTNSDSDIK	QAEITRLNEIDRVSGOTQFNGVKVLAODNTL	TIQVANDGETTIDILQ	NPLKIDDAALAOVDLRLSLGAVORRFNSATTLNGVTNLSARSRIE			
trjO6V217	DDAAGQAIANRFANIKGLTQASRNDGISIAQTTEGALNEINNLQVRELSVATNGTNSDSDIK	QAEITRLNEIDRVSGOTQFNGVKVLAODNTL	TIQVANDGETTIDILK	NPLKIDDAALAOVDLRLSLGAVORRFNSATTLNGVTNLSARSRIE			
trjO6V214	DDAAGQAIANRFANIKGLTQASRNDGISIAQTTEGALNEINNLQVRELSVATNGTNSDSDIK	QAEITRLNEIDRVSGOTQFNGVKVLAODNTL	TIQVANDGETTIDILK	NPLKIDDAALAOVDLRLSLGAVORRFNSATTLNGVTNLSARSRIE			
trjO6V216	DDAAGQAIANRFANIKGLTQASRNDGISIAQTTEGALNEINNLQVRELSVATNGTNSDSDIK	QAEITRLNEIDRVSGOTQFNGVKVLAODNTL	TIQVANDGETTIDILQ	NPLKIDDAALAOVDLRLSLGAVORRFNSATTLNGVTNLSARSRIE			
trjO6V2K0	DDAAGQAIANRFANIKGLTQASRNDGISIAQTTEGALNEINNLQVRELSVATNGTNSDSDIK	QAEITRLNEIDRVSGOTQFNGVKVLAODNTL	TIQVANDGETTIDILQ	NPLKIDDAALAOVDLRLSLGAVORRFNSATTLNGVTNLSARSRIE			
trjO6V2K2	DDAAGQAIANRFANIKGLTQASRNDGISIAQTTEGALNEINNLQVRELSVATNGTNSDSDIK	QAEITRLNEIDRVSGOTQFNGVKVLAODNTL	TIQVANDGETTIDILQ	NPLKIDDAALAOVDLRLSLGAVORRFNSATTLNGVTNLSARSRIE			
trjO6V2K6	DDAAGQAIANRFANIKGLTQASRNDGISIAQTTEGALNEINNLQVRELSVATNGTNSDSDIK	QAEITRLNEIDRVSGOTQFNGVKVLAODNTL	TIQVANDGETTIDILQ	NPLKIDDAALAOVDLRLSLGAVORRFNSATTLNGVTNLSARSRIE			
trjO6V2M1	DDAAGQAIANRFANIKGLTQASRNDGISIAQTTEGALNEINNLQVRELSVATNGTNSDSDIK	QAEITRLNEIDRVSGOTQFNGVKVLAODNTL	TIQVANDGETTIDILQ	NPLKIDDAALAOVDLRLSLGAVORRFNSATTLNGVTNLSARSRIE			
trjO6V2M3	DDAAGQAIANRFANIKGLTQASRNDGISIAQTTEGALNEINNLQVRELSVATNGTNSDSDIK	QAEITRLNEIDRVSGOTQFNGVKVLAODNTL	TIQVANDGETTIDILK	NPLKIDDAALAOVDLRLSLGAVORRFNSATTLNGVTNLSARSRIE			
trjO6V2M5	DDAAGQAIANRFANIKGLTQASRNDGISIAQTTEGALNEINNLQVRELSVATNGTNSDSDIK	QAEITRLNEIDRVSGOTQFNGVKVLAODNTL	TIQVANDGETTIDILK	NPLKIDDAALAOVDLRLSLGAVORRFNSATTLNGVTNLSARSRIE			
trjO6V2N1	DDAAGQAIANRFANIKGLTQASRNDGISIAQTTEGALNEINNLQVRELSVATNGTNSDSDIK	QAEITRLNEIDRVSGOTQFNGVKVLAODNTL	TIQVANDGETTIDILK	NPLKIDDAALAOVDLRLSLGAVORRFNSATTLNGVTNLSARSRIE			
trjO6V2N9	DDAAGQAIANRFANIKGLTQASRNDGISIAQTTEGALNEINNLQVRELSVATNGTNSDSDIK	QAEITRLNEIDRVSGOTQFNGVKVLAODNTL	TIQVANDGETTIDILK	NPLKIDDAALAOVDLRLSLGAVORRFNSATTLNGVTNLSARSRIE			
trjO6V2P2	DDAAGQAIANRFANIKGLTQASRNDGISIAQTTEGALNEINNLQVRELSVATNGTNSDSDIK	QAEITRLNEIDRVSGOTQFNGVKVLAODNTL	TIQVANDGETTIDILK	NPLKIDDAALAOVDLRLSLGAVORRFNSATTLNGVTNLSARSRIE			
trjO6V2V7	DDAAGQAIANRFANIKGLTQASRNDGISIAQTTEGALNEINNLQVRELSVATNGTNSDSDIK	QAEITRLNEIDRVSGOTQFNGVKVLAODNTL	TIQVANDGETTIDILK	NPLKIDDAALAOVDLRLSLGAVORRFNSATTLNGVTNLSARSRIE			
trjO6V2Z2	DDAAGQAIANRFANIKGLTQASRNDGISIAQTTEGALNEINNLQVRELSVATNGTNSDSDIK	QAEITRLNEIDRVSGOTQFNGVKVLAODNTL	TIQVANDGETTIDILK	NPLKIDDAALAOVDLRLSLGAVORRFNSATTLNGVTNLSARSRIE			
trjO9KV15	DDAAGQAIANRFANIKGLTQASRNDGISIAQTTEGALNEINNLQVRELSVATNGTNSDSDIK	QAEITRLNEIDRVSGOTQFNGVKVLAODNTL	TIQVANDGETTIDILK	NPLKIDDAALAOVDLRLSLGAVORRFNSATTLNGVTNLSARSRIE			
trjR7RH9K	DDAAGQAIANRFANIKGLTQASRNDGISIAQTTEGALNEINNLQVRELSVATNGTNSDSDIK	QAEITRLNEIDRVSGOTQFNGVKVLAODNTL	TIQVANDGETTIDILK	NPLKIDDAALAOVDLRLSLGAVORRFNSATTLNGVTNLSARSRIE			
trjR7RN44	DDAAGQAIANRFANIKGLTQASRNDGISIAQTTEGALNEINNLQVRELSVATNGTNSDSDIK	QAEITRLNEIDRVSGOTQFNGVKVLAODNTL	TIQVANDGETTIDILK	NPLKIDDAALAOVDLRLSLGAVORRFNSATTLNGVTNLSARSRIE			

Fig. 3. Multiple sequence alignment of D1 in *S. enterica* subsp. *enterica* serovars involved in TLR-5 binding. The red amino acids interact with TLR-5 electrostatically, while the yellow amino acids participate in Van der Waals binding. The green vertical line separates the N-terminal and C-terminal ends of the D1 domain. The positions of Arg90, Gln97 (R- and Q-amino acids, left), and Asn440 (N-amino acid, right) are shown in red. Q97 and N440 are conserved, while there is an amino acid substitution in R90 (R90S).

[Downloaded from vacres.pasteur.ac.ir on 2025-05-17] [DOI: 10.52547/vacres.7.2.85]

Multiple Sequence Analysis and the Phylogenic Tree of FliC

Multiple sequence alignment revealed that there are two conserved (D0 and D1 domains) and hypervariable regions (middle part of FliC proteins). The FliC of *S. enterica* subsp. *enterica* serovars represented high heterogeneity of sequence diversity. The multiple sequence alignment of 392 FliC proteins and NEWICK file were deposited in supplemental data. In this part, the obtained results from the neighbor-joining phylogenetic tree showed that the sequence of FliC can be used

to distinguish the *Enterobacteriaceae* family. Our data indicated that there are three major clades of *S. enterica* subsp. *enterica* serovars compared to one clade of *E. coli*. The distance matrix showed that the FliC proteins of *E. coli* are more homogeneous compared to *S. enterica*. Neighbour-joining phylogenetic tree showed that several FliC types from the genus *Escherichia* have merged into *S. enterica*. See Fig. 4. In addition, the FliC of *Y. enterocolitica* has low diversity of sequence variation compared to other *Enterobacteriaceae* families. See Fig. 4.

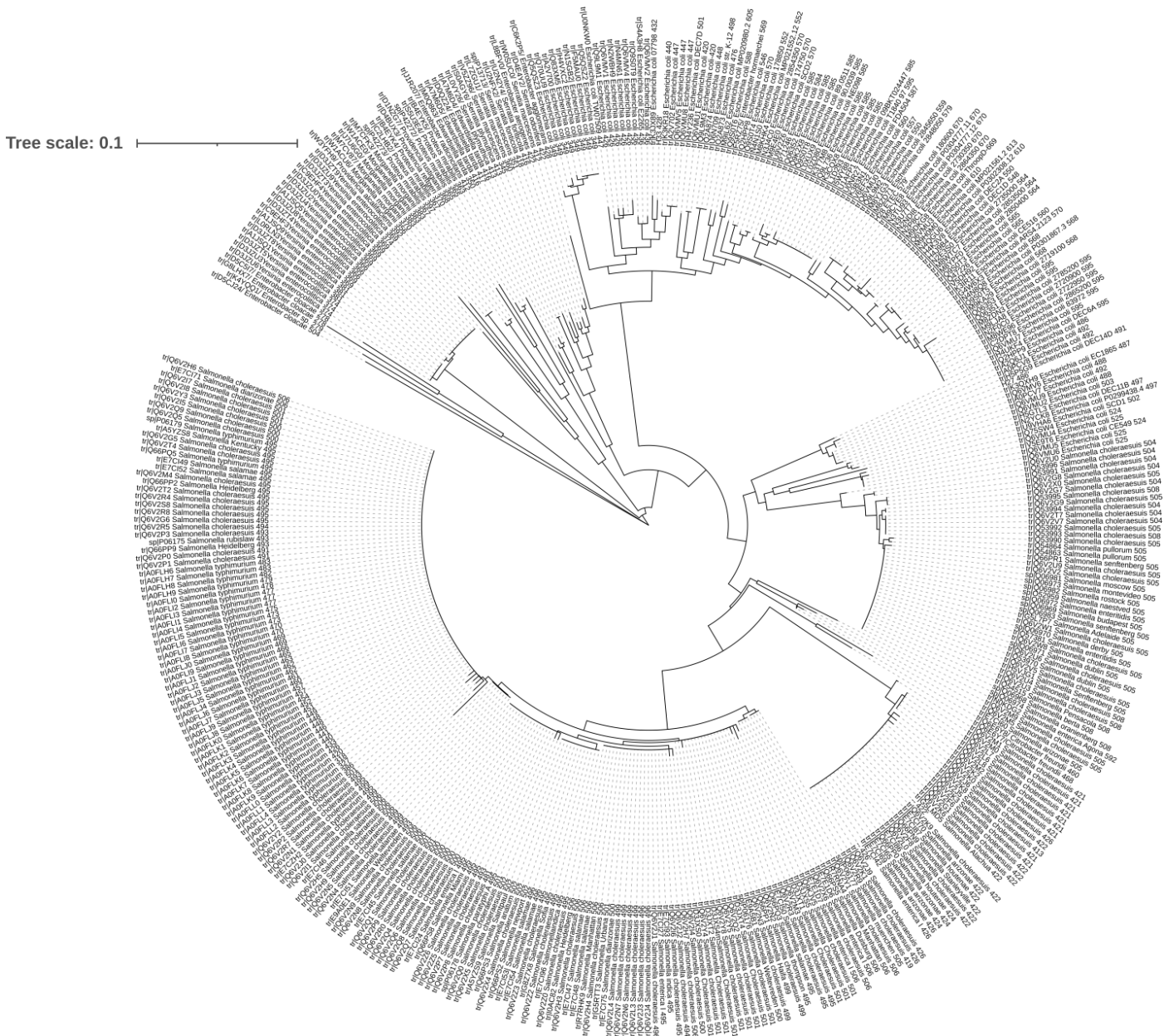


Fig. 4. Neighbour-joining phylogenetic tree of 392 non-redundant full-length FliC proteins extracted from *Enterobacteriaceae* family. This tree shows there are three major clades of *S. enterica* subsp. *enterica* serovars compared to one clade of *E. coli*. The distance matrix showed that the FliC proteins of *E. coli* are more homogeneous compared to *S. enterica*. It seems that several FliC types from the genus *Escherichia* have merged into *S. enterica*.

Interaction of FliC Protein of Enterobacteriaceae with TLR-5

The total interaction of FliC proteins with TLR-5 are as follows: *M. morganii* (-23.555), *Y. enterocolitica* (-27.772), *E. coli* (-22.639), *S. enterica* subsp. *enterica* (-29.774),

P. mirabilis (-22.968), *E. tarda* (-24.099), and *E. cloacae* (-23.703). According to obtained results, it seems that the FliC protein of *S. enterica* subsp. *enterica* had the most stable interaction with TLR-5 and could be considered as the putative vaccine candidate. The full results of this part was shown in Table 3.

Table 3. The interaction of FliC proteins of Enterobacteriaceae family with TLR-5

Bacterium	Electrostatic interaction	Desolvation	Van der Waals forces	Total
<i>M. morganii</i>	-21.420	-3.779	16.444	-23.555
<i>Y. enterocolitica</i>	-20.762	-11.731	47.208	-27.772
<i>E. coli</i>	-25.132	5.837	-33.441	-22.639
<i>S. enterica</i> subsp. <i>enterica</i>	-16.207	-16.194	26.275	-29.774
<i>P. mirabilis</i>	-19.084	-2.801	-10.835	-22.968
<i>E. tarda</i>	-25.302	-1.661	28.641	-24.099
<i>E. cloacae</i>	-7.243	-20.158	36.976	-23.703

DISCUSSION

The term *Enterobacteriaceae* was first proposed in 1936 [21]. This bacterial family is one of the most common causes of severe nosocomial and community-acquired bacterial infections in humans. Antibiotic resistance of these bacteria has become a growing problem in healthcare systems [22].

The flagellum is a whip-like appendage that allows bacteria to move. Flagellin is a component protein of the flagellum. Previously, flagellin was thought to be only a virulence factor that aids in host cell adhesion and invasion. However, it was discovered that it is a potent immune activator that affects both the innate and adaptive arms of immunity during microbial infections [23, 24]. Several genes are involved in flagella expression; however, FliC encodes the flagellin subunits that make up the bulk of a flagellum's structure [25]. In addition, FliC has specific roles in bacterial virulence. For example, in Enteropathogenic *E. coli* (EPEC), FliC induces the release of IL-8 in the T84 cell lineage [26]. In addition, Yang He and colleagues confirmed that FliC plays diverse roles in flagellar function, bacterial growth, protein secretion by TTSS, and virulence in *E. tarda* [27]. On the other hand, Horstmann et al. showed that post-translational methylation of flagellin facilitates adhesion to host cell surfaces and contributes to efficient intestinal colonization and host infection [28].

The efficacy of FliC as an adjuvant has been demonstrated in a growing body of research. An *in silico* study of FliCs from *S. typhimorium*, *P. aeruginosa* and *E. coli* and their use as adjuvants showed that FliC from *S. typhimorium* has a higher quality of physicochemical properties and a higher affinity for TLR-5 than the other FliCs, making its use more desirable as an adjuvant [29]. A study conducted with *S. typhimurium* and *E. coli* flagellin protein showed that purified recombinant FliC proteins can be used for DNA vaccine studies, therapeutic and multi-epitope platform purposes [30]. TLR-5 has three structural domains: extracellular leucine-rich repeats, a transmembrane domain, and a TIR domain. TLR-5 uses a group of relatively conserved amino acid residues located on both the N- and C-terminal domains of flagellin to recognize it [24]. Our data revealed some key residues of FliC protein associated with TLR-5 interaction had high conservation. On the other hand, molecular docking of FliC with TLR-5 showed that *S. enterica* subsp. *enterica* and *E. coli* had the strongest and weakest interactions with TLR-5, respectively. However, the scores of

these interactions in *Enterobacteriaceae* have narrow range (from -22.639 to -29.774).

Comparison of the flagellin amino acid sequence of many bacterial species revealed the unique domain structure of the protein [31]. The N- and C-terminal parts of the molecule responsible for secretion and polymerization are conserved between species, while the central region where the antigenic part exposed on the surface of the flagellar filaments is produced varies greatly within and between species [32]. A study by Reid *et al.*, on the sequencing of the fliC gene in pathogenic strains of *E. coli* with different H antigens, showed that the N- and C-terminal regions of the alleles are largely conserved, while the central region is more ethical and very diverse [33]. Sequencing of fliC genes related to flagellar antigens H25 and H28 revealed the genetic heterogeneity of fliC H25 and fliC H28 gene sequences in *E. coli*. Based on the allelic discrimination of these fliC genes, it is easy to develop a real-time PCR test to identify EHEC O145: H25 and O145: H28 [34]. In this study, the distance matrix showed a narrow difference between FliC proteins in *E. coli*. Potentially, PCR-sequencing or real-time PCR could be applied to detect pathotypes of this bacterium based on FliC sequence variation. However, further analysis needs to be performed.

The flagellar genes including *fliC* and *fliB*, which encode the Salmonella phase 1 and phase 2 antigens, respectively, are considered the Salmonella serotype-determining genes and contribute to the pathogenesis of Salmonella. Sequence diversity of flagellar genes suggested that there were two distinct evolutionary patterns of Salmonella flagellin gene clusters [35]. Mortimer et al. proposed a DNA-sequence based approach to serotyping of *Salmonella enterica*. It seems that sequencing and characterization of fliC can perform a molecular serotyping method [36]. The FliC protein shows great diversity in the middle of this protein among *S. enterica* subsp. *enterica* serovars, as revealed by our multiple sequencing analysis. Indeed, this bacterium exploits this property for its pathogenicity and outwits the immune system [35]. Moreover, *S. enterica* serovars were shown to reduce the expression of FliC below the level required to trigger TLR-5 to minimize identification by innate host defenses [37]. Although resistance to innate immunity is critical for the early survival of

Salmonella, bacterial mechanisms that prevent recognition by adaptive immunity can significantly promote survival [38].

CONCLUSION

In conclusion, FliC protein of *S. enterica* subsp. *enterica* could be considered as a promising epitope-delivery platform. It had the strongest interaction with TLR-5. It is clear that the higher affinity for TLR-5, making FliC use more desirable as an adjuvant. Moreover, the results showed that FliC is an acceptable marker for distinguishing genera in the *Enterobacteriaceae* family. However, these phylogenetic relationships inferred based on FliC sequences should be treated with caution.

ACKNOWLEDGMENT

The authors would like to thank the staff of the Department of Bacteriology of the Pasteur Institute of Iran for their help. No funding was allocated for this study.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest

REFERENCES

- Moens S, Vanderleyden J. Functions of bacterial flagella. *Critical reviews in microbiology*. 1996;22(2):67-100. doi: <https://doi.org/10.3109/10408419609106456>.
- Macnab RM. How bacteria assemble flagella. *Annual Reviews in Microbiology*. 2003;57(1):77-100. doi: <https://doi.org/10.1146/annurev.micro.57.030502.090832>.
- Das C, Mokashi C, Mande SS, Saini S. Dynamics and control of flagella assembly in *Salmonella typhimurium*. *Frontiers in cellular and infection microbiology*. 2018;8:36. doi: <https://doi.org/10.3389/fcimb.2018.00036>.
- Berg HC. The rotary motor of bacterial flagella. *Annual review of biochemistry*. 2003;72. doi: <https://doi.org/10.1146/annurev.biochem.72.121801.161737>.
- Yonekura K, Maki-Yonekura S, Namba K. Complete atomic model of the bacterial flagellar filament by electron cryomicroscopy. *Nature*. 2003;424(6949):643-50. doi: <https://doi.org/10.1038/nature01830>.
- Samatey FA, Imada K, Nagashima S, Vonderviszf F, Kumasaka T, Yamamoto M et al. Structure of the bacterial flagellar protofilament and implications for a switch for supercoiling. *Nature*. 2001;410(6826):331-7. doi: <https://doi.org/10.1038/35066504>.
- Andersen-Nissen E, Smith KD, Strobe KL, Barrett SLR, Cookson BT, Logan SM et al. Evasion of Toll-like receptor 5 by flagellated bacteria. *Proceedings of the National Academy of Sciences*. 2005;102(26):9247-52. doi: <https://doi.org/10.1073/pnas.0502040102>.
- Duan Q, Zhou M, Zhu L, Zhu G. Flagella and bacterial pathogenicity. *Journal of basic microbiology*. 2013;53(1):1-8. doi: <https://doi.org/10.1002/jobm.201100335>.
- Erdem AL, Avelino F, Xicohtencatl-Cortes J, Girón JA. Host protein binding and adhesive properties of H6 and H7 flagella of attaching and effacing *Escherichia coli*. *Journal of bacteriology*. 2007;189(20):7426-35. doi: <https://doi.org/10.1128/JB.00464-07>.
- Lockner JW, Eubanks LM, Choi JL, Lively JM, Schlosburg JE, Collins KC et al. Flagellin as carrier and adjuvant in cocaine vaccine development. *Molecular pharmaceutics*. 2015;12(2):653-62. doi: [10.1021/mp500520r](https://doi.org/10.1021/mp500520r).
- Tran B-CT, Vo-Nguyen H-V, Nguyen V-A, Tran TL, Tran-Van H. FliC-delta220-320 from *Salmonella enteritidis* as an Adjuvant for Vaccine Development. 2015. doi: <http://dx.doi.org/10.21276/SSR-IIJLS.2020.6.2.3>.
- Yazdani Z, Rafiei A, Yazdani M, Valadan R. Design an efficient multi-epitope peptide vaccine candidate against SARS-CoV-2: An in silico analysis. *Infection and drug resistance*. 2020;13:3007. doi: [10.2147/IDR.S264573](https://doi.org/10.2147/IDR.S264573)
- UniProt C. The Universal Protein Resource (UniProt) 2009. *Nucleic acids research*. 2009;37(Database issue):169-74. doi: <https://doi.org/10.1093/nar/gkn664>.

- Fereshteh S, Abdoli S, Shahcheraghi F, Ajdary S, Nazari M, Badmasti F. New putative vaccine candidates against *Acinetobacter baumannii* using the reverse vaccinology method. *Microbial pathogenesis*. 2020;143:104114. doi: <https://doi.org/10.1016/j.micpath.2020.104114>.
- Laskowski RA, Swindells MB. LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *ACS Publications*; 2011. doi: <https://doi.org/10.1021/ci200227u>.
- Thompson JD, Gibson TJ, Higgins DG. Multiple sequence alignment using ClustalW and ClustalX. *Current protocols in bioinformatics*. 2003(1):2.3. 1-2.3. 22. doi: <https://doi.org/10.1002/0471250953.bi0203s00>.
- Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular biology and evolution*. 2007;24(8):1596-9. doi: <https://doi.org/10.1093/molbev/msm092>.
- Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic acids research*. 2019;47(W1):W256-W9. doi: <https://doi.org/10.1093/nar/gkz239>.
- Jiménez-García B, Pons C, Fernández-Recio J. pyDockWEB: a web server for rigid-body protein-protein docking using electrostatics and desolvation scoring. *Bioinformatics*. 2013;29(13):1698-9. doi: <https://doi.org/10.1093/bioinformatics/btt262>.
- Yang J, Yan H. TLR5: beyond the recognition of flagellin. *Cellular & molecular immunology*. 2017;14(12):1017-9. doi: <https://doi.org/10.1038/cmi.2017.122>.
- Oliveira RV, Oliveira MC, Pelli A. Disease infection by *Enterobacteriaceae* family in fishes: a review. *J Microbiol Exp*. 2017;4(5):00128. doi: [10.15406/jmen.2017.04.00128](https://doi.org/10.15406/jmen.2017.04.00128).
- Pitout JD. Multiresistant *Enterobacteriaceae*: new threat of an old problem. *Expert review of anti-infective therapy*. 2008;6(5):657-69. doi: <https://doi.org/10.1586/14787210.6.5.657>.
- Beutin L, Strauch E, Zimmermann S, Kaulfuss S, Schaudinn C, Männel A et al. Genetical and functional investigation of fliC genes encoding flagellar serotype H4 in wildtype strains of *Escherichia coli* and in a laboratory *E. coli* K-12 strain expressing flagellar antigen type H48. *BMC microbiology*. 2005;5(1):1-11. doi: <https://doi.org/10.1186/1471-2180-5-4>.
- Hajam IA, Dar PA, Shahnaz I, Jaume JC, Lee JH. Bacterial flagellin—a potent immunomodulatory agent. *Experimental & molecular medicine*. 2017;49(9):e373-e. doi: <https://doi.org/10.1038/emm.2017.172>.
- Schwan WR, Flohr NL, Multerer AR, Starkey JC. GadE regulates fliC gene transcription and motility in *Escherichia coli*. *World journal of clinical infectious diseases*. 2020;10(1):14. doi: [10.5495/wjcid.v10.i1.14](https://doi.org/10.5495/wjcid.v10.i1.14).
- Zhou X, Girón JA, Torres AG, Crawford JA, Negrete E, Vogel SN et al. Flagellin of enteropathogenic *Escherichia coli* stimulates interleukin-8 production in T84 cells. *Infection and immunity*. 2003;71(4):2120-9. doi: <https://doi.org/10.1128/IAI.71.4.2120-2129.2003>.
- He Y, Xu T, Fossheim LE, Zhang X-H. FliC, a flagellin protein, is essential for the growth and virulence of fish pathogen *Edwardsiella tarda*. *PLoS One*. 2012;7(9):e45070. doi: <https://doi.org/10.1371/journal.pone.0045070>.
- Horstmann JA, Lunelli M, Cazzola H, Heidemann J, Kühne C, Steffen P et al. Methylation of *Salmonella Typhimurium* flagella promotes bacterial adhesion and host cell invasion. *Nature communications*. 2020;11(1):1-11. doi: <https://doi.org/10.1038/s41467-020-15738-3>.
- Bameri Z, Asadi Karam MR, Habibi M, Oloomi M, Ehsani P, Bouzari S. In silico Studies for Comparison of FliCs from *Salmonella typhimurium*, *Pseudomonas aeruginosa*, and *Escherichia coli* and their Implication as an Adjuvant. *International Biological and Biomedical Journal*. 2018;4(2):82-8. doi: <http://ibbj.org/article-1-158-fa.html>.
- Makvandi M, Teimoori A, Nahad MP, Khodadadi A, Zandi M. Expression of *Salmonella typhimurium* and *Escherichia coli* flagellin protein and its functional characterization as an adjuvant. *Microbial pathogenesis*. 2018;118:87-90. doi: <https://doi.org/10.1016/j.micpath.2018.03.016>.
- Macnab R. *Escherichia coli* and *Salmonella typhimurium*: Cellular and Molecular Biology (Neidhardt, FC, Ed.). 1996. doi: <https://doi.org/10.1086/416059>.
- Winstanley C, Morgan JAW. The bacterial flagellin gene as a biomarker for detection, population genetics and epidemiological analysis. *Microbiology*. 1997;143(10):3071-84. doi: <https://doi.org/10.1099/00221287-143-10-3071>.
- Reid SD, Selander RK, Whittam TS. Sequence diversity of flagellin (fliC) alleles in pathogenic *Escherichia coli*. *Journal of bacteriology*. 1999;181(1):153-60. doi: <https://doi.org/10.1128/JB.181.1.153-160.1999>.
- Beutin L, Delannoy S, Fach P. Sequence Variations in the Flagellar Antigen Genes fliC H25 and fliC H28 of *Escherichia coli* and Their Use in Identification and Characterization of Enterohemorrhagic *E. coli* (EHEC)

O145: H25 and O145: H28. PLoS One. 2015;10(5):e0126749. doi: <https://doi.org/10.1371/journal.pone.0126749>.
35. Liu Y, Zhang D-F, Zhou X, Xu L, Zhang L, Shi X. Comprehensive analysis reveals two distinct evolution patterns of Salmonella flagellin gene clusters. *Frontiers in microbiology*. 2017;8:2604. doi: <https://doi.org/10.3389/fmicb.2017.02604>.
36. Mortimer CK, Peters TM, Gharbia SE, Logan JM, Arnold C. Towards the development of a DNA-sequence based approach to serotyping of Salmonella enterica. *BMC microbiology*. 2004;4(1):1-10. doi: <https://doi.org/10.1186/1471-2180-4-31>.

37. Pan Z, Cong Q, Geng S, Fang Q, Kang X, You M et al. Flagellin from recombinant attenuated Salmonella enterica serovar Typhimurium reveals a fundamental role in chicken innate immunity. *Clinical and Vaccine Immunology*. 2012;19(3):304-12. doi: <https://doi.org/10.1128/CVI.05569-11>.
38. Olsen JE, Hoegh-Andersen KH, Casadesús J, Rosenkrantz JT, Chadfield MS, Thomsen LE. The role of flagella and chemotaxis genes in host pathogen interaction of the host adapted Salmonella enterica serovar Dublin compared to the broad host range serovar S. Typhimurium. *BMC microbiology*. 2013;13(1):1-11. doi: <https://doi.org/10.1186/1471-2180-13-67>.