

# Predicted structure of the 89-kDa invasion-tip protein of *Holospira obtusa*, an endonuclear symbiont of *Paramecium caudatum*, suggests a pH-dependent conformational change

Yasuhiro Takenaka <sup>1</sup>, Toshiyuki Murai <sup>2</sup>, Toru Nakatsu <sup>3</sup>, Yasushi Shigeri <sup>4</sup> and Masahiro Fujishima <sup>5,6,\*</sup>

<sup>1</sup> Department of Bioregulatory Science, Graduate School of Medicine, Nippon Medical School, 1-25-16 Nezu, Tokyo 113-0031, Japan; yasuihiro-takenaka@nms.ac.jp

<sup>2</sup> Graduate School of Medicine, The University of Osaka, 2-2 Yamada-oka, Suita, Osaka 565-0871, Japan; murai@or-gctl.med.osaka-u.ac.jp

<sup>3</sup> School of Pharmaceutical Sciences, Wakayama Medical University, 25-1 Shichibancho, Wakayama, Wakayama 640-8156, Japan; nakatsu@wakayama-med.ac.jp

<sup>4</sup> Department of Chemistry, School of Medicine, Wakayama Medical University, 580 Mikazura, Wakayama, Wakayama 641-0011, Japan; yshigeri@wakayama-med.ac.jp

<sup>5</sup> Research Center for Thermotolerant Microbial Resources, Yamaguchi University, Yoshida 1677-1, Yamaguchi 753-8512, Japan; fujishim@yamaguchi-u.ac.jp

<sup>6</sup> Institute of Environmental Radioactivity, Fukushima University, Kanayagawa 1, Fukushima 960-1296, Japan

\* Correspondence: fujishim@yamaguchi-u.ac.jp

## Abstract

The endonuclear symbiont *Holospira obtusa* invades the macronucleus of *Paramecium caudatum* using an infectious-form invasion-tip containing a major 89-kDa periplasmic protein implicated in escape from the host digestive vacuole and macronuclear invasion. Sequence-based analyses placed this protein within a *Holospira*-specific family and predicted an N-terminal membrane-anchoring module and a long  $\alpha$ -helical C-terminal region. AlphaFold2 modeling followed by molecular dynamics indicated that the C-terminal region forms an antiparallel, spectrin-repeat-like helical solenoid containing a localized three-helix coiled-coil. The protein exhibits a large pH-dependent charge difference, becoming 55 units more positive at pH 4 than at pH 7. Independent simulations at these pH values differed mainly in the relative orientation of the N-terminal core and C-terminal rod, described by DynDom as an approximately 35° rigid-body rotation about a hinge centered on Gly291. However, none of the nearby glutamates (Glu273, Glu275, and Glu302) formed a pH-dependent salt bridge, and no specific protonation-dependent interaction responsible for the conformational difference was identified. We nonetheless propose, as a testable hypothesis, that acidification of the digestive vacuole may promote this reorientation, thereby repositioning the N-terminal membrane-anchoring module at the bacterial surface as an early step toward escape from the vacuole.

**Keywords:** *Holospira obtusa*; *Paramecium caudatum*; digestive vacuole; endosymbiosis; invasion-tip protein; pH-dependent conformational change; molecular dynamics; DynDom; spectrin-repeat; AlphaFold2

## 1. Introduction

*Holospira obtusa* is a Gram-negative alphaproteobacterial endonuclear symbiont of the ciliate *Paramecium caudatum*. A highly specialized infection strategy has evolved that enables it to invade the host macronucleus while surviving entirely within the intracellular environment [1–6]. Unlike free-

living bacteria, *Holospira* species have undergone substantial genome reduction [2,3,7] and cannot be cultivated outside their host cells, which indicates a high degree of adaptation to intracellular life. During their life cycle, they show two morphologically different forms: a reproductive short form (RF, 1.5–2  $\mu\text{m}$  in length) and an infectious long form (IF, 10–15  $\mu\text{m}$  in length) [8–10]. When the host is growing by binary fission, the RF cell proliferates by binary fission within the macronucleus. If the host cell starves, the RF ceases the binary fission and differentiates into the IF cell via intermediate forms. The IF cell forms a characteristic structure consisting of two regions: one containing cytoplasm and the other a large periplasmic region with an electron-lucent invasion-tip [1,11–15]. Infection can be initiated synchronously by mixing isolated IF cells from host cell homogenates [16] with uninfected host cells. The *Holospira*–*Paramecium* system therefore provides an exceptional experimental model for investigating the earliest molecular events that enable endosymbiotic bacteria to recognize environmental cues, escape from a host digestive compartment, and invade its specific intracellular niche.

Shortly after uptake by the host, IF cells are enclosed within newly formed digestive vacuoles (DVs). DVs of *Paramecium* species have been classified into four stages [16–18]. The DV-I is a young vacuole produced by pinching off from a cytopharynx. The DV-I fuses with acidosomes and transforms into a contracted and acidified DV-II (inside  $\text{pH} < 3$ ), which then fuses with primary lysosomes and differentiates to DV-III. DV-IV is the final stage of the DV before defecation and lacks acid phosphatase activity. In the DV-I, the IF cells rapidly undergo a series of coordinated morphological and physiological changes that transform them into activated forms (AFs), and this transformation is inducible if IFs are treated with a low pH buffer ( $\text{pH} 4$ ) for 15 min [19]. The AF cell escapes only from the DV-I, migrates through the host cytoplasm, and penetrates the macronuclear envelope by leading with the invasion-tip. Escape occurs only during the brief period in which the DV lumen is undergoing acidification. Bacterial escape is completely blocked once the vacuole matures to the more acidic DV-II stage. Furthermore, inhibition of vacuolar acidification by the V-type  $\text{H}^+$ -ATPase inhibitors concanamycin A or concanamycin B abolishes both escape from the DV and subsequent nuclear infection [20]. These observations have established transient acidification in the DV-I as the essential environmental signal that initiates the infection process. However, the bacterial molecule that senses this pH change and converts it into the structural changes required for membrane engagement and host-cell invasion is unknown.

To identify the bacterial factor responsible for this acid-triggered activation process, we previously generated a monoclonal antibody against a major 89-kDa protein localized specifically within the lumen of the invasion-tip and subsequently identified its full-length gene [14]. Sequence analysis revealed a hydrophobic N-terminal segment, annotated at the time as a signal peptide, followed immediately by two putative actin-binding motifs (WH2 and FH2). This suggests that this protein might directly connect environmental sensing with interactions between the bacterial surface and the host cytoskeleton. Immunofluorescence microscopy further demonstrated that, although the protein is initially distributed uniformly throughout the invasion-tip lumen in freshly isolated IFs, its N-terminal region rapidly becomes exposed on the outer surface of the bacterial outer membrane immediately after bacterial uptake into host DVs. At the same stage, transmission electron microscopy revealed numerous filamentous structures bridging the bacterial surface and the DV membrane [14]. Following escape into the host cytoplasm, host actin accumulated around the exposed N-terminal region [16,21], and similar actin-associated structures remained at the site where the bacterium penetrated the macronuclear envelope [16]. Together, these observations identified the 89-kDa protein as a central component throughout the entire infection process, including escape from the DV, migration through the host cytoplasm, and penetration of the target nucleus.

Nonetheless, one fundamental question has remained unanswered. Although the N-terminal region of the 89-kDa protein clearly becomes exposed during infection, the molecular mechanism by which acidification of the DV triggers this dramatic structural rearrangement has not been determined. Specifically, there has been no explanation of how exposure to the acidic environment of the

DV-I converts a protein that is initially confined within the invasion-tip lumen into one capable of interacting with the bacterial outer membrane and, subsequently, with the host actin cytoskeleton. The molecular basis linking DV acidification to activation of the invasion machinery has thus remained unresolved for nearly two decades.

To elucidate the molecular mechanism underlying acid-triggered activation of *H. obtusa* infection, we investigated the three-dimensional structure and pH-dependent conformational dynamics of the 89-kDa periplasmic protein using AlphaFold2-based structural prediction and molecular dynamics simulations. Our objective was to determine whether pH-dependent conformational changes in this protein could provide a structural basis for activation of the bacterial invasion machinery. In addition, to determine whether proteins with similar amino acid sequences have been identified in other organisms, we also performed an updated homology search of the 89-kDa protein sequence using current sequence databases.

## 2. Materials and Methods

### 2.1. Phylogenetic Analyses

A BLASTP search against the nr database (accessed July 2026; default parameters) was used to assess similarity to proteins inside and outside the genus *Holospora*. Multiple alignment was performed using MUSCLE with default parameters in MacVector version 18.8.3, using a log-expectation profile. The best-fit substitution model was selected by model testing in MEGA version 12.1.2 [22], and the JTT+G+I model, which showed the lowest Bayesian information criterion score, was used for tree construction.

### 2.2. Computational Modeling and Molecular Dynamics Simulations

The three-dimensional structure of the 89-kDa periplasmic protein (BAE80122.1; 750 residues) was predicted using ColabFold [23] with AlphaFold2 [24] default parameters (run October 2025). The top-ranked model was selected for subsequent simulations.

pH-dependent protonation states were assigned using PDB2PQR version 3.7.1 with PROPKA [25,26] at pH 4.0 and pH 7.0. At pH 4.0, seven histidine residues (His96, His133, His364, His542, His588, His615, His735) were treated as doubly protonated (HIP), eight aspartate residues (Asp125, Asp225, Asp411, Asp444, Asp551, Asp585, Asp609, Asp611) were assigned as neutral protonated forms (ASH), and 39 glutamate residues were treated as neutral protonated forms (GLH), resulting in a net charge shift of +55 elementary charges relative to the pH 7.0 state.

Each protonation variant was parameterized with the AMBER ff14SB force field [27] and solvated in an explicit TIP3P water box [28] with a minimum padding of 10 Å between the protein surface and the box boundary. Counter-ions (Cl<sup>-</sup>) were added to electro-neutralize the system (59 Cl<sup>-</sup> for pH 4.0; 4 Cl<sup>-</sup> for pH 7.0). MD simulations were performed using OpenMM version 8.3.1 [29]. Each system was energy-minimized (up to 10,000 iterations), followed by stepwise heating from 0 to 300 K in increments of 50 K (5,000 steps per increment at 1 fs timestep). Equilibration was then carried out for 100 ps at 300 K with a 2 fs timestep. Production runs of 10 ns were performed under NPT conditions (300 K, 1 atm) using a LangevinMiddle integrator (friction coefficient 1 ps<sup>-1</sup>, timestep 2 fs) with a Monte Carlo barostat (coupling every 25 steps). Particle mesh Ewald summation [30] was used for long-range electrostatics (cutoff 1.0 nm, Ewald error tolerance 0.0005), with hydrogen bond lengths constrained and rigid water geometry enforced. Coordinates were saved every 10 ps, yielding 1,000 frames per simulation.

### 2.3. Secondary-structure, Coiled-coil and Membrane-topology Prediction

The amino-acid sequence of the 89-kDa periplasmic protein of *H. obtusa* (750 aa) was analyzed with a panel of sequence-based predictors using the MPI Bioinformatics Toolkit (Tübingen, Germany) [31]. Consensus secondary structure was obtained with Quick2D, which integrates PSIPRED [32],

SPIDER3 [33], PSSpred, DeepCNF/RaptorX-Property [34] and NetSurfP-2.0 [35]. Coiled-coil propen-  
sity was obtained with COILS and PCOILS [36,37] (window size 28) and MARCOIL [38]; transmem-  
brane segments with TMHMM 2.0 [39]. Intrinsic disorder was obtained with NetSurfP-2.0,  
DISOPRED3 [40], SPOT-Disorder and IUPred. Coiled-coil propensity was additionally evaluated with  
DeepCoil2 [41], reporting residues with a coiled-coil probability  $\geq 0.5$ . A predicted signal peptide  
(PSG, GvH), transmembrane segment (ALOM), membrane topology (MTOP), coiled-coil region (Lu-  
pas COIL) and subcellular localization (k-NN) were obtained with PSORT II [42]. All servers were run  
with default parameters (accessed June 2026).

Structure-based analyses were performed on the final frame of the pH 7 molecular-dynamics tra-  
jectory. Secondary structure was assigned from  $\alpha$  coordinates using the P-SEA method [43], giving a  
three-state (helix/strand/coil) assignment equivalent to DSSP [44]. Knobs-into-holes packing was iden-  
tified with SOCKET2 (SOCKET v3.02) [45] using a packing-cutoff of 7.0 Å on the heavy-atom model.  
 $\alpha$ - $\alpha$  contact maps (contacts defined at  $< 12$  Å) and an internal structural-repeat scan of the C-termi-  
nal rod (residues 198–720), based on sliding-window superposition and contact-map autocorrelation,  
were computed in-house with Python 3 and NumPy to detect tandem antiparallel  $\alpha$ -helical modules.

## 2.4. Homology Searches

Remote homology was assessed with HHpred (MPI Bioinformatics Toolkit; accessed May 2026)  
using the amino-acid sequence of BAE80122.1 as the query. Structural similarity was assessed with  
Foldseek (databases AFD50 and MGnify-ESM Atlas), using the final frame of the pH 7 molecular-  
dynamics trajectory as the query structure. The ten highest-scoring hits per structural class or data-  
base are listed in Tables S1 and S2, respectively.

## 2.5. Structural Analysis

The final frame of each 10-ns trajectory was extracted as the representative equilibrated structure  
(Files S1 and S2). Per-residue  $\alpha$  RMSF was calculated from the last 500 frames (final 5 ns) of each tra-  
jectory. Structures were superimposed on the C-terminal rod or on the N-terminal core using Biopy-  
thon's Superimposer module prior to  $\alpha$  displacement calculations. Domain boundaries and hinge  
regions were identified using DynDom [46] (<http://dyndom.cmp.uea.ac.uk/dyndom/>). Structural fig-  
ures were prepared with UCSF ChimeraX version 1.10.1.

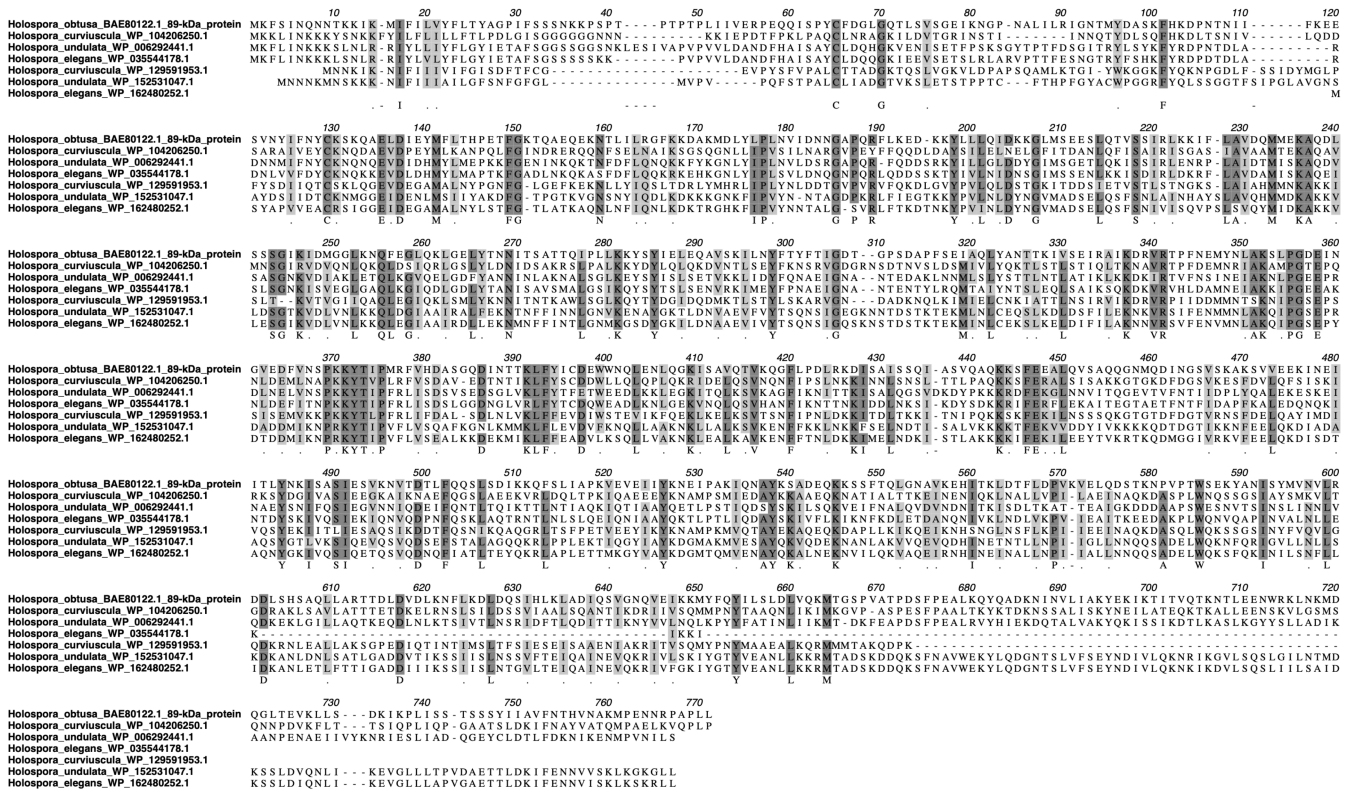
# 3. Results

## 3.1. Phylogenetic analysis of the 89-kDa periplasmic protein from *Holospira obtusa*

To examine the evolutionary relationship of the 89-kDa periplasmic protein from *Holospira obtusa*  
BAE80122.1, we first searched for homologous amino acid sequences and performed multiple se-  
quence alignment. A BLASTP search of BAE80122.1 against the nr database using default parameters  
showed no hits to proteins outside the genus *Holospira*, so subsequent analyses were restricted to  
*Holospira* homologs. The alignment revealed several conserved amino acid residues or short con-  
served regions among the *Holospira* proteins, suggesting that these residues may contribute to struc-  
tural or functional features shared by this protein family (Figure 1A).

We next constructed a maximum-likelihood phylogenetic tree based on the amino acid align-  
ment. Bootstrap analysis with 1,000 replications showed that the *Holospira* homologs were separated  
into two major clades (Figure 1B). The 89-kDa periplasmic protein from *H. obtusa* BAE80122.1 clus-  
tered with one group of *Holospira* proteins, together with homologs from *H. curviuscula*, *H. undulata*,  
and *H. elegans*. A second clade contained additional homologous proteins from these species, suggest-  
ing that multiple related proteins may be present within individual *Holospira* species. These results  
indicate that the 89-kDa periplasmic protein belongs to a conserved *Holospira*-specific protein family.  
The separation of the homologs into two clades suggests possible diversification of this protein family,  
potentially through gene duplication, although this possibility requires further genomic and

A



B

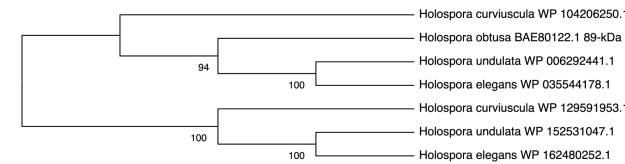


**Figure 1.** Phylogeny of the 89-kDa periplasmic invasion-tip protein from *Heliospora obtusa*. (A) Multiple amino acid sequence alignment of the 89-kDa periplasmic protein and *Heliospora* homologs. Darkly and lightly shaded text represents at least 80% identical and similar amino acid residues, respectively. NCBI accession numbers are indicated with species names. (B) The ML consensus tree obtained from bootstrap analysis with 1,000 replications of the 89-kDa periplasmic protein based on amino acid sequence alignments shown in (A). Bootstrap values > 60% are given to the left of selected nodes.

phylogenetic analysis. Notably, a corresponding type-II homolog has not yet been identified for *H. obtusa*, which may reflect incomplete sequence availability or, alternatively, true absence of this paralog in the currently available dataset. Further analysis of the conserved residues identified in the alignment, particularly in relation to the domain organization and pH-dependent structural rearrangement described below, may help to identify residues that are important for the structural behavior of BAE80122.1.

To characterize the protein in the absence of a close homologue, we applied a panel of primary-sequence-based prediction tools. Consensus secondary-structure prediction (PSIPRED, SPIDER3, PSSpred, DeepCNF and NetSurfP-2.0) divided the protein into a short N-terminal module (residues 1–197) of mixed  $\alpha/\beta$  character with low-complexity, partly disordered segments and only ~29% helix, and a long C-terminal region (residues 198–747) predicted to be almost entirely  $\alpha$ -helical (~82%

helix; Figure S1A). The N-terminal module contained a single strongly hydrophobic segment predicted as a transmembrane/type-II signal-anchor helix (residues 15–31; TMHMM and PSORT II/ALOM) rather than a cleavable signal peptide. This segment was originally annotated as a signal peptide [14]; the two assignments differ only in whether a cleavage site is present, and current predictors favor the non-cleaved form. Coiled-coil prediction assigned propensity to a single short locus near residues 405–437, on which DeepCoil2, COILS, MARCOIL and the Lupas COIL algorithm (PSORT II) agreed; only PCOILS, a known over-predictor on continuous  $\alpha$ -helical regions, predicted a broader region as ‘coiled-coil’ (Figure S1B). These predictions indicated that the C-terminal two-thirds of the protein forms an elongated  $\alpha$ -helical region with a single coiled-coil locus. However, we could not distinguish a continuous coiled-coil from an antiparallel helical-repeat, so we structurally examined the region using a three-dimensional model (below).

### 3.2. Molecular modeling and structure-based analysis of the C-terminal rod

To characterize the higher-order architecture of the 89-kDa periplasmic protein, we examined computational models of the protein. The structure of BAE80122.1 (750 residues) was predicted with AlphaFold2, and the ribbon model was colored by the per-residue confidence score (pLDDT) (Figure S2A). Most of the protein body is predicted with high confidence (pLDDT > 90, blue) and forms an elongated array of  $\alpha$ -helices. In contrast, the N-terminal region (residues 1–30) shows markedly low pLDDT scores (< 50, yellow), so the conformation of this segment in the model is not well determined. The putative transmembrane segment (residues 15–31) lies within this low-confidence region. Ionization states were then assigned with PDB2PQR/PROPKA, using the same procedure at pH 7 and at pH 4 (Figure S2B). Of the 55 residues predicted to differ between the two conditions, 39 Glu and eight Asp residues are neutral at pH 4 and negatively charged at pH 7, seven His residues carry a positive charge at pH 4 and are neutral at pH 7, and one lysine (Lys155) is protonated at pH 4 and neutral at pH 7. The net charge of the protein is therefore 55 units more positive at pH 4 than at pH 7. These residues are distributed along the entire length of the molecule. Using these assignments as input, we performed independent 10-ns MD simulations at each pH. Backbone C $\alpha$  RMSD analysis showed an initial relaxation phase during the first 5 ns in both simulations, with a modest late-stage drift in the pH 4 trajectory; the final 500 frames (5–10 ns) were therefore used for all subsequent analyses (Figure S2C).

We first used the pH 7 MD-relaxed model to resolve the architecture of the C-terminal rod. Structure-based secondary-structure assignment (DSSP/P-SEA) reproduced the two-region partition suggested by the sequence-based predictions (Figure S1), with residues 198–747 forming a predominantly  $\alpha$ -helical (~75% helix), markedly elongated rod (~270 Å; Figure S3A). Knobs-into-holes analysis with SOCKET2 detected a single antiparallel three-helix coiled-coil bundle (helices ~404–437, 448–475 and 483–510; knobs-into-holes core ~408–425) and no canonical knobs-into-holes packing elsewhere in the rod (Figure S3B), matching the single coiled-coil locus predicted from sequence (Figure S1B). The C $\alpha$ –C $\alpha$  contact map of the rod instead showed a tandem series of ~6 antiparallel helical modules (the characteristic diamond pattern; Figure S3C), indicating an antiparallel helical repeat rather than a parallel coiled-coil arrangement. Together, these analyses define the C-terminal region as an antiparallel  $\alpha$ -helical, spectrin-repeat-like solenoid that contains one localized coiled-coil.

### 3.3. Profile- and structure-based homology searches

We then asked whether this architecture has recognizable counterparts elsewhere. Profile-based searching with HHpred and structural searching with Foldseek, the latter using the pH 7 MD-relaxed model as the query, returned only weak hits, and these were confined to the C-terminal region; the N-terminal module returned no informative matches. The highest-scoring hits were to elongated  $\alpha$ -helical proteins, a trypanosomatid microtubule-associated protein (HHpred) and a dystrophin-like protein from a chigger mite (*Leptotrombidium deliense*; Foldseek), aligning over residues 198–720, the same region defined above as the spectrin-repeat-like rod (Tables S1 and S2).

### 3.4. Structural analysis of the 89-kDa periplasmic protein at pH 4 and pH 7

We next compared the equilibrated conformations obtained at the two pH values. Rigid-body domain analysis using DynDom partitioned the protein into an N-terminal core (Domain 1, residues 48–286/299–395), an N-terminal arm (Domain 2, residues 6–30) and a C-terminal rod (Domain 3, residues 287–298/396–747). Two hinge regions were identified: hinge-1 (bending residues 30–48, centered on Pro43) and hinge-2 (bending residues 286–302 and 395–396, centered on Gly291) (Figure 2A).

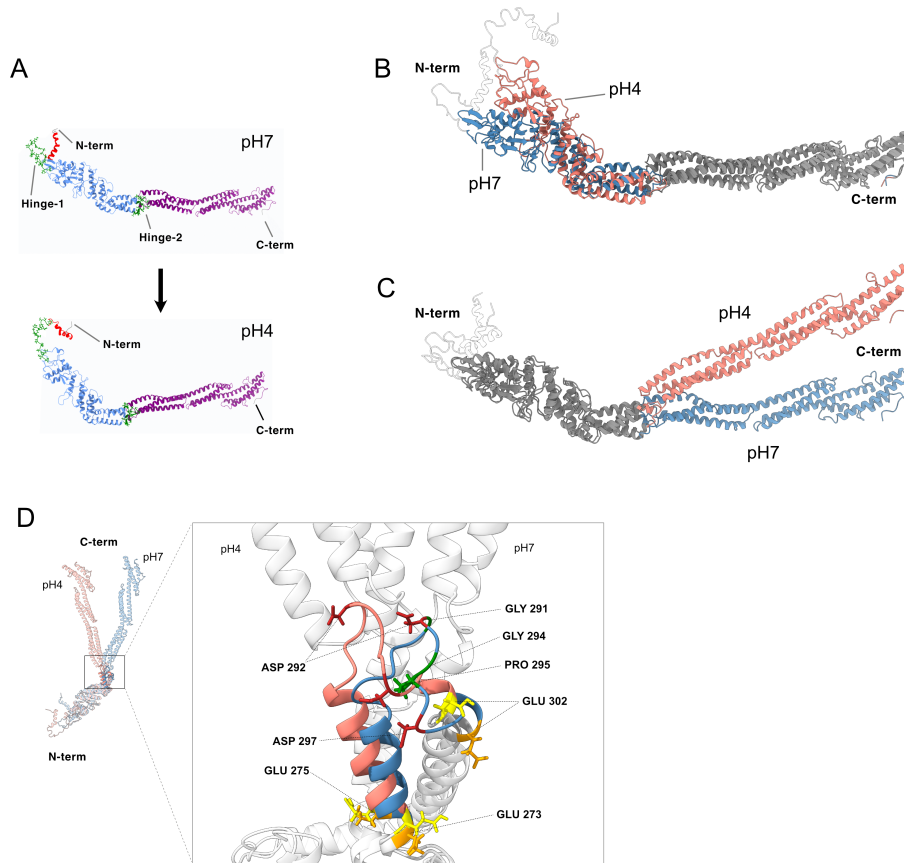


**Figure 2.** pH-dependent conformational change of the 89-kDa periplasmic invasion-tip protein. (A) Ribbon representations of the equilibrated MD structures at pH 7 (top) and pH 4 (bottom), colored according to DynDom domain assignments. Domain 1 (N-terminal core, residues 48–286 and 299–395; blue), Domain 2 (N-terminal arm, residues 6–30; red) and Domain 3 (C-terminal rod, residues 287–298 and 396–747; purple) are shown. Hinge-1 (residues 30–48) and hinge-2 (residues 286–302 and 395–396) are highlighted as green sticks. Unassigned terminal residues are shown in gray. The arrow indicates the pH 7 → pH 4 transition modeled in this study. (B, C) Superposition of the equilibrated pH 7 (blue) and pH 4 (salmon) structures on the C-terminal rod (C $\alpha$  atoms of residues 396–747; RMSD 3.4 Å over 352 atoms) (B) or on the N-terminal core (C $\alpha$  atoms of residues 48–286/299–395; RMSD 3.1 Å over 336 atoms) (C). In each panel, the region used as the superposition reference is shown in gray. Residues 1–47, which fall within the low-confidence N-terminal region (Figure S2A), are shown as semi-transparent and are not interpreted here. (D) Close-up of the hinge-2 region, comparing the pH 7 (blue) and pH 4 (salmon) conformations superposed on the N-terminal core (DynDom Domain 1, residues 48–286/299–395). Left (inset): full-length superposition with the N- and C-termini indicated; the boxed area is enlarged at right. Right: residues 273–302 are colored by state (pH 7, blue; pH 4, salmon) and the remainder is light gray. Side chains are shown for the Gly/Pro hinge residues (Gly291, Gly294, Pro295; green), the flanking glutamates Glu273, Glu275 and Glu302 (charged at pH 7, orange; protonated at pH 4, yellow), and Asp292 and Asp297 (firebrick). Protonation states are PROPKA/PDB2PQR assignments used as simulation inputs.

To characterize this rearrangement further, we superposed the two structures on each of the two large domains in turn. Superposition on the C-terminal rod gave a C $\alpha$  RMSD of 3.4 Å over 352 atoms (Figure 2B), and superposition on the N-terminal core gave 3.1 Å over 336 atoms (Figure 2C). In each case the domain used as the reference was closely superimposed, while the remainder of the molecule was displaced. The corresponding per-residue displacement profiles are approximately mirror images of one another, the displacement progressively increasing with distance from the domain junction in both reference frames (Figure S4A). Both domains therefore retain essentially the same internal structure at the two pH values, and the difference between the two states arises predominantly from a change in their relative orientation about the hinge-2 region that connects them, which DynDom quantified as a rigid-body rotation of approximately 35° (Figure 2A). The position of the N-terminal arm is poorly determined in the model (Figure S2A), so its apparent displacement is not interpreted here. Although the precise trajectory of the N-terminal arm cannot be inferred from the present model, exposure of the N-terminal region beyond the bacterial outer membrane during infection has been demonstrated experimentally [14]. A morph animation interpolating between the two models is provided as Video S1; the interpolation is geometric and does not represent a simulated transition pathway. Per-residue RMSF (Figure S4B) showed that hinge-1 and the N-terminal arm substantially fluctuate at both pH values, which is consistent with the low prediction confidence in this region. Both the hinge-2 loop (residues 286–302) and the core–rod junction and proximal rod (residues ~390–520) showed lower RMSF at pH 4 than at pH 7, with no consistent trend further along the rod (residues >520). Rather than reflecting increased disorder, this suggests that the pH 4 conformation is comparatively well ordered at the hinge-2 loop and along the proximal rod.

We next examined the hinge-2 region in structural detail (Figure 2D). The divergent segment (residues 273–302) contains a glycine/proline-rich loop (Gly291, Gly294, Pro295), flanked by three glutamates (Glu273, Glu275, Glu302) that are predicted to be neutral at pH 4, whereas two aspartates within the same loop (Asp292, Asp297) remain charged at both pH values. DynDom located hinge-2 from the two sets of coordinates alone, without reference to sequence. This geometrically defined hinge coinciding with a glycine/proline-rich loop means that two independent criteria (backbone geometry and amino-acid composition) identify the same segment as the flexible connection between the N-terminal core and the C-terminal rod. Within the loop itself, the conformational change is small: superposed on residues 286–302 alone, the pH 7 and pH 4 models differ by 1.8 Å C $\alpha$  RMSD, and the crossing angle between the helices flanking the loop changes by ~5°. None of the three flanking glutamates forms a salt bridge with a basic residue at either pH.

The connection is not a single joint. DynDom assigns residues 287–298 to the C-terminal rod (Domain 3) even though they lie within the core in sequence, and in the model this segment packs directly against the rod (closest heavy-atom distance 2.8 Å at pH 7 and 2.7 Å at pH 4). The core and the rod are therefore joined at more than one point: the chain passes between them at 286/287, 298/299 and 395/396, rather than through a single linker. The sequence-level features, structural homology, architecture and DynDom domain assignments described above are summarized on a common residue axis in Figure 3.

## 4. Discussion

Sequence-based homology searches identified confident orthologues only within the genus *Holospira*, with no hits to proteins from other taxa (see Results). This suggests that the 89-kDa protein has either diverged extensively or evolved largely without recognizable homologues outside this genus. This lack of similarity was not uniform along the sequence: weak but detectable hits were recovered for the C-terminal region, whereas the N-terminal module returned no informative matches (Tables S1 and S2). If a bona fide N-terminal homologue is eventually identified in other organisms, it may help to clarify the evolutionary origin of this module. Conversely, if it genuinely remains absent from other lineages, this would be consistent with either an origin unique to *Holospira* or, more speculatively,

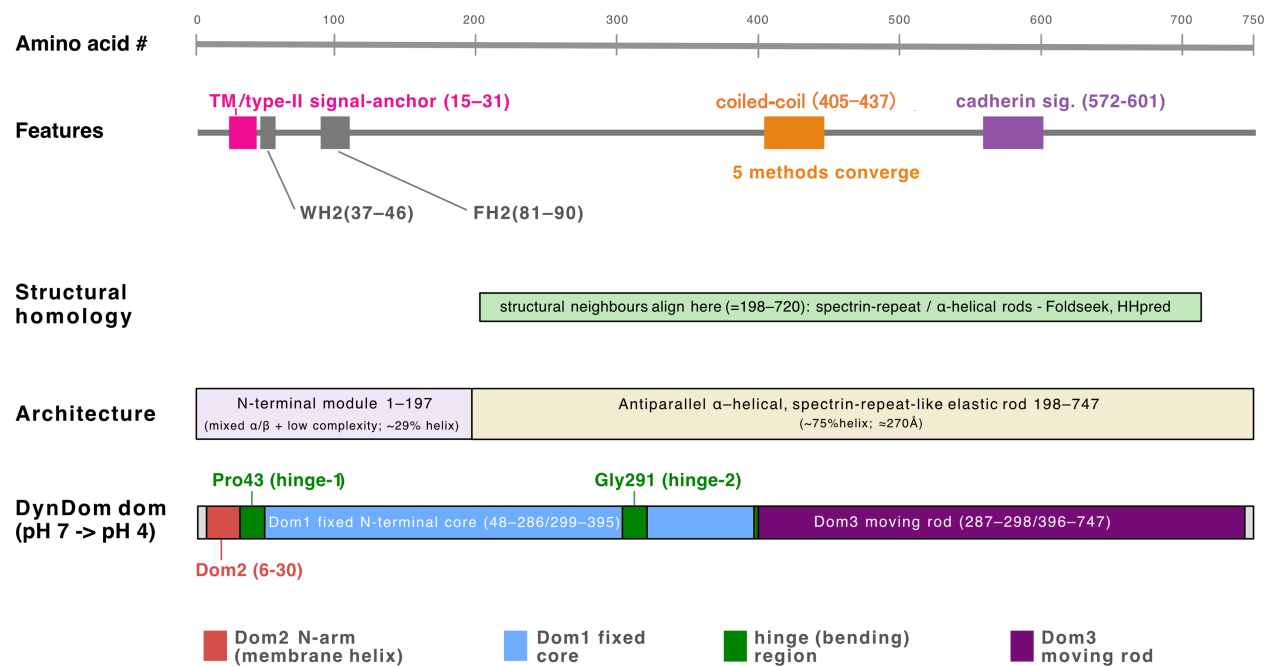


**Figure 3.** Integrated domain/feature map of the 89-kDa protein, summarizing the homology, primary-to-tertiary structure, and pH-dependent conformational analyses. Tracks are ordered top to bottom from earliest to most recent evidence. Features: sequence-level motifs, including a transmembrane/type-II signal-anchor helix (15–31; TMHMM and PSORT II/ALOM), the two putative WH2-like and FH2-like actin-binding motifs (37–46 and 81–90, respectively) and a cadherin signature (572–601) originally proposed from sequence [14], and a single coiled-coil (405–437) on which five independent methods converge (Figure S1B). Structural homology: the residue range (≈198–720) over which structural neighbors (spectrin-repeat /  $\alpha$ -helical rods) align in Foldseek and HHpred searches. Architecture: secondary-structure assignment partitions the protein into an N-terminal module (1–197; mixed  $\alpha/\beta$  with low-complexity segments, ~29% helix) and a C-terminal antiparallel  $\alpha$ -helical, spectrin-repeat-like elastic rod (198–747; ~75% helix, ≈270 Å) (Figures S1A and S3A). DynDom domains: rigid-body decomposition of the pH 7  $\rightarrow$  pH 4 conformational change into a fixed N-terminal core (Domain 1, residues 48–286/299–395; blue), a mobile N-terminal arm carrying the membrane helix (Domain 2, residues 6–30; red) and a mobile C-terminal rod (Domain 3, residues 287–298/396–747; purple; rotation 35°), hinged at Pro43 (hinge-1) and Gly291 (hinge-2; green). Dom1–Dom3, Domains 1–3.

an evolutionary history in which the N- and C-terminal modules were assembled from sequences of separate origin.

The two C-terminal hits noted above are individually weak, and they should be treated with caution. However, both point toward proteins that act as cytoskeletal or membrane-cortex scaffolds through interaction with partner proteins. Dystrophin in particular anchors the actin cytoskeleton to the plasma membrane via a spectrin-repeat rod. Combined with our finding that the C-terminal region of the 89-kDa protein is itself an antiparallel, spectrin-repeat-like solenoid (Figure 3 and Figure S3), this raises the possibility that the protein does not act alone, but rather assembles with other, as yet unidentified proteins, into a structural complex beneath the *H. obtusa* outer membrane. A further observation bears on this possibility. In addition to the 89-kDa protein, a 76-kDa protein has been detected in micro-dissected invasion-tips of the *H. obtusa* IF cell, and immunofluorescence indicates that it is confined to the invasion-tip (M. Fujishima, unpublished data); its gene has not yet been identified. We have no direct evidence of a physical interaction between the two proteins, but the presence of a second, invasion-tip-specific protein in the same compartment would be consistent with the 89-kDa protein functioning within a complex. If such a complex exists, it could contribute to retaining the C-

terminal rod within the periplasmic space while the N-terminal segment becomes exposed at the bacterial surface.

The MD and DynDom analyses indicate that the difference between the pH 7 and pH 4 models is a rigid-body reorientation of the N-terminal core and the C-terminal rod about hinge-2, rather than a deformation of either domain (Figures 2B and 2C). Two features of this arrangement are worth noting. First, the hinge identified from backbone geometry alone coincides with a glycine/proline-rich loop, so that geometry and sequence composition independently point to the same segment as the flexible connection between the two domains (Figure 2D). Second, the core and the rod are joined at two points rather than one: DynDom assigns residues 287–298 to the rod domain although they lie within the core in sequence, and this segment packs directly against the rod in the model. The rod is ~270 Å long, so a small change at this junction is expressed as a displacement of up to ~80 Å at its distal end (Figure S4A).

We were unable to identify the interaction through which protonation acts at this junction. None of the three glutamates flanking the hinge-2 loop forms a salt bridge at either pH, and the local rearrangement within the loop is small (see Results). The pH dependence of the reorientation may therefore reflect the distributed change in charge across the molecule as a whole (55 residues; Figure S2B) rather than the rupture or formation of a specific ion pair. Distinguishing between these possibilities will require experiment.

The biological significance of this predicted conformational change lies in its potential connection to the earliest experimentally defined step of *H. obtusa* infection. Previous cell-biological studies have shown that acidification of the host DV is required for bacterial escape [20] and that, during this period, the N-terminal region of the 89-kDa invasion-tip protein becomes exposed beyond the bacterial outer membrane [14]. The pH-dependent reorientation predicted here therefore provides a possible structural link between an environmental cue generated by the host DV and activation of the bacterial invasion apparatus (Figure 4). pH-triggered conformational switches are well established in other systems, notably in viral fusion glycoproteins, in which acidification promotes large-scale structural reorganization [47]. Although the 89-kDa protein is unrelated to these proteins in sequence and fold, this precedent illustrates how acidification can gate a large-scale conformational reorientation. The pH modeled here corresponds to the transient acidification of the DV-I stage, from which escape occurs, and to the condition (pH 4 for 15 min) shown experimentally to induce transformation of the infectious form into the activated form [19]; the more acidic DV-II stage (pH < 3) blocks escape [20].

Per-residue RMSF analysis showed that both the hinge-2 loop and the core–rod junction are more ordered at pH 4 (Figure S4B), suggesting that the pH 4 conformation may represent a discrete state rather than a transiently strained intermediate. Host actin filaments associate with the exposed N-terminal segment following vacuolar escape [16,21]. Actin engagement might therefore restrain the N-terminal arm after entry into the near-neutral host cytoplasm, making an intrinsically reversible conformational switch functionally unidirectional in vivo. This possibility remains speculative, particularly because the position of the N-terminal arm is poorly determined in the present model (Figure S2A).

Several limitations should be noted in relation to the current study. The model rests entirely on computational prediction and simulation. The three-dimensional structure was generated by AlphaFold2 without an experimental template, and protonation states were assigned computationally (PDB2PQR/PROPKA) rather than measured. The residues identified as titratable and their assigned pKa values may not correspond exactly to their behavior in the native protein, particularly in atypical local environments. Each pH condition was sampled from a single 10-ns trajectory, so we cannot distinguish the differences reported here from the variation that would be expected between independent simulations of the same system. Replicate simulations and longer sampling would thus be required to establish that the reorientation is reproducible. The two states were compared as static

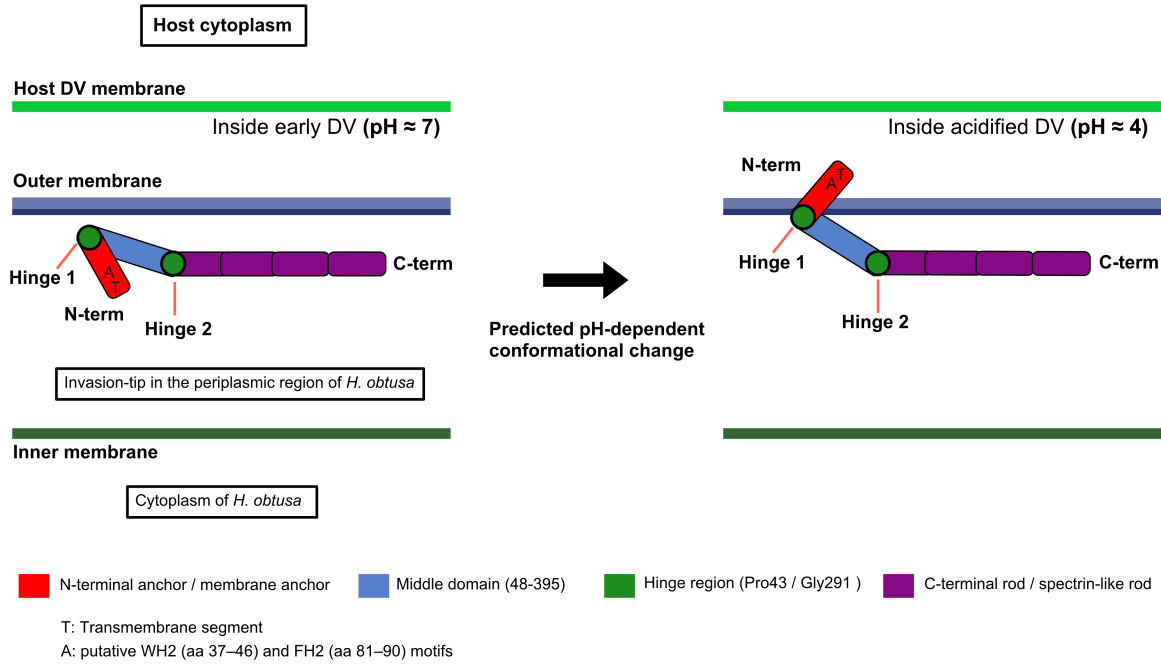


**Figure 4.** Proposed model of pH-dependent conformational change in the 89-kDa periplasmic invasion-tip protein during escape of the infectious form of *H. obtusa* from the host digestive vacuole (DV). Schematic cross-sections show, from top to bottom, the host cytoplasm, the host DV membrane, the outer and inner membranes of *H. obtusa*, and the intervening periplasmic space, where the invasion-tip protein resides. Left: inside the early, near-neutral DV (pH ≈ 7), the protein adopts an overall extended conformation: the N-terminal membrane anchor (red; T, transmembrane segment; A, putative WH2/FH2-like actin-binding motifs, residues 37–46 and 81–90) lies close to the outer membrane, while the middle domain (blue, residues 48–395) and the C-terminal spectrin-like rod (purple) extend away within the periplasmic space, held in this configuration by hinge-1 (Pro43) and hinge-2 (Gly291; green). Right: upon acidification of the DV (pH ≈ 4), the molecule adopts a more compact, curved overall conformation. In the model, this is accompanied by a reorientation of the N-terminal anchor about hinge-1. The conformation of this segment is poorly determined (Figure S2A), so the schematic shows its repositioning qualitatively rather than to scale.

endpoints rather than through free-energy or pathway analysis, and the observed difference was not shown to be caused by protonation. The results should therefore be regarded as a structural prediction and a testable hypothesis rather than a demonstrated mechanism. Direct experimental structure determination by X-ray crystallography or cryo-EM at both pH values will be required to test the conformational states proposed here. Finally, subsequent passage of the protein across the host macronuclear envelope appears, from available evidence, to be a separate process that does not obviously depend on the change described here, and is not addressed by the present model.

Experimental testing of this model is constrained by the biology of the organism: *H. obtusa* is an obligate intracellular symbiont that cannot be cultured outside *Paramecium*, and no system for its genetic manipulation is currently available, so direct mutational analysis in the bacterium is not presently feasible. A more accessible route would be to examine the recombinant protein, or fragments of it, in vitro. Whether its overall shape changes between pH 7 and pH 4 could be tested by circular dichroism, size-exclusion chromatography or hydrogen-deuterium exchange mass spectrometry. The contribution of the hinge-2 residues (Glu273, Glu275, Glu302; Gly291, Gly294, Pro295) could be assessed in the same assays using site-directed mutants.

## 5. Conclusions

The 89-kDa invasion-tip protein of *H. obtusa* is predicted to consist of an N-terminal membrane-anchoring module and a long C-terminal antiparallel, spectrin-repeat-like helical solenoid containing a localized coiled-coil. Molecular dynamics simulations at pH 7 and pH 4 revealed a difference in the relative orientation of the N-terminal core and C-terminal rod, described as an approximately 35° rigid-body rotation about a hinge centered on Gly291. Previous cell-biological experiments showed that, after *H. obtusa* IF cell is taken up into a host DV, the N-terminal region of the 89-kDa protein becomes exposed beyond the bacterial outer membrane during vacuolar acidification [14]. The pH-dependent conformational difference predicted in the present study is consistent with this experimentally observed N-terminal exposure and provides a possible structural basis for this early step in the infection process. Although the present computational analyses do not establish that protonation causes this conformational difference, they provide a testable structural hypothesis linking acidification of the host DV to activation of the bacterial invasion apparatus. Experimental structural analyses will be required to determine whether this predicted pH-dependent reorientation occurs during *H. obtusa* infection.

**Supplementary Materials:** **Figure S1:** Sequence-based secondary-structure, coiled-coil and transmembrane prediction. **Figure S2:** AlphaFold2 model, protonation state and molecular-dynamics setup. **Figure S3:** Structure-based analysis of the C-terminal rod on the molecular-dynamics model. **Figure S4:** Displacement and fluctuation analysis of the pH 7 and pH 4 MD models. **Table S1:** HHpred search of the query BAE80122.1 (89-kDa periplasmic protein, *Holospira obtusa*) against the PDB (job 6845740). **Table S2:** Foldseek 3Di/AA structural search (no iterative search) of the MD-refined model of the query (BAE80122.1, 89-kDa periplasmic protein, *Holospira obtusa*; 750 aa) against AFDB UniProt50 and MGNIFY\_ESM30, with no taxonomic filter. **File S1:** Equilibrated structure of the 89-kDa protein at pH 7 (final frame of the 10-ns MD trajectory; PDB format). **File S2:** Equilibrated structure of the 89-kDa protein at pH 4 (final frame of the 10-ns MD trajectory; PDB format). **Video S1:** Morph animation of the predicted pH-dependent conformational change of the 89-kDa periplasmic invasion-tip protein (pH 7 → pH 4), generated with ChimeraX. Coloring follows Figure 2A: N-terminal core (Domain 1, blue), N-terminal arm (Domain 2, red), C-terminal rod (Domain 3, purple), hinge regions (green). The interpolation is geometric and does not represent a simulated transition pathway.

**Author Contributions:** Conceptualization, M.F. and Y.T.; methodology, Y.T. and M.F.; software, Y.T.; validation, T.M., T.N. and Y.S.; investigation, Y.T.; resources, M.F.; data curation, Y.T.; writing—original draft preparation, M.F. and Y.T.; writing—review and editing, M.F., Y.T., T.M., T.N. and Y.S.; visualization, M.F. and Y.T.; supervision, M.F.; project administration, M.F. and Y.T. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The equilibrated pH 7 and pH 4 structures analyzed in this study are provided as Supplementary Materials. The full molecular-dynamics trajectories will be made available by the authors on request.

**Acknowledgments:** We acknowledge proofreading and editing by Benjamin Phillis, a Board-Certified Editor in the Life Sciences (BELS), at Wakayama Medical University.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

|    |                   |
|----|-------------------|
| RF | Reproductive form |
| IF | Infectious form   |
| DV | Digestive vacuole |
| AF | Activated form    |

## References

1. Abamo, F.; Dohra, H.; Fujishima, M. Fate of the 63-kDa periplasmic protein of the infectious form of the endonuclear symbiotic bacterium *Holospora obtusa* during the infection process. *FEMS Microbiol. Lett.* **2008**, *280*, 21–27. <https://doi.org/10.1111/j.1574-6968.2007.01023.x>
2. Lang, B.F.; Brinkmann, H.; Koski, L.B.; Fujishima, M.; Görtz, H.-D.; Burger, G. On the origin of mitochondria and Rickettsia-related eukaryotic endosymbionts. *Jpn. J. Protozool.* **2005**, *38*, 171–183. [https://doi.org/10.18980/jjprotozool.38.2\\_171](https://doi.org/10.18980/jjprotozool.38.2_171)
3. Garushyants, S.K.; Beliavskaia, A.Y.; Malko, D.B.; Logacheva, M.D.; Rautian, M.S.; Gelfand, M.S. Comparative genomic analysis of *Holospora* spp., intranuclear symbionts of paramecia. *Front. Microbiol.* **2018**, *9*, 738. <https://doi.org/10.3389/fmicb.2018.00738>
4. Fokin, S.I.; Lebedeva, N.A.; Potekhin, A.; Gammuto, L.; Petroni, G.; Serra, V. *Holospora*-like bacteria "Candidatus Gortzia yakutica" and *Preeria caryophila*: Ultrastructure, promiscuity, and biogeography of the symbionts. *Eur. J. Protistol.* **2023**, *90*, 125998. <https://doi.org/10.1016/j.ejop.2023.125998>
5. Giovannini, M.; Petroni, G.; Castelli, M. Novel evolutionary insights on the interactions of the *Holosporales* (*Alphaproteobacteria*) with eukaryotic hosts from comparative genomics. *Environ. Microbiol.* **2024**, *26*, e16562. <https://doi.org/10.1111/1462-2920.16562>
6. Castelli, M.; Petroni, G. An evolutionary-focused review of the *Holosporales* (*Alphaproteobacteria*): Diversity, host interactions, and taxonomic re-ranking as *Holosporineae* Subord. *Nov. Microb. Ecol.* **2025**, *88*, 15. <https://doi.org/10.1007/s00248-025-02509-0>
7. Dohra, H.; Tanaka, K.; Suzuki, T.; Fujishima, M.; Suzuki, H. Draft genome sequences of three *Holospora* species (*Holospora obtusa*, *Holospora undulata*, and *Holospora elegans*), endonuclear symbiotic bacteria of the ciliate *Paramecium caudatum*. *FEMS Microbiol. Lett.* **2014**, *359*, 16–18. <https://doi.org/10.1111/1574-6968.12577>
8. Görtz, H.-D. Nucleus-specific symbionts in *Paramecium caudatum*. In *Endosymbiosis and Cell Biology*; Schwemmler, W., Schenk, H.E.A., Eds.; Walter de Gruyter & Co.: Berlin, Germany, 1980; 1, pp. 381–392.
9. Gromov, B.V.; Ossipov, D.V. *Holospora* (ex Hafkine 1890) nom. rev., a genus of bacteria inhabiting the nuclei of paramecia. *Int. J. Syst. Evol. Microbiol.* **1981**, *31*, 348–352. <https://doi.org/10.1099/00207713-31-3-348>
10. Fujishima, M.; Nagahara, K.; Kojima, Y. Changes in morphology, buoyant density and protein composition in differentiation from the reproductive short form to the infectious long form of *Holospora obtusa*, a macronucleus-specific symbiont of the ciliate *Paramecium caudatum*. *Zool. Sci.* **1990**, *7*, 849–860. <https://iss.ndl.go.jp/books/R000000004-I3695221-00>
11. Fujishima, M.; Hoshide, K. Light and electron microscopic observations of *Holospora obtusa*: a macronucleus-specific bacterium of the ciliate *Paramecium caudatum*. *Zool. Sci.* **1988**, *5*, 791–799.
12. Görtz, H.-D.; Wiemann, M. Route of infection of the bacteria *Holospora elegans* and *Holospora obtusa* into the nuclei of *Paramecium caudatum*. *Eur. J. Protistol.* **1989**, *24*, 101–109. [https://doi.org/10.1016/S0932-4739\(89\)80037-9](https://doi.org/10.1016/S0932-4739(89)80037-9)
13. Dohra, H.; Fujishima, M. Cell structure of the infectious form of *Holospora*, an endonuclear symbiotic bacterium of the ciliate *Paramecium*. *Zool. Sci.* **1999**, *16*, 93–98. <https://doi.org/10.2108/zsj.16.93>
14. Iwatani, K.; Dohra, H.; Lang, B.F.; Burger, G.; Hori, M.; Fujishima, M. Translocation of an 89-kDa periplasmic protein is associated with *Holospora* infection. *Biochem. Biophys. Res. Commun.* **2005**, *337*, 1198–1205. <https://doi.org/10.1016/j.bbrc.2005.09.175>

15. Fujishima, M.; Kawano, H.; Miyakawa, I. A 63-kDa periplasmic protein of the endonuclear symbiotic bacterium *Holospora obtusa* secreted to the outside of the bacterium during the early infection process binds weakly to the macronuclear DNA of the host *Paramecium caudatum*. *Microorganisms* **2023**, *11*, 155. <https://doi.org/10.3390/microorganisms11010155>
16. Fujishima, M. Infection and maintenance of *Holospora* species in *Paramecium caudatum*. In *Endosymbionts in Paramecium*. Microbiology Monographs; Fujishima, M.; Ed.; Springer: Berlin/Heidelberg, Germany, 2009; Volume 12, pp. 201–225. [https://doi.org/10.1007/978-3-540-92677-1\\_8](https://doi.org/10.1007/978-3-540-92677-1_8)
17. Fok, A.K.; Lee, Y.; Allen, R.D. The correlation of digestive vacuole pH and size with the digestive cycle in *Paramecium caudatum*. *J. Protozool.* **1982**, *29*, 409–414. <https://doi.org/10.1111/j.1550-7408.1982.tb05423.x>
18. Kodama, Y.; Fujishima, M. Symbiotic *Chlorella* sp. of the ciliate *Paramecium bursaria* do not prevent acidification and lysosomal fusion of host digestive vacuoles during infection. *Protoplasma* **2005**, *225*, 191–203. <https://doi.org/10.1007/s00709-005-0087-5>
19. Kawai, M.; Fujishima, M. Monoclonal antibody specific for activated form of *Holospora obtusa*, a macronucleus-specific bacterium of *Paramecium caudatum*. In *Eukaryotism and Symbiosis*. Schenk, H.E.A.; Herrmann, R.G.; Jeon, K.W.; Müller, N.E.; Schwemmler, W.; Eds.; Springer: Berlin, Heidelberg, Germany, 1997, pp. 371–374. [https://doi.org/10.1007/978-3-642-60885-8\\_29](https://doi.org/10.1007/978-3-642-60885-8_29)
20. Fujishima, M.; Kawai, M. Acidification in digestive vacuoles is an early event required for *Holospora* infection of *Paramecium* nucleus. In *Eukaryotism and Symbiosis*. Schenk, H.E.A.; Herrmann, R.G.; Jeon, K.W.; Müller, N.E.; Schwemmler, W.; Eds.; Springer: Berlin, Heidelberg, Germany, 1997, pp. 367–370. [https://doi.org/10.1007/978-3-642-60885-8\\_28](https://doi.org/10.1007/978-3-642-60885-8_28)
21. Sabaneyeva, E.V.; Derkacheva, M.E.; Benken, K.A.; Fokin, S.I.; Vainio, S.; Skovorodkin, I.N. Actin-based mechanism of *Holospora obtusa* trafficking in *Paramecium caudatum*. *Protist* **2009**, *160*, 205–219. <https://doi.org/10.1016/j.protis.2008.11.006>
22. Stecher, G.; Suleski, M.; Tao, Q.; Tamura, K.; Kumar, S. MEGA 12.1: Cross-platform release for macOS and Linux operating systems. *J. Mol. Evol.* **2026**, *94*, 14–18. <https://doi.org/10.1007/s00239-025-10287-z>
23. Mirdita, M.; Schütze, K.; Moriwaki, Y.; Heo, L.; Ovchinnikov, S.; Steinegger, M. ColabFold: making protein folding accessible to all. *Nature Methods* **2022**, *19*, 679–682. <https://doi.org/10.1038/s41592-022-01488-1>
24. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Zidek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S.A.A.; Ballard, A.J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A.W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. <https://doi.org/10.1038/s41586-021-03819-2>
25. Dolinsky, T.J.; Czodrowski, P.; Li, H.; Nielsen, J.E.; Jensen, J.H.; Klebe, G.; Baker, N.A. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res.* **2007**, *35*, W522–525. <https://doi.org/10.1093/nar/gkm276>
26. Olsson, M.H.M.; Søndergaard, C.R.; Rostkowski, M.; Jensen, J.H. PROPKA3: Consistent treatment of internal and surface residues in empirical pKa predictions. *J. Chem. Theory Comput.* **2011**, *7*, 525–537. <https://doi.org/10.1021/ct100578z>
27. Maier, J.A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K.E.; Simmerling, C. ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713. <https://doi.org/10.1021/acs.jctc.5b00255>
28. Jorgensen, W.L.; Chandrasekhar, J.; Madura, J.D.; Impey, R.W.; Klein, M.L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935. <https://doi.org/10.1063/1.445869>
29. Eastman, P.; Galvelis, R.; Pelaez, R.P.; Abreu, C.R.A.; Farr, S.E.; Gallicchio, E.; Gorenko, A.; Henry, M.M.; Hu, F.; Huang, J.; Kramer, A.; Michel, J.; Mitchell, J.A.; Pande, V.S.; Rodrigues, J.P.; Rodriguez-Guerra, J.; Simmonett, A.C.; Singh, S.; Swails, J.; Turner, P.; Wang, Y.; Zhang, I.; Chodera, J.D.; De Fabritiis, G.; Markland, T.E. OpenMM 8: Molecular dynamics simulation with machine learning potentials. *J. Phys. Chem. B* **2024**, *128*, 109–116. <https://doi.org/10.1021/acs.jpcc.3c06662>

30. Essmann, U.; Perera, L.; Berkowitz, M.L.; Darden, T.; Lee, H.; Pedersen, L.G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577–8593. <https://doi.org/10.1063/1.470117>
31. Gabler, F.; Nam, S.Z.; Till, S.; Mirdita, M.; Steinegger, M.; Soding, J.; Lupas, A.N.; Alva, V. Protein sequence analysis using the MPI Bioinformatics toolkit. *Curr. Protoc. Bioinform.* **2020**, *72*, e108. <https://doi.org/10.1002/cpbi.108>
32. Jones, D.T. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **1999**, *292*, 195–202. <https://doi.org/10.1006/jmbi.1999.3091>
33. Heffernan, R.; Yang, Y.; Paliwal, K.; Zhou, Y. Capturing non-local interactions by long short-term memory bidirectional recurrent neural networks for improving prediction of protein secondary structure, backbone angles, contact numbers and solvent accessibility. *Bioinformatics* **2017**, *33*, 2842–2849. <https://doi.org/10.1093/bioinformatics/btx218>
34. Wang, S.; Peng, J.; Ma, J.; Xu, J. Protein secondary structure prediction using deep convolutional neural fields. *Sci. Rep.* **2016**, *6*, 18962. <https://doi.org/10.1038/srep18962>
35. Klausen, M.S.; Jespersen, M.C.; Nielsen, H.; Jensen, K.K.; Jurtz, V.I.; Sonderby, C.K.; Sommer, M.O.A.; Winther, O.; Nielsen, M.; Petersen, B.; Marcatili, P. NetSurfP-2.0: Improved prediction of protein structural features by integrated deep learning. *Proteins* **2019**, *87*, 520–527. <https://doi.org/10.1002/prot.25674>
36. Lupas, A.; Van Dyke, M.; Stock, J. Predicting coiled coils from protein sequences. *Science* **1991**, *252*, 1162–1164. <https://doi.org/10.1126/science.252.5009.1162>
37. Gruber, M.; Soding, J.; Lupas, A.N. Comparative analysis of coiled-coil prediction methods. *J. Struct. Biol.* **2006**, *155*, 140–145. <https://doi.org/10.1016/j.jsb.2006.03.009>
38. Delorenzi, M.; Speed, T. An HMM model for coiled-coil domains and a comparison with PSSM-based predictions. *Bioinformatics* **2002**, *18*, 617–625. <https://doi.org/10.1093/bioinformatics/18.4.617>
39. Krogh, A.; Larsson, B.; von Heijne, G.; Sonnhammer, E.L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **2001**, *305*, 567–580. <https://doi.org/10.1006/jmbi.2000.4315>
40. Jones, D.T.; Cozzetto, D. DISOPRED3: precise disordered region predictions with annotated protein-binding activity. *Bioinformatics* **2015**, *31*, 857–863. <https://doi.org/10.1093/bioinformatics/btu744>
41. Ludwiczak, J.; Winski, A.; Szczepaniak, K.; Alva, V.; Dunin-Horkawicz, S. DeepCoil-a fast and accurate prediction of coiled-coil domains in protein sequences. *Bioinformatics* **2019**, *35*, 2790–2795. <https://doi.org/10.1093/bioinformatics/bty1062>
42. Nakai, K.; Horton, P. PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends Biochem. Sci.* **1999**, *24*, 34–36. [https://doi.org/10.1016/s0968-0004\(98\)01336-x](https://doi.org/10.1016/s0968-0004(98)01336-x)
43. Labesse, G.; Colloc'h, N.; Pothier, J.; Mornon, J.P. P-SEA: a new efficient assignment of secondary structure from C alpha trace of proteins. *Comput. Appl. Biosci.* **1997**, *13*, 291–295. <https://doi.org/10.1093/bioinformatics/13.3.291>
44. Kabsch, W.; Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22*, 2577–2637. <https://doi.org/10.1002/bip.360221211>
45. Kumar, P.; Woolfson, D.N. Socket2: a program for locating, visualizing and analyzing coiled-coil interfaces in protein structures. *Bioinformatics* **2021**, *37*, 4575–4577. <https://doi.org/10.1093/bioinformatics/btab631>
46. Hayward, S.; Berendsen, H.J.C. Systematic analysis of domain motions in proteins from conformational change: New results on citrate synthase and T4 lysozyme. *Proteins* **1998**, *30*, 144–154. [https://doi.org/10.1002/\(SICI\)1097-0134\(19980201\)30:2<144::AID-PROT4>3.0.CO;2-N](https://doi.org/10.1002/(SICI)1097-0134(19980201)30:2<144::AID-PROT4>3.0.CO;2-N)
47. Bullough, P.A.; Hughson, F.M.; Skehel, J.J.; Wiley, D.C. Structure of influenza haemagglutinin at the pH of membrane fusion. *Nature* **1994**, *371*, 37–43. <https://doi.org/10.1038/371037a0>