 doi:10.20944/preprints201905.0105.v1

# Sequence-based analysis of *zonula occludens* toxins identified by comparative genomics in non-toxigenic *Vibrio parahaemolyticus* strains isolated in Southern Chile.

Diliana Pérez-Reytor[1], Daniel Castillo[2], Carlos J. Blondel[1], Sebastián Ramírez-Araya[1,3], Nicolás Plaza[1], Alequis Pavón[1], Gino Corsini[1], Víctor Jaña[4], Leonardo Pavéz[4,5], Roberto Bastías[6], Paola Navarrete[7,8], Katherine García[1*]

1. *Instituto de Ciencias Biomédicas, Facultad de Ciencias de la Salud, Universidad Autónoma de Chile. DPR: d.perez@uautonoma.cl; CJB: carlos.blondel@uautonoma.cl; NP: nicolas.plaza@uautonoma.cl; AP: alequis.pavon@uautonoma.cl; GC: gino.corsini@uautonoma.cl; KG: katherine.garcia@uautonoma.cl.*
2. *Marine Biological Section, University of Copenhagen, Strandpromenaden 5, DK-3000, Helsingør, Denmark. DC: daniel.castillo@bio.ku.dk.*
3. *Departamento Ciencias Básicas, Facultad de Ciencias, Universidad Santo Tomás, Av. Ejército Libertador 146, Santiago de Chile. SR: seba264@gmail.com.*
4. *Facultad de Medicina Veterinaria y Agronomía, Universidad de Las Américas, Santiago, Chile. VJ: victor.jgaray@gmail.com; LP: lpavez@udla.cl.*
5. *Departamento de Ciencias Químicas y Biológicas, Universidad Bernardo O'Higgins, Santiago, Chile. LP: leonardopavez@docente.ubo.cl.*
6. *Laboratory of Microbiology, Institute of Biology, Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile. RB: roberto.bastias@pucv.cl.*
7. *Laboratory of Microbiology and Probiotics, Institute of Nutrition and Food Technology (INTA), University of Chile, Santiago, Chile. PN: pnavarrete@inta.uchile.cl.*
8. *Millenium nucleus in the Biology of Intestinal Microbiota, Santiago, Chile.*

* Correspondence: katherine.garcia@uautonoma.cl, *Instituto de Ciencias Biomédicas, Facultad de Ciencias de la Salud, Universidad Autónoma de Chile. El Llano Subercaseaux 2801, San Miguel, Santiago de Chile.*

## Abstract

Gastroenteritis cases associated with non-toxigenic strains of *Vibrio parahaemolyticus* have been reported in many countries, suggesting the contribution of novel virulence factors. One candidate is *zonula occludens* toxin (Zot), which increases the intestinal permeability by other bacteria. Recently we identified prophages belonging to the *Inoviridae* family encoding putative Zot-like toxins in Chilean strains. Based on this information we performed sequence-based analyses of these toxins, followed by phylogenetic and structural analyses using computational tools. Our results showed that Zots found in Chilean *V. parahaemolyticus* strains are grouped into three different phylogenetic clusters, sharing two conserved motifs (Walker A and B) in their N-terminal region. These motifs are also conserved in Zots from the human pathogens *Vibrio cholerae*, *Neisseria meningitidis* and *Campylobacter concisus*. Although Zots of *V. parahaemolyticus* do not possess the FCIGRL sequence responsible for the effects produced by *V. cholerae*, they do possess a conserved secondary structure within their C-terminal region with Zots proteins able to disrupt the intestinal barrier, which is interesting since it has been suggested that the structure and not the Zot sequence would be responsible for the biological effects. This preliminary study provides the

 doi:10.20944/preprints201905.0105.v1

basis to study the function of Zots found in *V. parahaemolyticus* on the intestinal barrier and their possible role as a virulence factor.

**Keywords:** *Vibrio parahaemolyticus*, non-toxigenic strains, Zot, *zonula occludens* toxin, *Vibrio cholerae*, *Campylobacter concisus*, *intestinal permeability*

**Key Contribution:** This study provides valuable information for a more in-depth examination of Zot sequences found in *Vibrio parahaemolyticus* strains, and additionally provides the basis for the study of their biological effects on the intestinal epithelial barrier and their definition as a possible new virulence factor in this species.

## 1. Introduction

Inshore marine waters around the world are densely populated with *Vibrio parahaemolyticus* [1], which is the leading cause of seafood-associated bacterial gastroenteritis [2]. However, only a few strains can cause infections in humans and most environmental strains are non-pathogenic [3]. The most characteristic virulence-associated factors are thermostable direct hemolysin (TDH) and TDH-related hemolysin (TRH), encoded by the *tdh* and *trh* genes, respectively [2–5]. However, *V. parahaemolyticus* remains pathogenic in the absence of these hemolysins, indicating that other virulence factors exist [2]. Analysis of the complete genome sequence of *V. parahaemolyticus* strain RIMD2210633 revealed the presence of other virulence factors such as the type III and VI secretion systems (designated T3SS and T6SS respectively), in both chromosome and in various genomic islands (VPaI) [6–8]. Studies reported that environmental isolates of *V. parahaemolyticus* lacking *tdh* and/or *trh* and T3SS2 can be highly cytotoxic to human gastrointestinal cells [9–11]. These results indicate that cytotoxicity and enterotoxicity of pathogenic *V. parahaemolyticus* are not explained only by classic virulence factors and suggest that one or more novel virulence factors could be responsible for its pathogenicity [12]

It is known that *Vibrio* species share virulence genes in estuarine environments where they live [8]. In fact, *Vibrio cholerae* has an arsenal of different toxins besides the classical cholera toxin (CT), including the *zonula occludens toxin* (Zot), the most important toxin in the absence of CT [13]. The *zot* gene was first described in *V. cholerae*; it is encoded by the CTX prophage [14,15]. The N-terminal domain of the *V. cholerae* Zot protein is involved in bacteriophage morphogenesis, while the C-terminal domain is cleaved and secreted into the intestinal lumen [15–17]. Structure-function analyses indicate that the biologically active fragment of Zot (FCIGRL) can be mapped to amino acids 288-293. FCIGRL is structurally similar to another motif (SLIGRL) that activates an intracellular signaling pathway by binding to proteinase-activated receptor-2 (PAR-2), a receptor that has been implicated in the regulation of paracellular permeability, inducing a transient reduction in transepithelial resistance and an increase in transepithelial flux along concentration gradients by affecting the tight junction (TJ) permeability [18–20]. It was recently reported that toxigenic *Campylobacter concisus* strains producing Zot have the potential to initiate inflammatory bowel disease or could be aggravators of Crohn disease [21,22]. This Zot protein

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 May 2019 doi:10.20944/preprints201905.0105.v1

causes sustained intestinal barrier damage, induces the liberation of proinflammatory cytokines and increases the response of macrophages to other microorganisms [23]. In our previous work we showed that a *V. parahaemolyticus zot* sequence is encoded into prophage f237 of the pandemic RIMD2210633 strain, which is different from the *V. cholerae* CTX prophage [11]. We also showed that the clinical strain PMC53.7, which does not possess any other known virulence factor in its genome, also possess a *zot*-encoding prophage [11]. However, whether these Zot proteins encode *bona-fide* enterotoxins remains unknown. It was recently reported that different *zot*-encoding prophages were found in 77.9% of the clinical isolates of *V. parahaemolyticus* [24]. These prophages belong to the *Inoviridae* family, which plays an important role in the evolution and pathogenesis of multiple bacteria, showing that Zot is highly prevalent in clinical strains of this species [24]. This suggests that Zots could have a possible role in the pathogenesis of *V. parahaemolyticus*. In this preliminary study we decided to perform sequence-based analyses of Zot toxins from non-toxigenic Chilean *V. parahaemolyticus* strains, followed by phylogenetic and structural analyses to identify their main features using computational tools. The phylogenetic analysis of Zot-like toxin proteins showed that Zot from Chilean non-toxigenic strains PMC53.7 and PMA2.15 belongs to the B4 phylogenetic group, while Zot of PMA3.15 belongs to B2 phylogenetic group defined by Castillo et al. 2018b [24]. We noted that two motifs in the N-terminal end (Walker A and Walker B) and the secondary structure of the C-terminal motif FCIGRL are highly conserved with Zots of *V. cholerae* and *C. concisus* strains, which are able to damage the intestinal barrier and disturb intestinal permeability. These results will provide useful information to study further a potential role of Zot in the pathogenesis of *V. parahaemolyticus*.

## 2. Results

### Phylogenetic analysis of *zot* sequences and comparison of prophages containing *zot* in *Vibrio* species

The *zonula occludens* toxin gene (*zot*) located in the CTX prophage has been associated with the pathogenicity of *V. cholerae* [25,26] (Figure 1A). Interestingly, a recent report has shown that *zot*-encoding prophages are widely distributed among *Vibrio* species, including the environmental human pathogen *V. parahaemolyticus* [24]. For example, the *zot* gene has been detected in the *V. parahaemolyticus* prophage f237 and in the filamentous phages VfO3:K6 and VfO4:K68. However, no homologs to *ctx* toxin genes were present (Figure 1A).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 May 2019 doi:10.20944/preprints201905.0105.v1

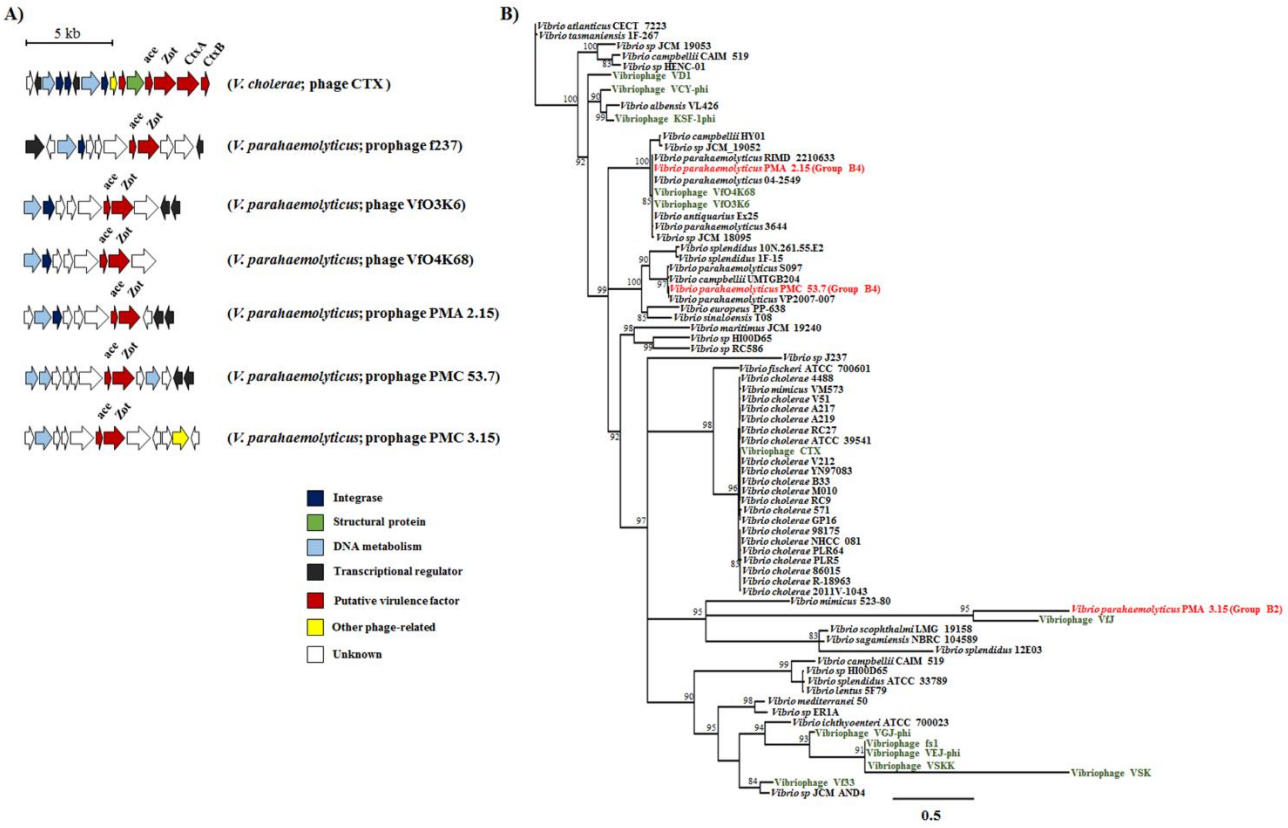



**Figure 1. Genomic organization of *zot*-encoding prophages in *Vibrio* species A)** Diagrammatic representation of *zot*-encoding prophages and phages of different pathogenic and environmental *Vibrio parahaemolyticus* isolates (See Supplementary Table 1S). **B)** Maximum likelihood tree based on the amino acid sequences of Zots found in different *Vibrio* species. *V. parahaemolyticus* Zots of Chilean non-toxigenic strains belonging to the B2 and B4 phylogenetic groups are highlighted. Bootstrap values <80% were removed from the tree. The horizontal bar at the base of the figure represents 0.5 substitutions per amino acid site.

A list of Zot-encoding prophages in *V. parahaemolyticus* and *V. cholerae* is shown in Table 1.

**Table 1. Prophages which harbor *zot* in *V. parahaemolyticus* and *V. cholerae* prophage strains**

| Phage | Host | Origin | Year | Size (bp) | Accession number |
|---|---|---|---|---|---|
| CTX | *Vibrio cholerae* strain KMN002 | Unknown | 1996 | 10,638 | HQ224500.1 |
| VFJ | *Vibrio cholerae* strain ICDC-4470 | Unknown | 2012 | 8,555 | KC357596.1 |
| VCY-phi | *Vibrio cholerae* strain 10E09PW02 | USA, MA | 2011 | 7,103 | JN848801.1 |

*Preprints* (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 May 2019 doi:10.20944/preprints201905.0105.v1

| | | | | | |
|---|---|---|---|---|---|
| VEJphi | *Vibrio cholerae* strain MO45 | India | 2010 | 6,849 | FJ904927 |
| KSF1 | *Vibrio cholerae* strain 55V71 | Bangladesh | 2005 | 7,107 | AY714348.1 |
| VSK | *Vibrio cholerae* O139 | India | 2001 | 6,882 | AF453500.3 |
| VGJphi | *Vibrio cholerae* strain SG25-1 | India | 2003 | 7,542 | AY242528.1 |
| fs1 | *Vibrio cholerae* strain A1-4450 | Unknown | 1997 | 6,341 | NC_004306.1 |
| fs2 | *Vibrio cholerae* strain MDO14 | Unknown | 1998 | 8,651 | AB002632.1 |
| Vf12 | *Vibrio parahaemolyticus* strain Vp12 | Unknown | 1998 | 7,965 | AB012574.1 |
| Vf33 | *Vibrio parahaemolyticus* strain Vp33 | Unknown | 1998 | 7,965 | AB012573.1 |
| VfO4K68 | *Vibrio parahaemolyticus* strain O4:K68 | Japan | 2002 | 6,891 | AB043679.1 |
| VfO3K6 | *Vibrio parahaemolyticus* strain O3:K6 | Japan | 2000 | 8,784 | AB043678.1 |

We constructed a phylogenetic tree based on the amino acid sequences of Zots found in various *Vibrio* species, including the three Chilean non-toxigenic *V. parahaemolyticus* strains PMA2.15, PMA3.15 and PMC53.7 (Supplementary Table S1). As shown in Figure 1B, the Zot toxin present in PMA2.15 grouped with Zots of the pandemic strain and the *V. parahaemolyticus* phages O3:K6 and O3:K48. In contrast, the Zot toxin present in PMC53.7 was different in sequence and had greater similarity to sequences encoding Zots present in *V. campbellii* and other *V. parahaemolyticus* strains. The toxin of PMA 3.15 was the most different in sequence and shared an ancestor in common with the *V. cholerae* phage VFJ (Figure 1B). It should also be noted that the clade formed by the coding sequences for Zot present in *V. cholerae* does not include *V. parahaemolyticus* sequences (Figure 1B). Interestingly, the phylogenetic analysis of these Zot-like toxin proteins showed four different clusters, B1–B4 defined by Castillo et al., 2018b, where the B4 group included the Zot toxin encoded by prophage VfO3K6 identified in the pandemic *V. parahaemolyticus* clone. Our analysis showed that Zots found in Chilean non-toxigenic strains PMC53.7 and PMA2.15 belong to B4 while the Zot of PMA3.15 belongs to the B2 phylogenetic group (Figure 1B). Interestingly, all these *zot*-encoding prophages contained the Accessory Cholera Enterotoxin gene (*ace*), which has also been described in the CTX prophage.

**Multiple sequence alignment (MSA) in different Zot proteins and their Walker A and Walker B motifs**

To detect conserved patterns present in the Zot protein sequences of *V. parahaemolyticus*, an MSA was performed comparing Zot of different species of human pathogens. Alignment using Zot amino acid sequences from *C. concisus*, *N. meningitidis* and *V. cholerae* strains demonstrated that the two highly conserved domains of these proteins, named Walker A (GXXXXGK[S/T] where X is any residue) and Walker B (HHHH[D/E] where H is a hydrophobic residue) [27], were also present in *V. parahaemolyticus* strains (Figure 2, black squares).

*Preprints* (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 May 2019 doi:10.20944/preprints201905.0105.v1


**Figure 2.** Walker A and walker B motifs identified in different Zot proteins (*N. meningitidis* MC58, *V. cholerae* El Tor Inaba N16961, *C. concisus* 13826) are also present in all sequences

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 May 2019 doi:10.20944/preprints201905.0105.v1

of *V. parahaemolyticus* Chilean strains PMC53.7, PMA2.15 and PMA3.15. Both Walker motifs are marked with black squares.

We noticed that three *V. parahaemolyticus* strains sequences have a glycine changed to a tyrosine in the Walker A motif (GXXXX**Y**K[S/T]) as also was observed in *V. cholerae*. Both Walker motifs were located at the N-terminal side prior to the transmembrane domains (approximately 1–270, as defined for *V. cholerae* Zot). As these Walker A and Walker B motifs belong to the proteins of the p-loop containing the nucleoside triphosphate hydrolase (p-loop NTPase) superfamily, we used the entry identity IPR027417 in InterPro database [23]. As expected, we identified that the Zot proteins of *V. parahaemolyticus* had p-loop NTPase domains, the most prevalent domain of the several distinct nucleotide-binding protein folds (Figure 2, black squares). All sequences of the Zot proteins identified in *V. parahaemolyticus* aligned with 100% identity with the previously identified active domain of *V. cholerae* FCIGRL located in the C-terminal domain (Figure 3, black square) [19]. However, neither the other Zot sequences of *C. concisus* or *N. meningitidis* had the FCIGRL domain (Figure 3, black square). Di Pierro and coworkers showed that the eight amino acids shared by Zot and zonulin represent the putative receptor-binding site, characterized by the motif: non-polar (G)/variable/non-polar/variable/non-polar (V)/polar (Q)/variable/non-polar (G) [26]. They also showed that the glycine residue in position 298 has a key role in the activation of the intercellular TJ opening. We showed that PMC53.7, PMA2.15 and PMA3.15 do not have a glycine in position 298 (Table 2), but also *C. concisus*. Instead, PMA2.15 and PMA3.15 have a serine (S) and asparagine (N) residue, respectively, in this position (Table 2).

**Table 2. Amino acid sequence of *V. parahaemolyticus* and *C. concisus* strains aligned with the octapeptide suggested as the Zot putative receptor-binding site of *V. cholerae*.**

| Strain | Non-polar (G) | Variable | Non-polar | Variable | Non-polar (V) | Polar (Q) | Variable | Non-polar (G) Position 298* |
|---|---|---|---|---|---|---|---|---|
| *Vc N16961* | **G** | **R** | **L** | **C** | **V** | **Q** | **D** | **G** |
| Human zonulin | **G** | G | V | L | **V** | **Q** | P | **G** |
| *Vp* PMC53.7 | N | T | V | A | N | T | H | - |
| *Vp* PMA2.15 | E | R | W | H | K | A | T | S |
| *Vp* PMA3.15 | E | S | S | M | N | P | P | N |
| *Cc 13826* | T | C | L | N | N | N | C | - |

*Vp: V. parahaemolyticus; Vc: V. cholerae; Cc: C. concisus*

Interestingly, a later comparison of the Zot amino acid sequence found in PMC53.7 strain against known toxin databases using BTXpred [45] showed that the sequence matched with an endotoxic bacterial toxin, while PMA2.15 matched with an enterotoxin which actives the guanylate cyclase and PMA3.15 has not-matched. As control, the sequence of *V. cholerae* matches with an exotoxin (data not shown).

*Preprints* (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 May 2019 doi:10.20944/preprints201905.0105.v1

**Secondary structure of the C-terminal region of *V. parahaemolyticus* Zot proteins.**

As was described above, the active domain of *V. cholerae* FCIGRL was absent from Zot proteins of *V. parahaemolyticus*, *C. concisus* and *N. meningitidis*. However, the secondary structure of the β-sheet of sequences that aligned with this fragment using PROMALS3D alignment [28,29] was highly conserved among *V. cholerae*, *N. meningitidis*, *C. concisus* and all the *V. parahaemolyticus* strains (Figure 3).

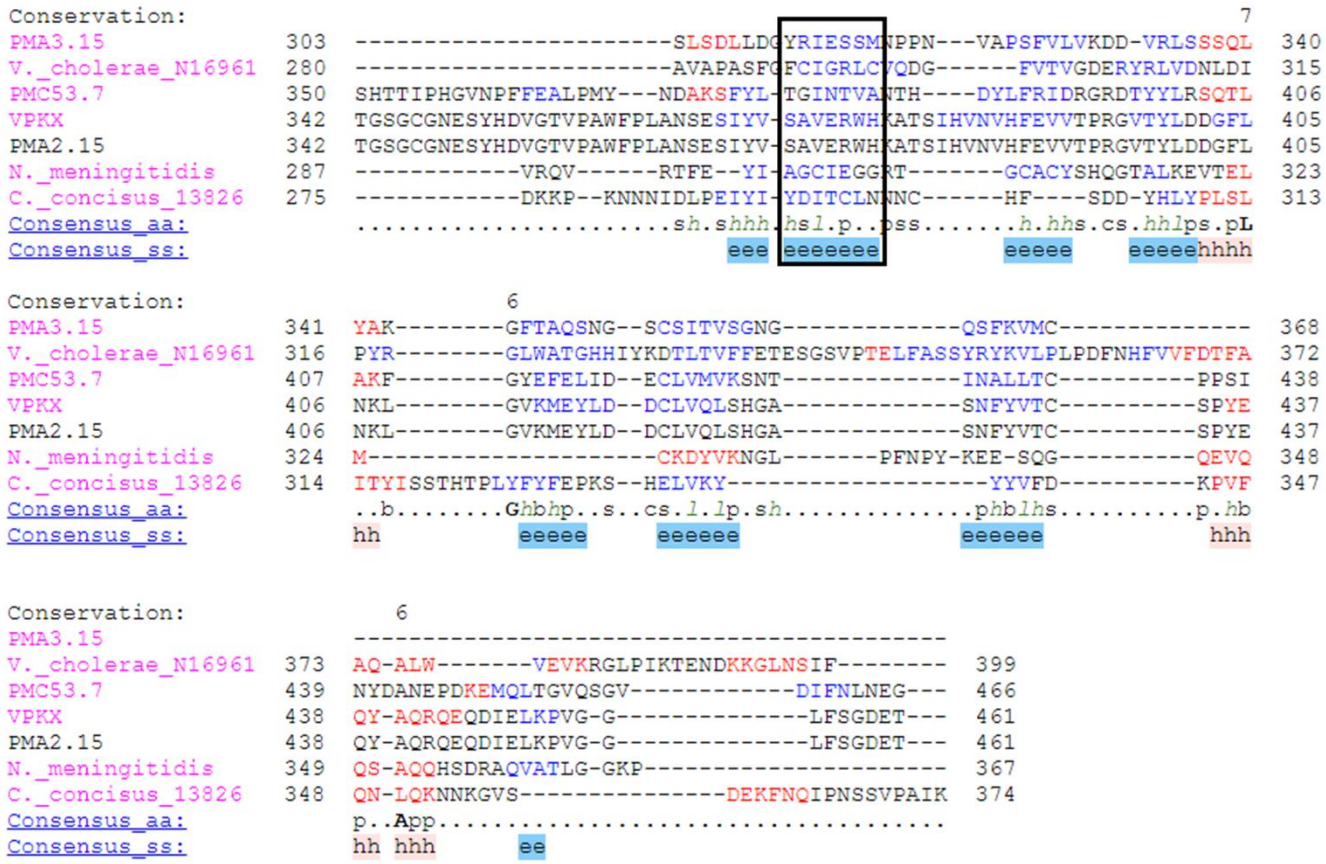



**Figure 3**. Zot alignment considering the C-terminal secondary structure of PMC53.7, PMA2.15, PMA3.15, VpKX and *N. meningitidis* MC58, *V. cholerae* El Tor Inaba N16961 and *C. concisus* 13826. The C-terminal was calculated based on the information of domains reported for *V. cholerae* [26]. The conserved structure of the α-helix and β-sheet can be observed under the alignment in red (hhhhhh) and blue (eeeee) respectively. The conserved structure of the FCIGRL fragment is marked with a black square. Complete alignment can be observed in Supplementary figure 1.

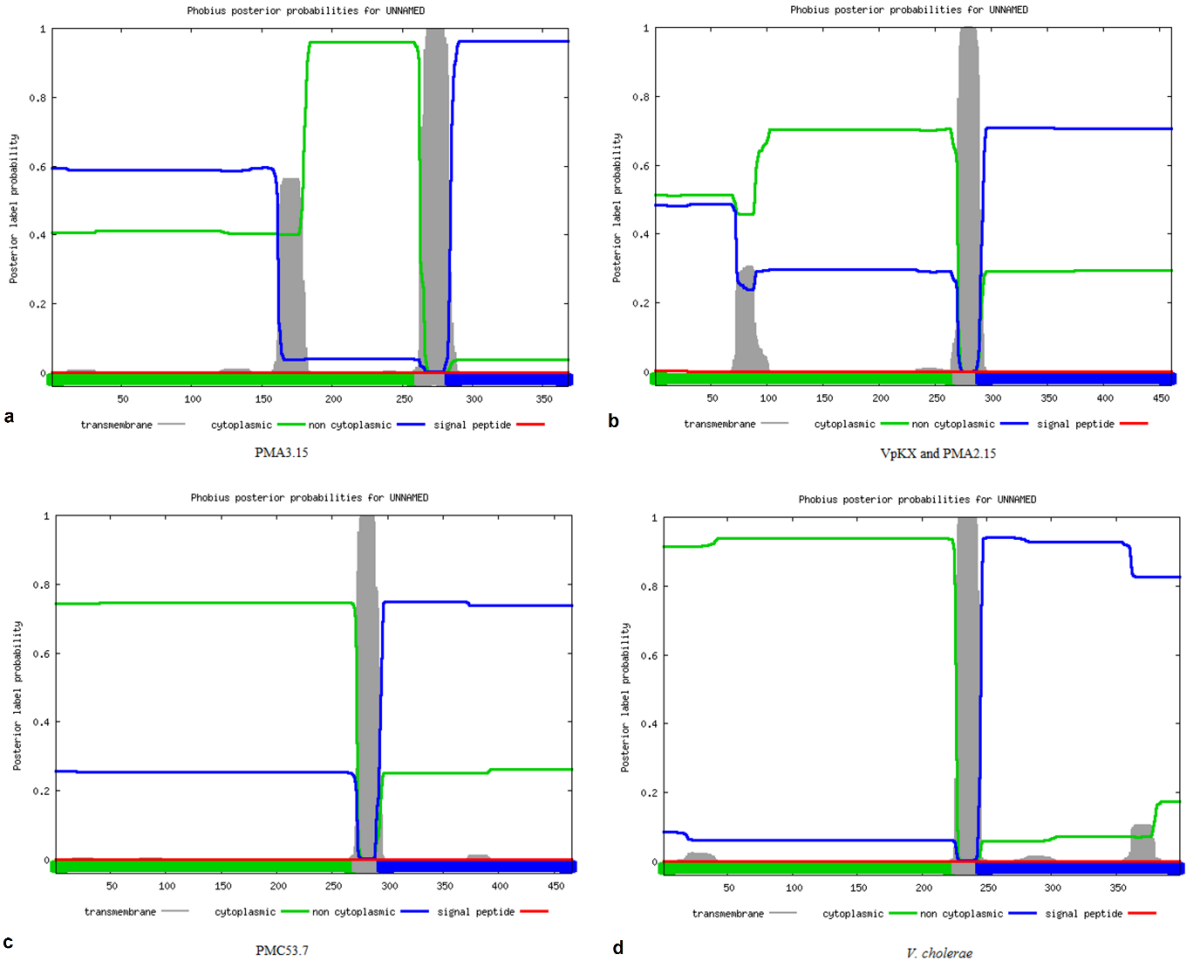
Interestingly, it has been previously suggested that the structure and not the sequence would be responsible for the biological effects of Zot on the epithelial barrier [30]. Considering the importance of the structure in the possible role of Zot, we performed a secondary structure

*Preprints* (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 May 2019 doi:10.20944/preprints201905.0105.v1

prediction using PSIPRED [31]. Zot proteins of *V. parahaemolyticus*, *V. cholerae* and *C. concisus* showed a conserved secondary structure in their C-terminal region (Supplementary Figure 2). Zots of PMC53.7, PMA2.15 and PMA3.15 contain 7, 8 and 9 α-helices, and 17, 19 and 17 β-strands respectively, compared to the 7 and 8 α-helices and 20 and 14 β-strands of *V. cholerae* and *C. concisus*. Interestingly, in all cases the region of the FCIGRL peptide of *V. cholerae* was part of conserved β-strand structure (see PMC53.7 in Supplementary Figure 2, PMA2.15 and PMA3.15 are not shown).

**Prediction of the transmembrane domain in the Zots of *V. parahaemolyticus***

Since the Zot of *V. cholerae* had 3 well-defined domains (amino-end, carboxy-end and transmembrane), we predicted the transmembrane domain Zots in *V. parahaemolyticus* strains using the Phobious server [32]. Similarly, all Zots of *V. parahaemolyticus* showed three defined domains: cytoplasmic in the N-terminal, non-cytoplasmic domain in the C-terminal and a transmembrane domain (Figure 4A, B and C). This last domain was predicted in amino acids 260 to 294 in the three Zots of *V. parahaemolyticus* (Figure 4A, B and C), while *V. cholerae* possesses this domain located within amino acid positions 227 to 245 (Figure 4D).



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 May 2019 doi:10.20944/preprints201905.0105.v1

**Figure 4.** Prediction of transmembrane domains in Zot of *V. parahaemolyticus* strains using the Phobius server. (a) *V. parahaemolyticus* PMA3.15*;* (b) *V. parahaemolyticus* VpKX and PMA2.15; (c) *V. parahaemolyticus* PMC53.7; (d) *V. cholerae* used as control.

**Structure prediction and 3D modeling**

The Zot protein homology was determined using HHpred [33]. Amino acids aligned with a probability of 99.72% in the N-terminal region of the "*Zonula Occludens* Toxin" from *N. meningitidis* MC58 (template 2R2A from the PDB database), which has a crystallized structure of 199 amino acids from its N-terminal domain. The tridimensional structure was predicted with Phyre2 [34] and Swiss-Modell [35] using the template 2R2A, which had sequence identity with the target protein. Since to date there is no model for the C-terminal region of this protein, we could only align the first 250 amino acids of *V. parahaemolyticus* Zots using the crystallized Zot of *N. meningitidis* as template (partially available N-terminal structure). A total of 8 three-dimensional models were obtained (Figure 5).

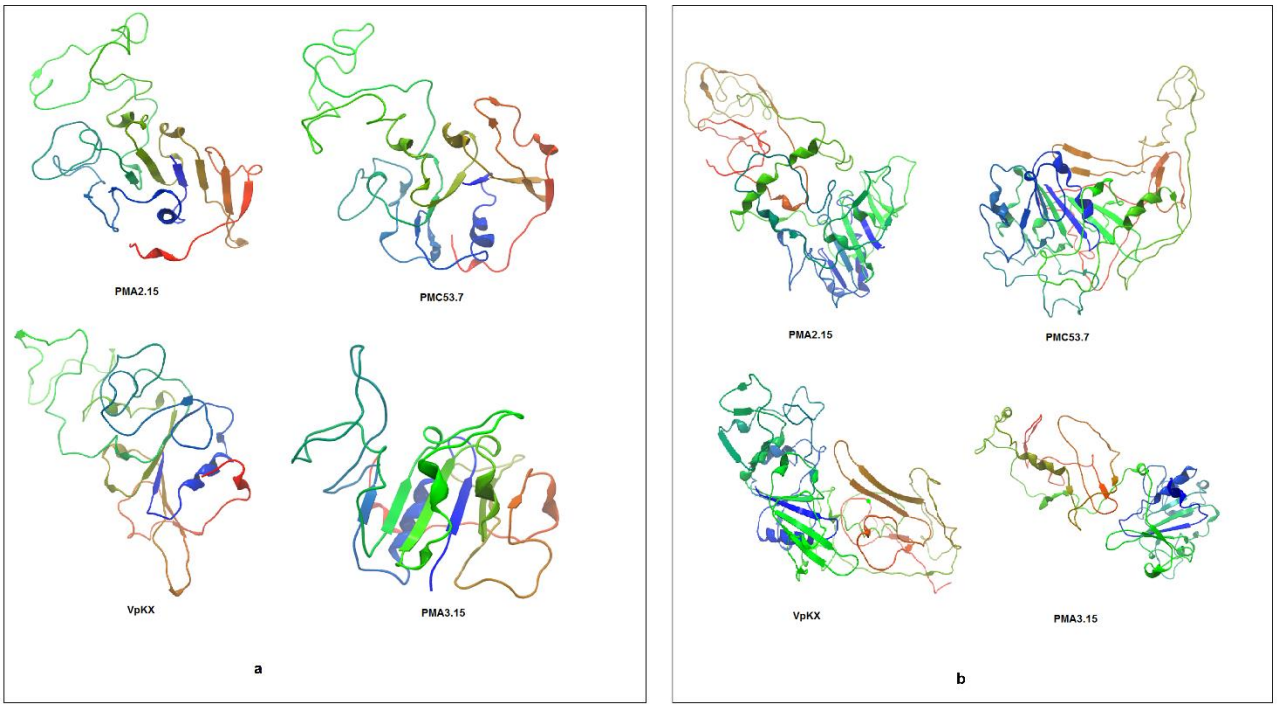



**Figure 5.** Tertiary structure of *V. parahaemolyticus* Zot. Crystallized Zot of *N. meningitidis* 2R2A was used as template. (a) Swiss-Model, (b) Phyre2

**Validity evaluation of the model**

To determine whether the model predicted by the homology modeling software is valid, quality and potential errors were calculated using the Protein Structure Analysis (ProSA) web server [36]. The quality index calculated by ProSA-web for a specific input structure is shown on a graph that gives the scores of all experimentally determined protein chains, and is currently available at PDB. This feature correlates the punctuation of a specific model with scores calculated from all

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 May 2019 doi:10.20944/preprints201905.0105.v1

experimental structures deposited in PDB. Z-scores outside a range characteristic for native proteins indicate erroneous structures [36,37]. The results obtained are shown in Table 3. Comparing with calculated z-score, the model is shown to be valid according to the quality of the structure (Figure 6).

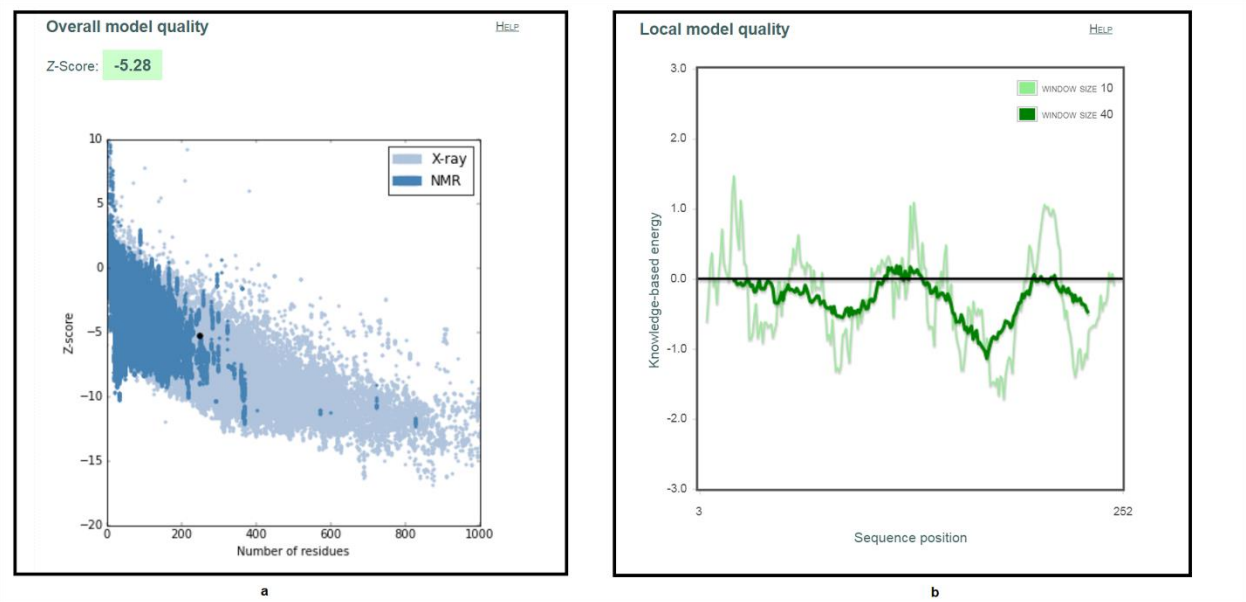



**Figure 6**. (a) ProSA analysis; Z-score plot (overall quality model) and (b) graphic plot (local model quality) of PMC53.7 Zot protein (Swiss-model 3D model).

**Table 3. ProSA-web z-score results for the Zot 3D structure**

| Z-score | PMC53.7 | PMA2.15 | PMA3.15 | VpKX |
|---|---|---|---|---|
| **Phyre2** | -4.21 | -2.91 | -4.47 | -3.75 |
| **Swiss-Model** | -5.28 | -2.57 | -3.64 | -2.57 |

To predict the stereochemical quality of the desired PMC53.7 protein model, PROCHECK server has been used [44]. From the server, in the Ramachandran plot, 168 residues (76.7 %) are plotted in the most favorable region (Figure 7) [44], which concludes the protein to be a good quality model.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 May 2019 doi:10.20944/preprints201905.0105.v1

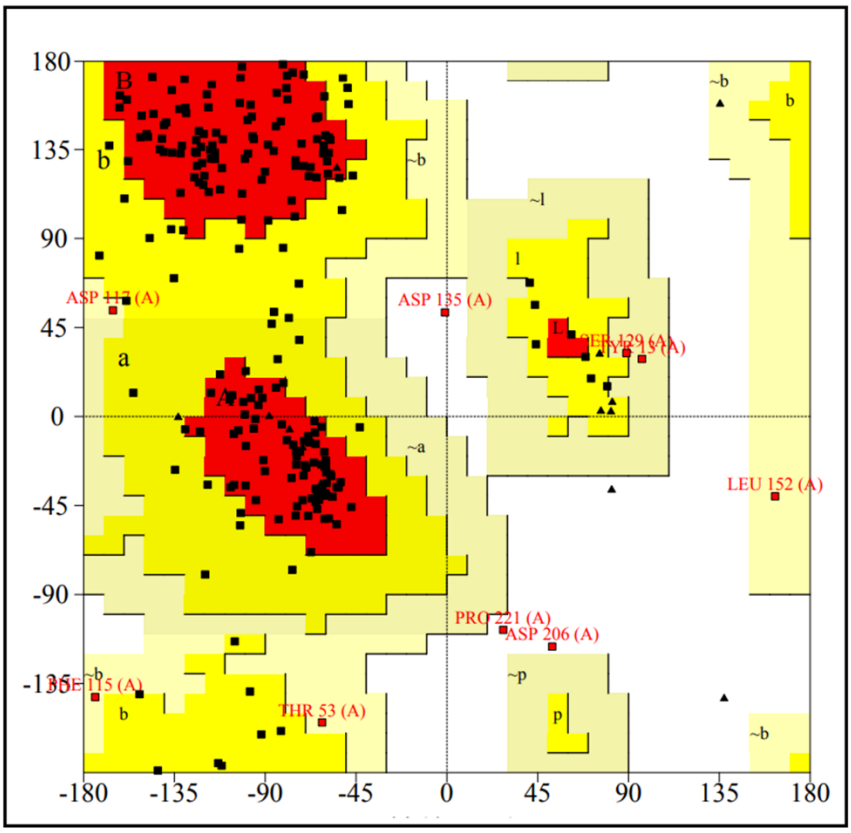



**Figure 7**. Ramachandran plot for PMC53.7 Zot (predicted 3D model) generated using PROCHECK server. The areas showing different colors i.e. red, yellow and light yellow represents most favored regions (76.7%), additional allowed regions (19.6%) and generously yellowed regions (2.7%) respectively. Residues in disallowed regions (0.9%).

## 3. Discussion

We are currently focused on the identification and characterization of new virulence factors that could explain the pathogenicity of non-toxigenic strains of *V. parahaemolyticus.* We have identified that some of these strains possess *zot* genes in their accessory genome associated with prophages and pathogenicity islands [11]. To understand better the relation among these toxins, we constructed a phylogenetic tree based on the amino acid sequences of Zots found in various *Vibrio* species [11]. The Zot toxin present in PMA2.15 grouped with Zot of the RIMD2210633 strain and Zots of other Vibriophages also found in pandemic strains. However, the Zot toxin present in PMC53.7 was different in sequence and had major similarity to sequences encoding Zots present in *V. campbellii* and other *V. parahaemolyticus* strains. It should also be noted that the clade formed by the coding sequences for Zot present in *V. cholerae* does not include *V. parahaemolyticus* sequences, which is interesting since the *V. cholerae* toxin Zot has been extensively studied and characterized, but this has not occurred in *V. parahaemolyticus.*

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 May 2019 doi:10.20944/preprints201905.0105.v1

Our previous studies demonstrated that *V. parahaemolyticus* Zot proteins had low (about 24% identity) similarity to *V. cholerae* Zot, but they share some conserved regions located toward the N-terminal domain [11]. One of these domains is the Pfam PF05707 Zot domain, named for the homologue in the Vibrio CTX phage, which is essential for the assembly and export of phage virions. All Zot proteins of *V. parahaemolyticus* also contained Walker A and Walker B motifs, which are conserved motifs of the p-loop NTPase superfamily [27]. P-loop NTPase binds to NTP, typically ATP or GTP, through the Walker A and B motifs. Specifically, the N-terminal of Zot is predicted to act as an ATPase, powering the assembly and transport of phages through the envelope, as has been observed for *Escherichia coli* Ff-type phages [38].

The change of a glycine (non-polar aliphatic amino acid) to a tyrosine (aromatic amino acid) into the Walker A motif (GXXXX**G**K[S/T]) observed in most of *V. parahaemolyticus* strains was also observed in *V. cholerae* (GXXXX**Y**K[S/T]) but non-observed in *V. parahaemolyticus* PMA3.15. This strain has a Glycine as also *N. meningitidis* and many *Campylobacter* species [23]. Despite this change, the Zot of *V. cholerae* maintains the functionality [15]. Also, a transmembrane domain was found in all *V. parahaemolyticus* Zot proteins, showing that Zots are transmembrane proteins in this species. We observed very low similarities (*a, b, c values*) among the C-terminal end of *V. parahaemolyticus* Zot proteins compared to *Campylobacter*, *V. cholerae* and *N. meningitidis* Zot proteins. No Zots of *V. parahaemolyticus* contained the active fragment described for *V. cholerae* in their sequences. However, since the Zot of *C. concisus*, lacking the FCIGRL fragment, also affects the paracellular pathway experimentally, it may be that the presence of this amino acid sequence would not be strictly necessary to perform the action of all Zot proteins. Similarly, a glycine in position 298 of *V. cholerae* Zot, with a crucial role in producing the opening of intracellular TJs, is also absent in *C. concisus*. These observations support that the structure and not the sequence would be responsible for the biological effects of Zot on the epithelial barrier. Indeed, all Zots of *V. parahaemolyticus* shared the secondary structure observed for the active fragment in *V. cholerae* and this β-sheet structure is also observed in *C. concisus*. In summary, our results showed high variability in the amino acid sequences of Zot proteins between different bacterial species and between strains of *V. parahaemolyticus*, but all of them shared a general similarity in the secondary structure. However, it must be considered that the comparison of Zot amino acid sequences of *V. parahaemolyticus* strains against known toxin databases suggests that not all Zots would have a function, or at least not the same function. Interestingly, the sequence of PMA3.15, having not-match with any toxin, was the most different in sequence.

Regrettably, although the tertiary structure was modeled we obtained results only with the first 250 amino acids (partial N-terminal structure) using crystallized Zot of *N. meningitidis* as template. The tertiary structure prediction indicated that Zot structures from these three bacterial species were highly variable. The accuracy of predicted models was confirmed through the online ProSA-web server. The scores determined by ProSA-web compare the tridimensional models obtained with existing models predicted by NMR or X-ray and verify the probability of mistakes that may exist in these predictions. The different models presented a *z*-score around -4.21 (Phyre2) and -5.29 (Swiss-model); this negative *z*-score is a reflection of the amino acid residues present in the

 doi:10.20944/preprints201905.0105.v1

N-terminal region. These values indicate that the predicted structures are located in the range of native protein and are close to the database average, allowing the choice of the better model. Altogether, our results provide useful information for further examination of *V. parahaemolyticus* Zot proteins as potential virulence factors in non-toxigenic strains. Future studies will be conducted to determine the role of Zot in *V. parahaemolyticus* and the mechanism that could affect human cells.

## 4. Materials and Methods

### Multiple sequence alignment (MSA) and phylogenetic analysis

The amino acid sequences of the Zot proteins from *V. cholerae*, *C. concisus* and *N. meningitidis* were obtained from UniprotKB. Multiple alignment was converted to PHYLIP format using Clustal Omega software [39]. It is shown in the InterPro database that the Zot family proteins (InterPro entry identity: IPR008900) belong to the p-loop NTPase superfamily. The proteins of the p-loop NTPase superfamily have Walker A and Walker B motifs [27]. Here we examined the presence of Walker A and Walker B motifs in *V. parahaemolyticus* Zot proteins by protein alignment using Clustal Omega software [39] and the PROMALS3D multiple sequence and structure alignment server [28,29] at http://prodata.swmed.edu/promals3d/promals3d.php. The Walker A motif has a sequence of GxxxxGK[S/T], where x is any residue, and the Walker B motif has a sequence of hhhh[D/E], where h is a hydrophobic residue [27]. The alignments for multiple protein sequences and secondary structure prediction were performed with PROMALS3D [28,29]. To reveal the phylogenetic relationship among genes encoding the identified Zot, amino acid sequences were aligned using ClustalW version 2.042 and phylogeny was inferred using Maximum Likelihood (1,000 bootstrap replicates) in Geneious version 10.1.338 [40].

### Sequence-based analyses of the Zot protein

Sequence-based analysis of the Zot protein was performed using different web-based tools: the InterProScan https://www.ebi.ac.uk/interpro/; the PSIPRED server [31] at http://bioinf.cs.ucl.ac.uk/psipred/; CD search tools [41,42] https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi and the NetPhos server [43] http://www.cbs.dtu.dk/services/NetPhos/ to detect domains and motifs, secondary structures, superfamily, and phosphorylation sites of the Zot protein, respectively. The prediction analysis of bacterial toxins was performed with the BTXpred program in SVM mode [45].

### Prediction of transmembrane domains in *V. parahaemolyticus* Zot

The Phobius server [32] was used for prediction of transmembrane topology and signal peptides at https://www.ebi.ac.uk/Tools/pfa/phobius/.

### Construction of the Zot protein structure using homology modelling

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 May 2019 doi:10.20944/preprints201905.0105.v1

Homology modeling of selected protein sequences was performed using Phyre2 [34] at http://www.sbg.bio.ic.ac.uk/phyre2/, the SWISS-MODEL automated comparative modeling server [35] at https://swissmodel.expasy.org/. Protein structure files were compiled from the protein data bank available at http://www.rcsb.org/pdb. Protein structures were viewed using CLC Sequence Viewer version 8.0.

**Validation of the Zot protein model generated**

The internal consistency and reliability of the model of the Zot protein (Swiss- Model and Phyre2) were evaluated using the ProSA-web server [36] at (https://prosa.services.came.sbg.ac.at/). The ProSA-web calculates the general quality score for a 3D structure. If the calculated scores are outside the native protein range, an error in the predicted structure is indicated. The overall quality of the targeted protein is validated with a graphical output map of local quality estimates. The z-score for homology modeling was calculated with a graphical plot, where the X-ray and NMR data from all the known protein sequences commencing to the PDB database are clearly depicted. Accordingly the residue score has been calculated using each amino acid sequence positions. Both 10 amino acid residue and 40 amino acid residue energy data are calculated for further consideration in experimental and theoretical structure validation [36]. Parallely the ProCheck server at http://servicesn.mbi.ucla.edu/PROCHECK/ has been applied for validation of the stereochemical quality of proteins structure using the Ramachandran plot [44].

**Supplementary Materials**

**Figure 1S.** Complete alignment of secondary structure of PMC53.7, PMA2.15, PMA3.15, VpKX and *N. meningitidis* MC58, *V. cholerae* El Tor Inaba N16961 and *C. concisus* 13826. The C-terminal was calculated based on the information of domains reported for *V. cholerae* [26]. The conserved structure of the α-helix and *B*-sheet can be observed under the alignment in red (hhhhhh) and blue (eeeee) respectively. The conserved structure of the FCIGRL fragment is marked with a black square.

**Figure 2S.** The secondary structure composition of the Zot protein of *V. cholerae*, *C. concisus* and *V. parahaemolyticus* PMC53.7 was predicted using the PSIPRED web server. The pink cylinders represent helices, the yellow arrows represent β-strands, and the black lines represent coiled structures. The height and color intensity of the blue bars indicate the confidence (conf) of the prediction. Pred: predicted secondary structure; AA: target sequence.

**Table 1S.** Amino acid sequences of Zots found in various *Vibrio* species.

**Author contribution**

KG conceived the idea; KG and DPR designed the experiments and wrote the manuscript. DC, SRA and NP made the phylogenetic analysis and comparison of prophages, VJ, LP and PN

performed the sequence alignment analysis and searching of conserved motifs, GC and RB performed the secondary structure analysis and determined the position of transmembrane domain, AP and CJB made the 3D structure prediction and validated the 3D models. All the authors read, discussed and approved the final version of this manuscript.

## Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Funding

The authors acknowledge Fondecyt Iniciación 11140257 and 11160901, Fondecyt Regular 1190957, CONICYT, Chile; REDI170296 and Competitive Funds of Universidad de Las Américas PI2018026, Chile.

## References

1. Letchumanan, V.; Chan, K.; Khan, T.M.; Bukhari, S.I.; Mutalib, N.A.; Goh, B.; Lee, L.; Lee, L. Bile Sensing : The Activation of *Vibrio parahaemolyticus* Virulence. **2017**, *8*, 1–6.
2. Raghunath, P. Roles of thermostable direct hemolysin (TDH) and TDH-related hemolysin (TRH) in *Vibrio parahaemolyticus*. **2015**, *5*, 2010–2013.
3. Shinoda, S. Sixty Years from the Discovery of *Vibrio parahaemolyticus* and Some Recollections. *Biocontrol Sci.* **2011**, *16*, 129–137.
4. Nishibuchi, M.; Fasano, A.; Russell, R.G.; Kaper, J.B.; Pediatria, C.; Pugliese, O. Enterotoxigenicity of *Vibrio parahaemolyticus* with and without Genes Encoding Thermostable Direct Hemolysin. **1992**, *60*, 3539–3545.
5. Zhang, L.; Orth, K. Virulence determinants for *Vibrio parahaemolyticus* infection. *Curr. Opin. Microbiol.* **2013**, *16*, 70–77.
6. Broberg, C.A.; Calder, T.J.; Orth, K. *Vibrio parahaemolyticus* cell biology and pathogenicity determinants. *Microbes Infect.* **2011**, *13*, 992–1001.
7. Yu, Y.; Yang, H.; Li, J.; Zhang, P. Putative type VI secretion systems of *Vibrio parahaemolyticus* contribute to adhesion to cultured cell monolayers. **2012**, 827–835.
8. Ceccarelli, D.; Hasan, N.A.; Huq, A.; Colwell, R.R. Distribution and dynamics of epidemic and pandemic *Vibrio parahaemolyticus* virulence factors. *Front. Cell. Infect. Microbiol.* **2013**, *3*, 1–9.
9. Mahoney, J.C.; Gerding, M.J.; Jones, S.H.; Whistler, C.A. Comparison of the Pathogenic Potentials of Environmental and Clinical *Vibrio parahaemolyticus* Strains Indicates a Role for Temperature Regulation in Virulence. **2010**, *76*, 7459–7465.
10. Wagley, S.; Borne, R.; Harrison, J.; Baker-Austin, C.; Ottaviani, D.; Leoni, F. *Galleria mellonella* as an infection model to investigate virulence of *Vibrio parahaemolyticus*. **2018**, *9*, 197–207.
11. Castillo, D.; Pérez-Reytor, D.; Plaza, N.; Ramírez-Araya, S.; Blondel, C.J.; Corsini, G.; Bastías, R.; Loyola, D.E.; Jaña, V.; Pavez, L.; et al. Exploring the genomic traits of non-

toxigenic *Vibrio parahaemolyticus* strains isolated in southern Chile. *Front. Microbiol.* **2018**, *9*, 1–15.

12. Pérez-Reytor, D.; García, K. *Galleria mellonella* : A model of infection to discern novel mechanisms of pathogenesis of non-toxigenic *Vibrio parahaemolyticus* strains. *Virulence* **2018**, *9*, 22–24.
13. Pérez-reytor, D.; Jaña, V.; Pavez, L.; Navarrete, P.; García, K. Accessory Toxins of *Vibrio* Pathogens and Their Role in Epithelial Disruption During Infection. **2018**, *9*, 1–11.
14. Fasano, A. Toxins and the gut: role in human disease. **2002**, 9–14.
15. Schmidt, E.; Kelly, S.M.; Walle, C.F. Van Der Tight junction modulation and biochemical characterization of the *zonula occludens* toxin C-and N-termini. **2007**, *581*, 2974–2980.
16. Uzzau, S. Purification and preliminary characterization of the *zonula occludens* toxin receptor from human (CaCo2) and murine (IEC6) intestinal cell lines. **2001**, *194*, 1–5.
17. Mahendran, V.; Liu, F.; Riordan, S.M.; Grimm, M.C.; Tanaka, M.M.; Zhang, L. Examination of the effects of *Campylobacter concisus zonula occludens* toxin on intestinal epithelial cells and macrophages. *Gut Pathog.* **2016**, 1–10.
18. Gopalakrishnan, S.; Pandey, N.; Tamiz, A.P.; Vere, J.; Carrasco, R.; Somerville, R.; Tripathi, A.; Ginski, M.; Paterson, B.M.; Alkan, S.S. Mechanism of action of ZOT-derived peptide AT-1002 , a tight junction regulator and absorption enhancer. **2009**, *365*, 121–130.
19. Goldblum, S.E.; Rai, U.; Tripathi, A.; Thakar, M.; Leo, L. De; Toro, N. Di; Not, T.; Ramachandran, R.; Puche, A.C.; Hollenberg, M.D.; et al. The active Zot domain (aa 288 – 293) increases ZO-1 and myosin 1C serine / threonine phosphorylation, alters interaction between ZO-1 and its binding partners, and induces tight junction disassembly through proteinase activated receptor 2 activation.
20. Vanuytsel, T.; Vermeire, S.; Cleynen, I. The role of Haptoglobin and its related protein, Zonulin, in inflammatory bowel disease. **2013**, 1–9.
21. Zhang, L.; Lee, H.; Grimm, M.C.; Riordan, S.M.; Day, A.S.; Lemberg, D.A. *Campylobacter concisus* and inflammatory bowel disease. **2014**, *20*, 1259–1267.
22. Kaakoush, N.O.; Mitchell, H.M.; Man, S.M. Role of Emerging *Campylobacter* Species in Inflammatory Bowel Diseases. **2014**, *20*, 2189–2197.
23. Liu, F.; Lee, H.; Lan, R.; Zhang, L. Zonula occludens toxins and their prophages in *Campylobacter* species. *Gut Pathog.* **2016**, 1–11.
24. Castillo, D.; Kauffman, K.; Hussain, F.; Kalatzis, P.; Rørbo, N.; Polz, M.F.; Middelboe, M. Widespread distribution of prophage-encoded virulence factors in marine Vibrio communities. *Sci. Rep.* **2018**, *8*, 2–10.
25. Fasano, A.; Fiorentini, C.; Donelli, G.; Uzzau, S.; Kaper, J.B.; Margaretten, K.; Ding, X.; Guandalini, S.; Comstock, L.; Goldblum, S.E. *Zonula occludens* toxin modulates tight junctions through protein kinase C-dependent actin reorganization, *in vitro*. *J. Clin. Invest.* **1995**, *96*, 710–720.
26. Di Pierro, M.; Lu, R.; Uzzau, S.; Wang, W.; Margaretten, K.; Pazzani, C.; Maimone, F.; Fasano, A. Zonula Occludens Toxin Structure-Function Analysis. *J. Biol. Chem.* **2002**, *276*, 19160–19165.
27. Hanson, P.I.; Whiteheart, S.W. Aaa+ proteins: have engine, will work. **2005**, *6*, 519–529.
28. Pei, J.; Kim, B.; Grishin, N. V PROMALS3D : a tool for multiple protein sequence and structure alignments. **2008**, *36*, 2295–2300.
29. Pei, J.; Grishin, N. V PROMALS3D: Multiple Protein Sequence Alignment Enhanced with Evolutionary and Three-Dimensional Structural Information. In; 2014; pp. 263–271 ISBN

9781627036467.

30. Kaakoush, N.O.; Man, S.M.; Lamb, S.; Raftery, M.J.; Wilkins, M.R.; Mitchell, H. The secretome of *Campylobacter concisus*. **2010**, *277*, 1606–1617.
31. Mcguffin, L.J.; Bryson, K.; Jones, D.T. The PSIPRED protein structure prediction server. **2000**, *16*, 404–405.
32. Käll, L.; Krogh, A.; Sonnhammer, E.L. A Combined Transmembrane Topology and Signal Peptide Prediction Method. *J. Mol. Biol.* **2004**, *338*, 1027–1036.
33. Zimmermann, L.; Stephens, A.; Nam, S.; Rau, D.; Kübler, J.; Lozajic, M.; Gabler, F.; Söding, J.; Lupas, A.N.; Alva, V. A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J. Mol. Biol.* **2018**, *430*, 2237–2243.
34. Kelley, L.A.; Mezulis, S.; Yates, C.M.; Wass, M.N.; Sternberg, M.J.E. Europe PMC Funders Group The Phyre2 web portal for protein modelling , prediction and analysis. **2017**, *10*, 845–858.
35. Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F.T.; Beer, T.A.P. De; Rempfer, C.; Bordoli, L.; et al. SWISS-MODEL : homology modelling of protein structures and complexes. **2018**, *46*, 296–303
36. Wiederstein, M.; Sippl, M.J. ProSA-web : interactive web service for the recognition of errors in three-dimensional structures of proteins. **2007**, *35*, 407–410.
37. Droppa-Almeida, D.; Franceschi, E.; Padilha, F.F. Immune-informatic analysis and design of peptide vaccine from multi-epitopes against Corynebacterium pseudotuberculosis. *Bioinform. Biol. Insights* **2018**, *12*, 25–29.
38. Feng JN, Russel M, Model P A permeabilized cell system that assembles filamentous bacteriophage. **1997**, *94*, 4068–4073.
39. Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T.J.; Karplus, K.; Li, W.; Lopez, R.; Thompson, J.D.; Higgins, D.G.; Mcwilliam, H.; et al. Fast , scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. **2011**.
40. Kearse, M.; Moir, R.; Wilson, A.; Stones-havas, S.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C.; Thierer, T.; et al. Geneious Basic : An integrated and extendable desktop software platform for the organization and analysis of sequence data. **2012**, *28*, 1647–1649.
41. Marchler-bauer, A.; Bryant, S.H. CD-Search : protein domain annotations on the fly. **2004**, *32*, 327–331.
42. Marchler-bauer, A.; Bo, Y.; Han, L.; He, J.; Lanczycki, C.J.; Lu, S.; Chitsaz, F.; Derbyshire, M.K.; Geer, R.C.; Gonzales, N.R.; et al. CDD / SPARCLE : functional classification of proteins via subfamily domain architectures. **2017**, *45*, 200–203.
43. Blom, N.; Sicheritz-Pontén, T.; Gupta, R.; Gammeltoft, S.; Brunak, S. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics* **2004**, *4*, 1633–1649.
44. Laskowski, R.A.; Rullmann, J.A.C.; MacArthur, M.W.; Kaptein, R.; Thornton, J.M. AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **1996**, *8*, 477–486.
45. Saha, S.; Raghava, G.P.S. BTXpred: prediction of bacterial toxins. *In Silico Biol.* **2007**, *7*, 405–12.