Biochemistry and Biophysics Reports 39 (2024) 101797

Contents lists available at ScienceDirect

Biochemistry and Biophysics Reports

journal homepage: www.elsevier.com/locate/bbrep

Deciphering the enigmatic PilY1 of *Acidithiobacillus thiooxidans*: An *in silico* analysis

Araceli Hernández-Sánchez [1], Edgar D. Páez-Pérez [*,1], Elvia Alfaro-Saldaña, J. Viridiana García-Meza [**]

Geomicrobiología, Metalurgia, UASLP, Sierra Leona 550, San Luis Potosí, 78210, SLP, Mexico

ARTICLE INFO

Keywords:
Acidophile
PilY1 interaction
vWA domain
Protein structure
Protein disorder
Circular dichroism spectra

ABSTRACT

Thirty years since the first report on the PilY1 protein in bacteria, only the C-terminal domain has been crystallized; there is no study in which the N-terminal domain, let alone the complete protein, has been crystallized. In our laboratory, we are interested in characterizing the Type IV Pili (T4P) of *Acidithiobacillus thiooxidans*. We performed an *in silico* characterization of PilY1 and other pilins of the T4P of this acidophilic bacterium. *In silico* characterization is crucial for understanding how proteins adapt and function under extreme conditions. By analyzing the primary and secondary structures of proteins through computational methods, researchers can gain valuable insights into protein stability, key structural features, and unique amino acid compositions that contribute to resilience in harsh environments. Here, it is presented a description of the particularities of *At. thiooxidans* PilY1 through predictor software and homology data. Our results suggest that PilY1 from *At. thiooxidans* may have the same role as has been described for other PilY1 associated with T4P in neutrophilic bacteria; also, its C-terminal interacts (interface interaction) with the minor pilins PilX, PilW and PilV. The N-terminal region comprises domains such as the vWA and the MIDAS, involved in signaling, ligand-binding, and protein-protein interaction. In fact, the vWA domain has intrinsically disordered regions that enable it to maintain its structure over a wide pH range, not only at extreme acidity to which *At. thiooxidans* is adapted. The results obtained helped us design the correct methodology for its heterologous expression. This allowed us partially experimentally characterize it by obtaining the N-terminal domain recombinantly and evaluating its acid stability through fluorescence spectroscopy. The data suggest that it remains stable across pH changes. This work thus provides guidance for the characterization of extracellular proteins from extremophilic organisms.

1. Introduction

As other extremophiles, *Acidithiobacillus thiooxidans* has a significant role in global biotechnologies as the acidophilic bioleaching process [1,2,3] and mining bioremediation [4]. The conditions and adaptive mechanisms which the *Acidithiobacillus* genus has developed to maintain the integrity of its cellular functions (pH < 3, high redox potentials and high concentrations of heavy metals) provide a good model for studying the resilience of extracellular proteins. For bioleaching, cell adherence and biofilm formation are key factors. Biofilm formation involves surface sensing and the production of extracellular polymeric substances (EPS) [5], exchange of flagellar and pili motility, and multiple cellular signaling cascades and gene regulation [6]. Previous studies have shown that Type IV Pili (T4P) are involved in biofilm formation in *Acidithiobacillus ferrooxidans* and *At. thiooxidans* [7]. Functional and biochemical assays as well as tertiary structure analyses of several pilins of T4P of various prokaryotic pathogens have been carried out, but not for *Acidithiobacillus* sp.

Based on information obtained from bacteria of clinical or biotechnological interest, more than a dozen different key proteins forming part

* Corresponding author. Geomicrobiología, Metalurgia, UASLP, Sierra Leona 550, San Luis Potosí, 78210, SLP, Mexico.
** Corresponding author.
E-mail addresses: edpaezperez@gmail.com (E.D. Páez-Pérez), jvgm@uaslp.mx, jviridianagm@gmail.com (J.V. García-Meza).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.bbrep.2024.101797
Received 30 May 2024; Received in revised form 18 July 2024; Accepted 19 July 2024

of the T4P structure and system are currently known, e.g. proteins called major pilins (PilA forming the stem of the pilum) and minor pilins on the pili tip (FimU, PilV, PilW, PilX, PilY1/PilC). In *At. thiooxidans*, a cluster with a structure similar to the canonical operon found in *Pseudomonas aeruginosa*, which primes the T4P, has been identified. This cluster encodes four minor pilins *fimT*, *pilV*, *pilW*, and *pilX*, in the same orientation with respect to *pilY1* [8]. PilY1 is a high molecular weight protein associated with the pili tip and found in extracellular membrane [9–12].

Specifically, PilY1 was identified by Alm et al. [13] in *Pseudomonas aeruginosa* and described as a PilC2 gonococcal homologue due to the fact that the C-terminal domain of both is similar. The 127 kDa PilY1 protein was found in membrane and in sheared pili, and the deletion of the *pilY1* gene triggers loss of twitching motility and the disruption of pili biogenesis in mutants, suggesting a priming role in pili assembly. Therefore, the authors suggested that PilY1 participates before and during pilus assembly and is part of pili [13]. One year earlier, Rudel et al. [14] reported PilC2 and identified it as the T4P tip adhesin from purified pilus in the genus *Neisseria*; they indicated that the PilC adhesion property is essential for cell adherence and subsequent infection. The C-terminal similarity and its location at the pili tip were the reasons why PilY1 was also defined as an adhesin.

The PilY1 C-terminal is a conserved domain and folds into an eight-beta-propeller structure (usually named *Neisseria* PilC beta-propeller domain) similar to that predicted in PilC2, which interacts with the minor pili complex. Additionally, this C-terminal domain of PilC2 and PilY1 contains an EF-hand Ca^{2+}-binding motif where the release of calcium ion (Ca^{2+}) serves as a signal for the regulation of pili retraction, affecting PilT ATPase [15–18].

The N-terminal of PilY1 is not conserved, but diverse bacteria contain a domain that participates in crucial roles such as sensing, cell recognition, or adhesion. This domain includes the von Willebrand A-like factor (vWA) and the metal ion-dependent adhesion site (MIDAS) motif [19–22]. The vWA "is structurally similar to the A2 domain of the human vW factor, a force-sensing glycoprotein important in stopping bleeding" [23]. In bacteria, this domain is involved in migration, mechanosensing, signaling, adhesion, and biofilm formation [23], while the MIDAS motif resembles the cellular adhesion role of the vWA in α-β integrin chains [24].

PilY1 contributes not only to the structure of pili but also to signaling cascades that regulate twitching motility, pilus biogenesis and retraction, and bacterial behavior on surfaces such as cell adhesion processes through molecular binding, sulfur bridges, hydrogen bonds, electrostatic forces, and sense of shear force [10,19,21,25–27]. To date, PilY1 has been well studied in pathogens [17,21,28–33], however only the tertiary structure of the PilY1 C-terminal domain of *Pseudomonas* is currently known [17] and no information regarding the N-terminal or full-length structure has been reported. However, the investigation in the PilY1 from extremophiles such as acidophiles has been limited to the reporting of their expression [20,34] and its primary and secondary structure analyses [8], nevertheless "extremophiles are already used in many fields of biotechnology and hold great potential for future exploitation" [3]. Specifically, the adhesin PilY1 of *At. thiooxidans* may play a significant role during cell attachment to the mineral surfaces during mineral biooxidation for metals bioleaching [5].

According to Koonin et al. [35], 'gene mining' must lies *in silico* to identify the sequence-function pipeline of novel genes in assembled metagenomic sequence datasets. We must consider that working with extremophilic proteins presents significant challenges related to data availability, purification, stability, structural plasticity, amino acid composition, adaptation mechanisms, and ecological niche factors. Overcoming these difficulties requires innovative approaches and continued research in this field [36–39]. Here, we conduct *in silico* analyses of the predicted tertiary structure of PilY1 of *At. thiooxidans*. First, we describe the conserved C-terminal and its homology with *P. aeruginosa* and *Neisseria* PilC2, enabling us to delineate a potential interaction interface with the minor pilins on the pili tip. We then describe the non-conserved and less known region of PilY1, the N-terminal domain, focusing on the vWA because its central roles, e.g. surficial mechanosensor and biofilm formation on minerals. The tertiary structure analyses and sequence alignments were conducted comparing vWA' PilY1 and its homologues in neutrophilic bacteria and eukaryotes. Further computational characterization must show that despite the complexity of this protein, it is designed to withstand the hostile environment where *At. thiooxidans* thrives. To corroborate these findings, the N-terminal domain of PilY1 was cloned and expressed in *Escherichia coli*. Although the protein was initially expressed as inclusion bodies, it was successfully solubilized using the detergent sarkosyl, followed by an on-column refolding process. Stability was assessed using fluorescence spectroscopy, which preliminarily indicated that the protein remains stable across different pH levels. Our results have facilitated the characterization and production of PilY1, offering valuable guidelines for designing robust methodologies to express homologous proteins from other extremophilic organisms in heterologous systems. This ensures the production of functional proteins for various downstream applications.

2. Methodology

2.1. Structure prediction and sequence analysis of PilY1

The sequence for the PilY1 protein of *At. thiooxidans* ATCC 19377 and of *P. aeruginosa* are AWP39905.1 and AAA93502.1, respectively; AFA53726.1 for PilC of *Neisseria meningitides* were also used for comparisons. Sequences were obtained from GenBank (https://www.ncbi.nlm.nih.gov/genbank/).

Secondary structure and structural disorder of PilY1 were predicted using, respectively, the FELLS server (http://old.protein.bio.unipd.it/fells/) [40] and the PONDR® server (http://www.pondr.com/), applying the VL-XT algorithm [41] (see Table 1).

Predictions of tertiary structure of the PilY1 and its vWA domain from *At. thiooxidans*, *P. aeruginosa* PAO1 and *N. meningitides* were obtained using the AlphaFold2 program ColabFold V1.5.2 (https://alphafold.ebi.ac.uk/) following default recommendations [42,43]. The complete amino acid sequence of PilY1 and PilC2 and the sequence of vWA without the signal peptide of both PilY1 were submitted and predicted with default parameters in each prediction run, yielding PDB files of each structure and the predicted local distance difference test (pLDDT). The C-terminal domain of PilY1 crystal structure (PDB ID: 3HX6) [17] was obtained from the RCSB Protein Data Bank (https://www.rcsb.org/). Tertiary structure alignments and domain localization were visualized and computed using UCSF Chimera 1.16 (www.cgl.ucsf.edu/chimera/) [44].

To align the tertiary predicted structure of PilC2 with similar proteins, we used FoldSeek Server (https://search.foldseek.com/search) [45]. This software compares tertiary structures with protein that have similar shapes, giving us quantitative data about TM-align and RMSD scores.

SignalP-6.0 (https://dtu.biolib.com/SignalP-6) was employed to identify the presence of a signal peptide in PilY1 and PilC2 [46]. The Ca^{2+} binding motifs were predicted using the MIB2 Metal Ion-Binding site prediction and modeling server (http://combio.life.nctu.edu.tw/MIB2/), which uses structural and sequence information to identify protein-metal interaction sites [47].

Table 1
Software and databases used for protein characterization.

Software	Brief description	Reference and URL (http://)
FELLS	Fast Estimator of Latent Local Structure: provides predictions for protein disorder, secondary structure, and aggregation propensity	[40] old.protein.bio.unipd.it/fells/
PONDR	Predictor of Natural Disordered Regions: disorder predictions from amino acid sequence	[41] www.pondr.com/
ColabFold2	Utilizes the MMseqs2 server to generate multiple sequence alignments (MSAs) for protein structure prediction from AlphaFold2	[42,43] colab.research.google.com/github/
UCSF Chimera	Primarily used for interactive visualization and analysis of 3D molecular models	[44] www.cgl.ucsf.edu/chimera/
FoldSeek	For comparisons of protein structures by using the MMseqs2 server, single-chain protein structures in PDB/mmCIF format against a target database	[45] search.foldseek.com/search
SignalP-6.0	Prediction of the presence of signal peptides, its location for cleavage sites in protein sequences	[46] dtu.biolib.com/SignalP-6
MIB2	Metal Ion-Binding site prediction and modeling server in proteins. Identify residues likely to bind metal ions.	[47] combio.life.nctu.edu.tw/MIB2/
BIPSPI	Interface Prediction of Specific Partner Interactions: uses a machine learning-based method to predict partner-specific protein-protein interaction sites on protein surfaces	[49] bipspi.cnb.csic.es/
MEGA 11	To analyze DNA and protein sequence data from various species; to tasks phylogenetic relationships, inference and calculation of evolutionary distance	[50] www.megasoftware.net/
PDB2PQR	Prepares molecular structures for electrostatics calculations converting PDB files into the PQR format, that includes optimized hydrogen bonding and charge assignments	[51] server.poissonboltzmann.org/pdb2pqr
WebGRO	To performe fully solvated molecular dynamics (MD) simulations of macromolecules analyzing the resulting trajectories	[52] simlab.uams.edu/index.php
PDBMD2CD	Prediction of circular dichroism (CD) spectra for multiple protein structures obtained from molecular dynamics (MD) simulations	[53] pdbmd2cd.cryst.bbk.ac.uk/
ProtParam	Expert Protein Analysis System: To compute various physical and chemical parameters for a given protein sequence. To calculate molecular weight, isoelectric point, extinction coefficient, and amino acid composition of a protein, among others	[55] web.expasy.org/protparam/
GenBank of NCBI	DNA and protein sequences from the International Nucleotide Sequence Database Collaboration, DNA DataBank of Japan, and European Nucleotide Archive	www.ncbi.nlm.nih.gov/g
STRING	Search Tool for the Retrieval of Interacting Genes/Proteins: encompasses both physical and functional associations between genes or proteins, encompasses both physical and functional associations to analyze protein-protein interaction networks	[48] and 2023 string-db.org/

2.2. Prediction of functional interaction partners and interface interaction with PilY1

To find potential interaction partners, the STRING database (https://string-db.org/) was utilized. STRING is a comprehensive database specialized in cataloging known and anticipated protein-protein interactions, including both direct (physical) and indirect (functional) associations. It amalgamates evidence from diverse sources such as automated text mining, interaction experiments, annotated complexes/pathways, and computational predictions. With its forthcoming version 11.5, STRING aims to achieve extensive coverage, encompassing over 14,000 organisms [48]. Subsequently, the BIPSPI (xgBoost Interface Prediction of Specific Partner Interactions) web server (https://bipspi.cnb.csic.es/) was used to identify which residues are involved in the interaction between PilY1 and the minor pilins predicted by STRING. It employs the XGBoost algorithm and is designed as a three-step workflow involving XGBoost classifiers and a scoring function to convert interacting pair predictions into binding site residue scores. BIPSPI has shown high performance compared to other methods [49].

2.3. Primary structure and phylogenetic analysis of vWA

We analyzed the phylogenetic relationships using the protein sequences of the vWA of the PilY1/PilC family of some *Acidithiobacillus* spp. To polarize the obtained tree, the vWA containing the integrin sequence of *Homo sapiens* was used. The sequences obtained from GenBank (https://www.ncbi.nlm.nih.gov/genbank/) were aligned with those obtained in this work using MEGA11 software [50].

2.4. Protonation and deprotonation of vWA domain at different pH

The protonation status of acidic and basic residues in the PilY1 vWA domain was analyzed using the PDB2PQR web server (https://server.poissonboltzmann.org/pdb2pqr) [51]. PDB2PQR utilizes the PROPKA algorithm alongside with the PARSE force field to calculate the ionization state of the protein at specific pH values. The submitted PilY1 vWA structure was modified by incorporating protons, which were added according to the ionization states of the titratable groups at pH 1, 2, 3, 4, 5, 6, 7, and 8.

2.5. Molecular dynamic and circular dichroism at different pH

Molecular dynamics (MD) simulations of the vWA domain exposed to different pH were carried out using WebGRO, an online automated program which uses the GROMACS package for MD simulations (https://simlab.uams.edu/index.php) [52]. The stability of PilY1 vWA was assessed based on changes, primarily in the RMSD (root mean square deviation), RMSF (root mean square fluctuation) and Rg (radius of gyration). MD runs were carried out with the default parameters, including the GROMOS96 43a1 force field, a triclinic water box with simple point charge (SPC), neutralized system with 150 mM of NaCl, the protein energy minimizations with a maximum of 5000 steps, and constant temperature (300 K) and pressure (1.0 bar). The trajectories were saved at 1000 frames in each simulation and the simulation time was 50 ns.

Circular dichroism (CD) spectra of the vWA domain were calculated with the PDBMD2CD program [53] at the web server (https://pdbmd2cd.cryst.bbk.ac.uk/), starting from the molecular dynamic simulation data (e.g. MD trajectories) and generating spectra for each of the structures provided at each pH evaluated.

2.6. PilY1 N-terminal domain purification

The cDNA of the PilY1 N-terminal domain (PilY1N) (residues 32–583) was obtained by RT-PCR. It was cloned between the EcoR1 and Sal1 sites of the pET-32b(+) vector, to express the recombinant protein as a fusion protein with thioredoxin (Trx). The resulting construct, pET32-*pilY1N*, was used to transform *E. coli* BL21-CodonPlus (DE3) cells (Agilent Technologies, Inc. 2015). A transformed colony was used to grow a pre-inoculum, which was incubated at 37 °C overnight. This preinoculum was used to inoculate 1 L of LB medium, and bacterial growth was monitored until an optical density (OD 600) of 0.8 was reached. Protein expression was then induced with isopropil ß-D-1-thiogalactopyranoside (IPTG) at a final concentration of 1 mM, followed by incubation at 16 °C overnight. The cells were harvested and stored at −20 °C until use. The bacterial pellet was resuspended in lysis buffer [50 mM HEPES pH 8.0, 100 mM NaCl, 0.1 mg/mL lysozyme] and disrupted by sonication (50 % amplitude; 15 s pulse on, 45 s pulse off). The soluble and insoluble fractions were separated by centrifugation at 15,000×*g* for 30 min at 4 °C. The supernatant was discarded, and the inclusion bodies were washed twice with 2 % (v/v) Triton X-100 and 1 M urea. The inclusion bodies were solubilized with solubilization buffer [50 mM Tris-HCl pH 8.0, 200 mM NaCl, 20 mM 2-mercaptoethanol, 2 % (w/v) sarkosyl]. The sample was stirred overnight, and the solubilized protein was then obtained by centrifugation at 15,000×*g* for 30 min at 4 °C.

The solubilized Trx-PilY1N was passed through a Ni column for refolding. The refolding process was performed using the method described by Oganesyan et al. [54]. Briefly, the bound protein was renatured by washing with the following buffer [20 mM Tris-HCl pH 8.0, 500 mM NaCl, 0.1 % (v/v) Triton X-100]. To remove the detergent, the column was washed with [20 mM Tris-HCl pH 8.0, 100 mM NaCl, 5 mM β-cyclodextrin]. To remove the cyclodextrin, the column was washed with 20 mM Tris-HCl pH 8.0, 100 mM NaCl, 0.5 % (w/v) CHAPS, 5 % (v/v) glycerol. Finally, the protein was eluted with the same buffer supplemented with 300 mM imidazole. Subsequently, the sample was desalted in desalting buffer [25 mM Tris-HCl pH 8.0, 150 mM NaCl, 2.5 mM $CaCl_2$, 0.5 % (w/v) CHAPS, 5 % (v/v) glycerol], and the fusion protein was cleaved with thrombin. The sample was then desalted in the same buffer without $CaCl_2$ and passed through an anion exchange column to separate PilY1N from Trx. Finally, the purified protein was concentrated and quantified by measuring the absorbance at 280 nm, using the extinction coefficient (78870 M^{-1} cm^{-1}) obtained from ExPASy ProtParam (https://web.expasy.org/protparam/) [55].

2.7. Fluorescence spectroscopy

The fluorescence emission spectra of PilY1N were measured at various pH levels using a fluorescence spectrophotometer (Agilent Technologies Cary Eclipse) with a 1 cm path length quartz cuvette. The PilY1N protein was diluted to 3 μM using different buffers at pH values of 2.8, 3.6 and 8.0. The buffers used (100 mM each) were glycine-HCl (pH 2.8), acetate (pH 3.6) and sodium phosphate (pH 8.0). Samples were excited at 280 nm, with emission spectra recorded in the 300–400 nm range, and both excitation and emission slit widths set to 5 nm. Prior to recording, samples were incubated in the dark at room temperature for 5 min. All values are reported as the mean of three independent measurements.

3. Results and discussion

3.1. The PilY1 structure

Upon analyzing and comparing the predicted three-dimensional (3D) structure of PilY1 of *At. thiooxidans*, *P. aeruginosa*, and PilC2 of *Neisseria meningitides,* two well differentiated domains can be identified: the N- and C-terminal. (Fig. 1). The pLDDT graphics for the adhesin models show peaks near 100 % with sudden drops. These drops are more common in the *At. thiooxidans* PilY1 model and in the first 400 residues of PilC2 model, indicating that some regions are difficult to predict (data not shown) because of the lack of counterparts/homologues. To quantify the similarity between N- and C- terminals of PilY1 proteins and PilC2, we use the FoldSeek Search algorithm to compare predicted PilC and PilY1 structures against other structures deposited in the European Bioinformatics Institute and the ESM Atlas [45]. The PilC2 predicted structure (Fig. 1c) was the query structure; the FoldSeek results showed that the query structure of PilC2 fits with the pilus assembly protein of *Neisseria gonorrhoeae* (PilC), covering the complete structure with a RMSD of 2.33; while for PilY1 of *P. aeruginosa* it is 6.35, due to the structure covering only the residues 347–1044 of the query structure; i. e. the N-terminal of our query structure does not overlap with it. These results confirm that the N-terminal domain of PilC proteins is dissimilar to the corresponding domain in PilY1.

3.2. The interactions of the C-terminal domain of PilY1

The C-terminal domains for PilY1 of *P. aeruginosa* (PDB ID: 3HX6) and the predicted with AlphaFold2 of *At. thiooxidans* were compared (Fig. 2). The β-propeller shape was found in both PilY1- but as expected, the PilY1 C-terminal domain of *P. aeruginosa* showed the higher match score (RMSD: 2.30; data not shown) and that of *At. thiooxidans* has the lowest (RMSD: 14.39; Fig. 2a). The difference between the RMSD values indicates that the arrangement of the secondary structures is not identical while the tertiary structure of the β-propeller folding is maintained.

According to the models, the PilY1 C-terminal domain of *At. thiooxidans* contains six well-formed blades and one incomplete blade consisting of only two β-strands (Fig. 2b) similar to the seven β-propeller fold in the PilY1 C-terminal domain of *P. aeruginosa* ([17] and Fig. 2c). The β-propeller is a circular fold formed by four to ten repeats that may result from the diversification of an ancestral four-stranded or β-blade implied in protein-protein interactions [56]. In *M. xanthus* and *P. aeruginosa*, the interaction between PilY1 and the minor pilins complex occurs among the PilY1 C-terminal and PilX and PilW [18,26],

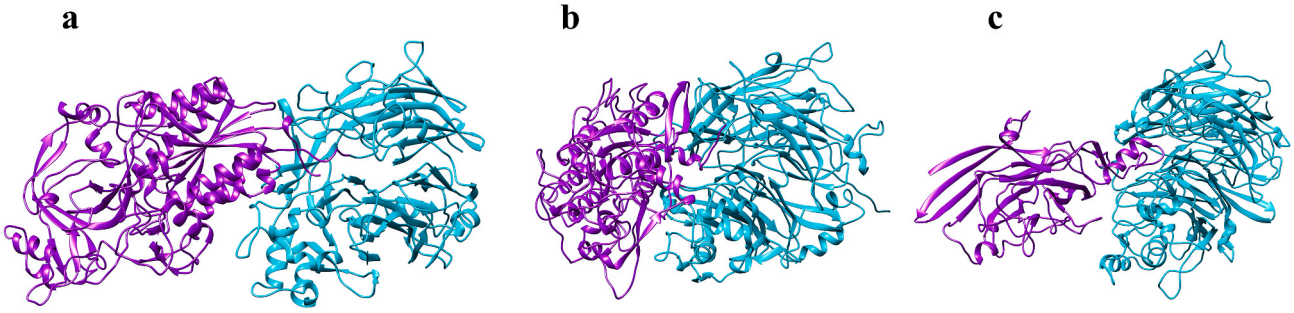

Fig. 1. AlphaFold2 predicted tertiary structure of (a) PilY1 of *At. thiooxidans* (1176 residues) (b) PilY1 of *P. aeruginosa* (1161) residues and (c) PilC2 of *N. meningitides* (1048 residues). The three structures showed the N- and C-termini in purple and cyan, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

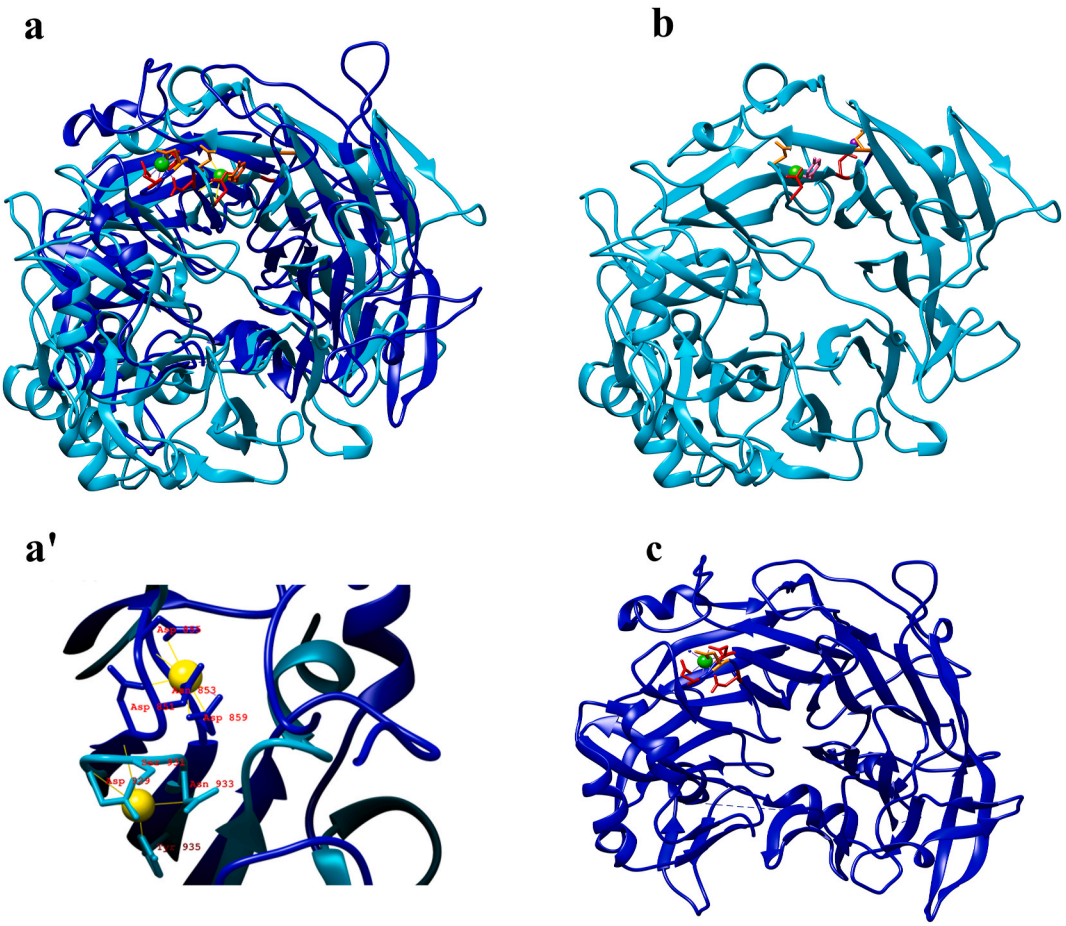

Fig. 2. 3D structure of C-terminal domain of PilY1 predicted by AlphaFold2 from *At. thiooxidans* (cyan) and from the crystal structure of *P. aeruginosa* PDB 3HX6 (in blue). Both structures are shown overlapping (a) and separately (b and c) with the EF-hand Ca^{2+}-binding motif highlighted; a close-up of such divalent cations (yellow spheres) within predicted binding sites is shown in (a'). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

which could suggest the role of the β-propeller in the interaction of PilY1 with minor pilins. In fact, STRING analyses of PilY1 from *At. thiooxidans*, show functional associations with various proteins, some of which are members of the PilMNOP machinery, which refers to a minimal system capable of assembling periplasmic T4P [57]. Interestingly, the other proteins with which an interaction is predicted are the minor pilins PilX, PilW, and PilV (Fig. 3), with scores (predicted functional partners) of 0.933, 0.903, and 0.833 respectively, these also being the highest in the interaction network predicted by this program. To date, the possible interaction interface of PilY1 with the minor pilins has not been previously reported *in silico*.

We are aware of the complications involved in using molecular docking to find potential interaction interfaces due to the lack of experimental information on this and other protein-protein interaction systems. To address this issue, we decided to utilize the BIPSPI method to identify the potential residues of PilY1 for interacting with the minor pilins PilW and PilX. Interestingly, despite the full-length structure of PilY1 being submitted to the BIPSPI server, the residues with the highest scores (threshold ≥0.6) for interaction between the minor pilins and the protein are located around a cavity on the outer face of the C-terminal domain (Fig. 3c).

The prediction showed that residues Y1000, Y1001, G1771, W1172, and K1173 of PilY1 could potentially interact with both minor pilins, with only residues Q1002 for PilW and residues L1174 and V1175 for PilX being different between the minor pilins. This delimited zone by these shared interaction residues between the two pilins indicates a likely hot spot for synergistic interaction of PilY1 with the minor pilins, since the results from BIPSPI regarding PilY1 and PilV using the same precision threshold also indicate that residues Q1002, G1771, and W1172 are predicted to interact with this minor pilin.

Additional insights into the C-terminus of PilY1 and the PilC protein were reported by Parker et al. [58], highlighting an EF-hand Ca^{2+}-binding motif. This motif is involved in Ca^{2+} ion signaling which regulates the extension and retraction of pili in *P. aeruginosa* [17] and *Kingella kingae* [59]. This Ca^{2+} binding motif is a conserved sequence (Dx[**DN**x]xD**G**xxD) in PilY1 homologues [58]. Thus, we specifically search for divalent cations (Ca^{2+}, Mg^{2+} and Mn^{2+}) binding motifs in the PilY1 C-terminal region (Fig. 2b) and comparing through structure homology with *P. aeruginosa* we found a possible motif **DASGN** (D929-N933), (Fig. 2a' where D: Asp, S: Ser and N: Asn); other suggested motif could bed **DNTGN** (D938-N942; data non show). The former suggested that the C-terminal domain of PilY1 of *At. thiooxidans* may have the same functions as have been reported for homologues of PilY1 and PilC of *P. aeruginosa* and *Neisseria* spp.

3.3. PilY1 of At. thiooxidans possess a signal peptide type I

For major and minor pilins, the matured form is due to the PilD peptidase which has cleavages in the well-conserved signal peptide type III. For the *P. aeruginosa* PilY1, the signal peptide type was predicted by Lewenza et al. [60] as a prepilin signal peptide with a type III signal peptide processed by PilD. However, type I signal peptides have been reported in PilY1 of *P. aeruginosa* [26] and in PilY1.1, PilY1.2 and PilY1.3 of *M. xanthus* [18]. Our results suggest that PilY1 of *At.*

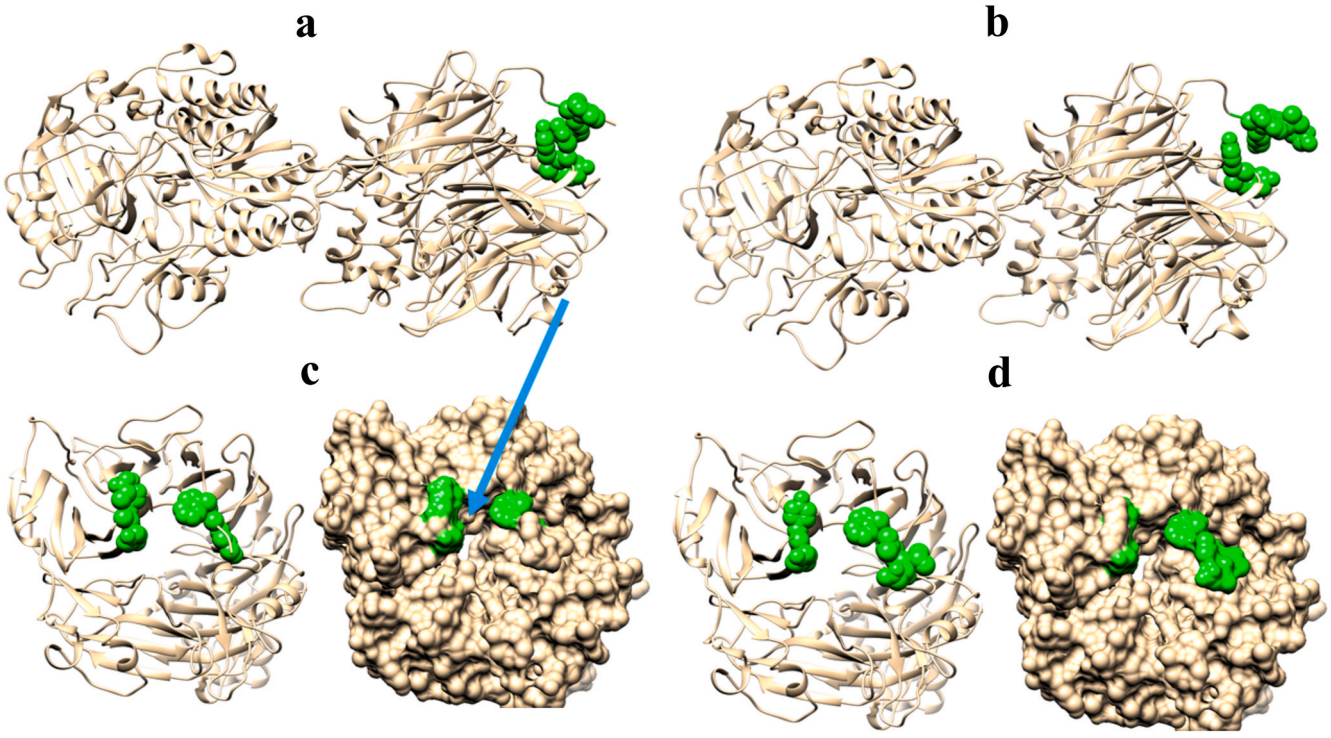

Fig. 3. Ribbon (left) and surface (right) representations of PilY1, highlighting the suggested residues (green spheres) involved in the interfacial interaction between the C-terminal domain of PilY1 and minor pilins. (a) and (b) lateral view; (c) and (d) front view. Blue arrow in (c) indicates the protein cavity. Panels (a) and (c) correspond to PilW, (b) and (d) to PilX. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

thiooxidans had a type I signal peptide (confidence score >96 %) and its predicted cleavage site occurs between A32–N33. We used the *P. aeruginosa* PilY1 to compare, and the predicted signal peptide is indeed type I, while the cleavage occurs between A32–L33. The above data suggests that *At. thiooxidans* PilY1 is matured and cleaved by a type I peptidase as its homologues *P. aeruginosa* and *M. xanthus*.

3.4. The vWA is a conserved domain in PilY1 of At. thiooxidans

In addition to the signal peptide type I, the PilY1N contains a vWA domain [18,20,21]. The vWA is widely spread along the three life domains and it had been implicated in sensing, cell recognition, and shear force [23,61–65].

The AlphaFold2 predicted vWA domains of PilY1 in both *At. thiooxidans* and *P. aeruginosa* (Fig. 4b) show the basic and global structure, but the vWA of *At. thiooxidans* has five β-sheets surrounded by six α-helices (Fig. 4a) known as α/β Rossman folds [66,67], and the vWA of *P. aeruginosa* has only four β-sheets and four surrounded α-helices (Fig. 4b). Remarkably, the folding of the vWA domain of human integrin alpha 1(PDB: 5HGJ) is similar to vWA domain of *At. thiooxidans* (RMSD:

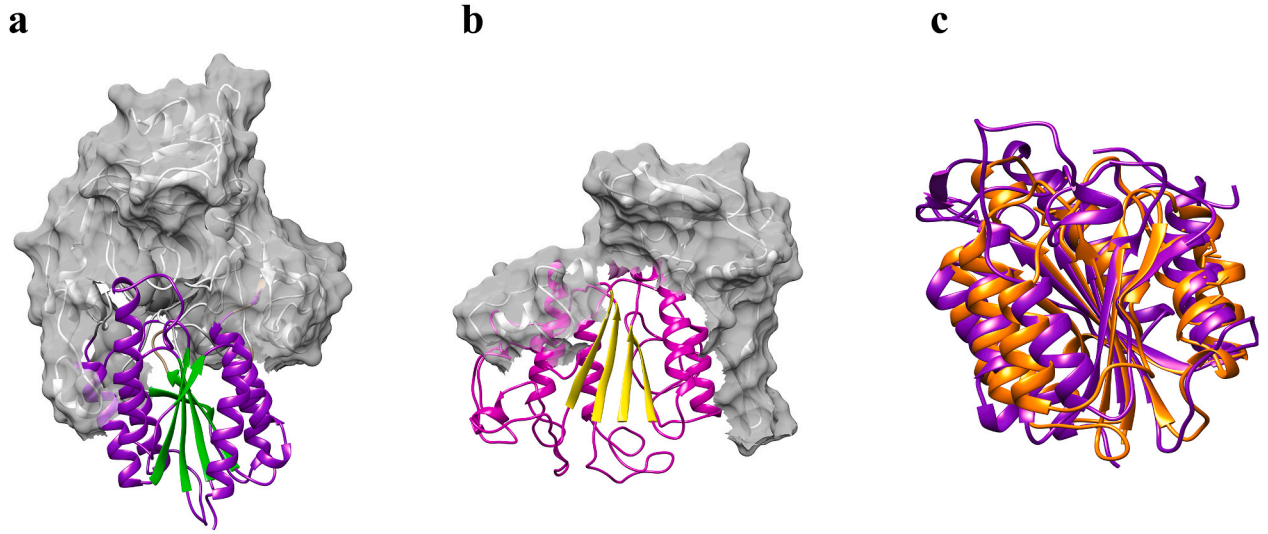

Fig. 4. The structure of the AlphaFold2 predicted vWA domain of PilY1. (a) The vWA of *At. thiooxidans* includes six α-helixes (purple) and five β-sheets (green). (b) The vWA structure of *P. aeruginosa* has only four α-helixes (magenta) and three β-sheets (yellow). (c) Merge between vWA domain of *At. thiooxidans* (purple) and the vWA domain of the integrin α1 for *H. sapiens* (orange) (PDB: 5HGJ). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

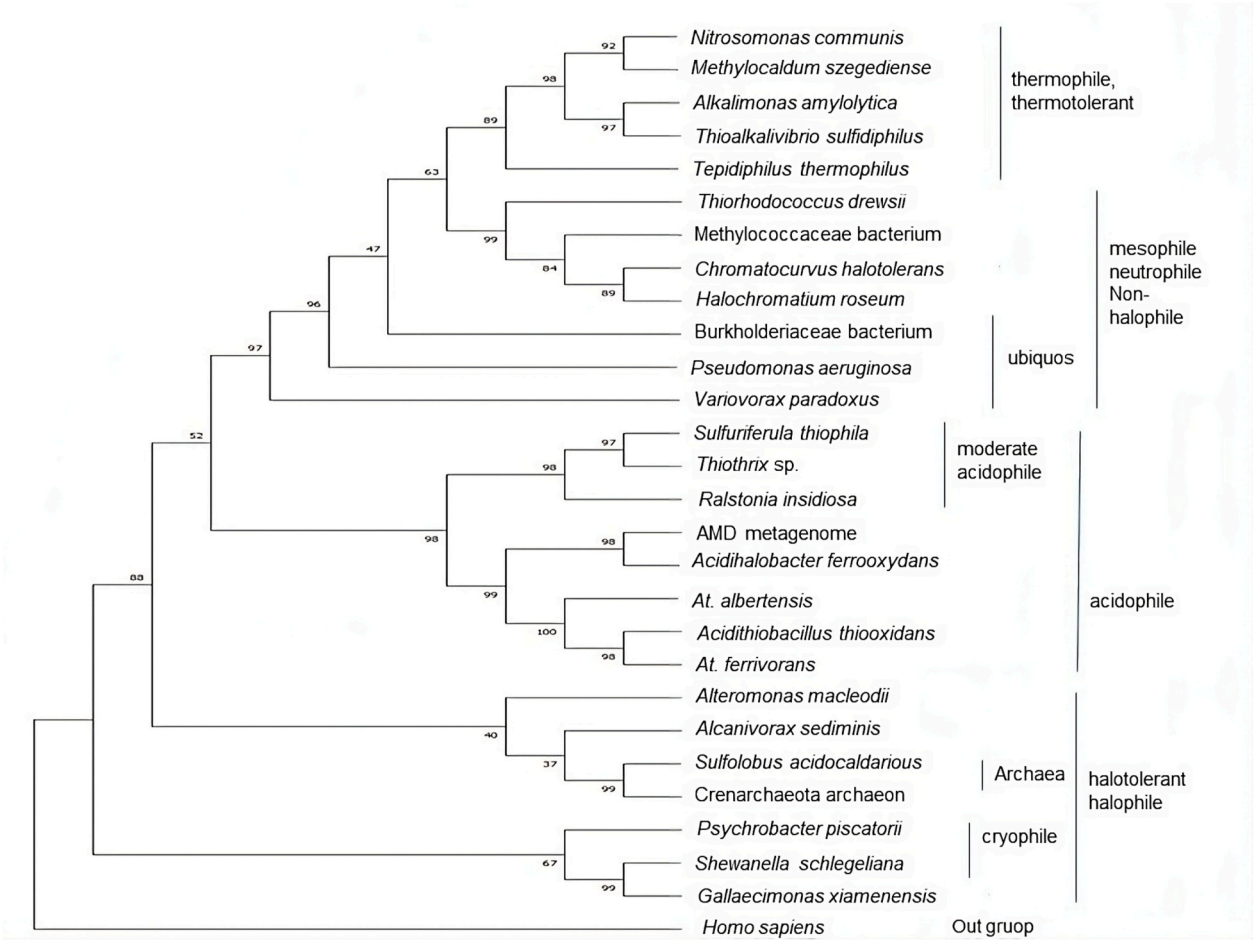

Fig. 5. Phylogenetic analysis of the vWA motif present in the PilY/PilC family of some Bacteria, in the archaellum regulatory network of an Archaea (*Sulfolobus acidocaldarious*) and in the integrin of an Eukarya, *Homo sapiens*. Evolutionary history obtained by ML log likelihood and the JTT matrix-based model. Bootstrap consensus tree inferred after 1000 replicates; the discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 3.2697)). The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 1.40 % sites). Analysis of 28 amino acid sequences, 713 positions in the final dataset. Evolutionary analyses were conducted in MEGA11 [50,71]. Accession numbers in Supplementary Table S1.

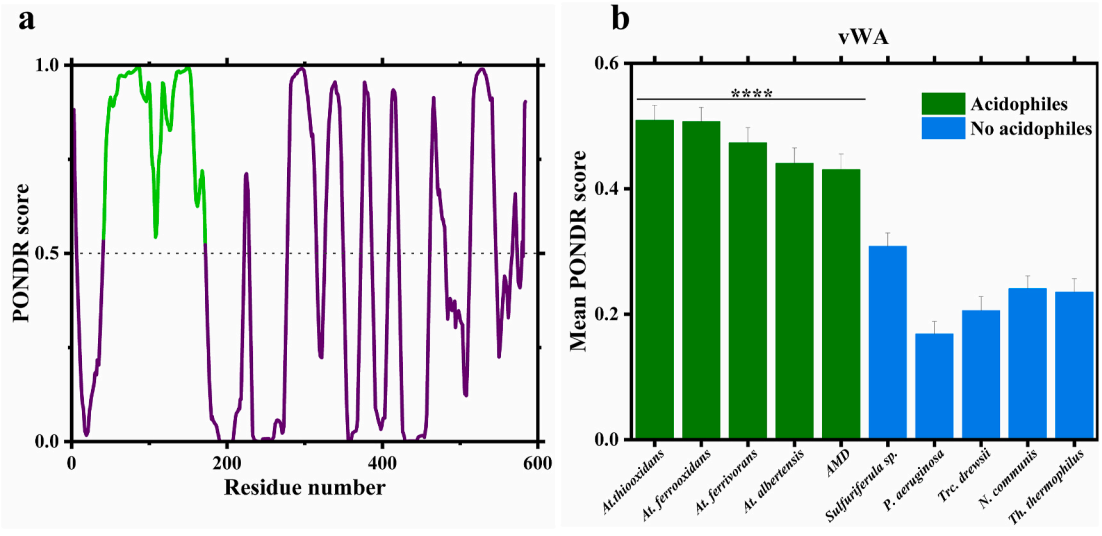

Fig. 6. PONDR disorder analysis of PilY1N. (a) Disorder prediction of PilY1 N-terminal (1–581). Values above dotted line indicate high potential for disorder. Green region in (a) indicates the order-disorder transition predicted by PONDR using VL-XT algorithm. (b) Comparison between predicted disorder in vWA domains from acidophiles (green) and their non-acidophile counterparts (blue). Graphic shows the one-way ANOVA comparison with $p < 0.0001$. AMD: acid mine drainage. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

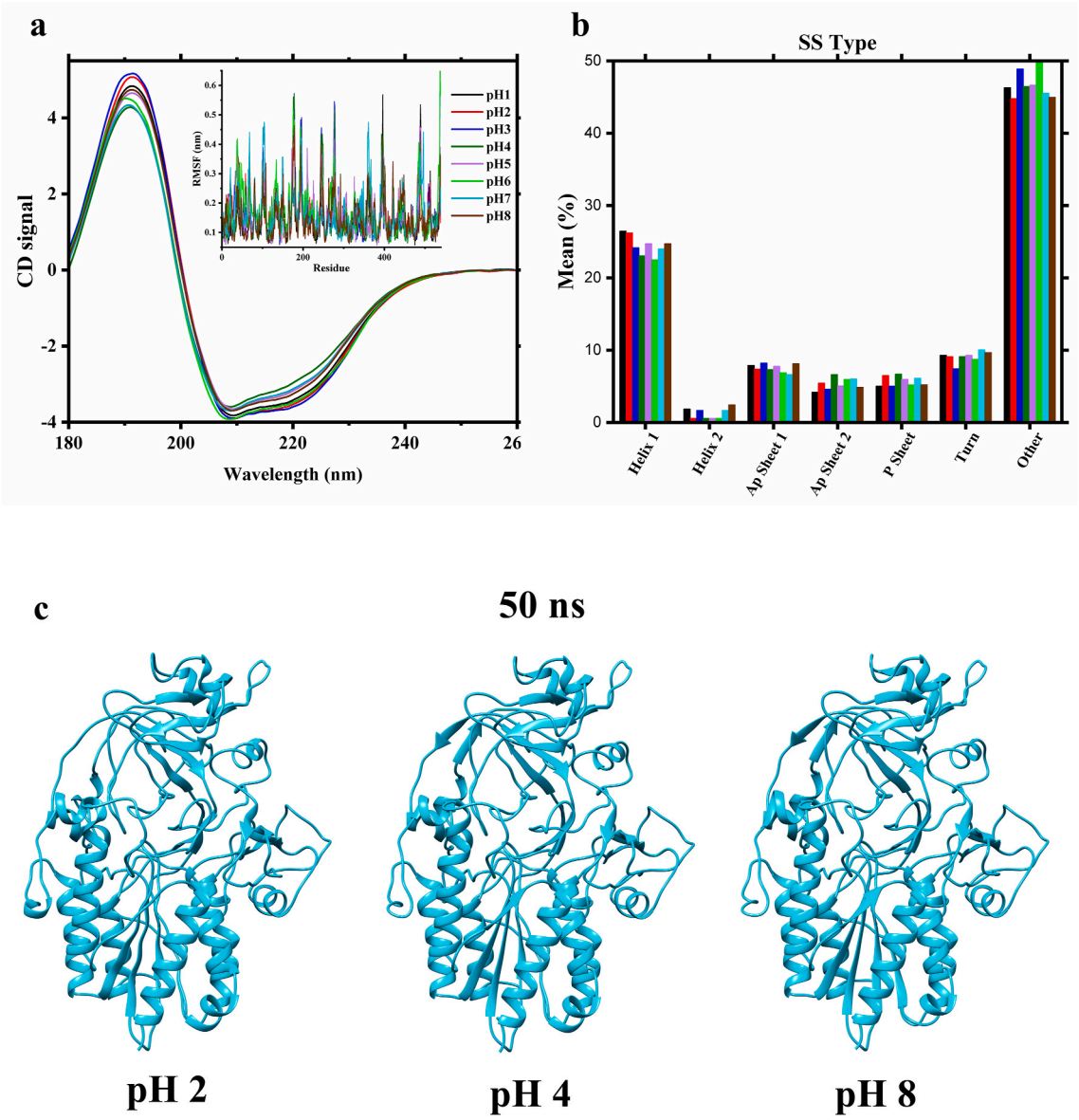

Fig. 7. Conserved structure of *At. thiooxidans* vWA domain. (a) CD spectrum for the vWA at different pH values. The inset represents the RMSF graph obtained from the MD (b) Percent of secondary structures predicted by CD. Horizontal axis shows secondary structure type and vertical axis the percent. Color code in (b) is the same as in (a). AP: antiparallel; P: parallel. (c) MD screenshots showing the whole protein surface at representative pH at the end of a 50 ns simulation. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

1.07) (Fig. 4c).

Analysis of the primary structure of the *At. thiooxidans* PilY1 reveals that the vWA domain (residues 42 to 581) includes the conserved metal-ion-dependent adhesion site (MIDAS) motif (DXSXS … T … D, residues 153–187) [20] (Fig. 4a). The MIDAS motif is a metal-ion binding site similar to the vWA domain found in the integrins I-domain of eukaryotes, voltage-gated Ca^{2+}, and other intra-extra cellular proteins [61,62,68], and it has been reported in other PilY1 vWA domains and other proteins involved in binding extracellular proteins acting as integrin-like switches [23,24,69,70]. Thus, during the mineral-*At. thiooxidans* interaction (adhesion, biofilm formation), the MIDAS motif may participate in binding divalent metallic ions such as Ca^{2+}, Mn^{2+} or Mg^{2+} through loops expanded from the vWA, exposed on the protein surface.

MIDAS is involved in conformational changes of the vWA structure, but the cysteines stabilize the domain [23]. Such cysteines are also involved in mechanosensing; e.g. the glycoprotein vWA factor of eukaryotic cells sense the force across the breaking of sulfide bonds that are usually formed via cysteines interaction, modifying the protein redox state [23]. The PilY1' vWA of *At. thiooxidans* comprises five of seven cysteines (C120, C133, C230, C244 and C444), which may play a vital role during its mechanosensing and attachment to inorganic surfaces such as sulfur minerals such as sphalerite or chalcopyrite (ZnS or $CuFeS_2$, respectively).

Previously, we reported that the phylogeny analysis of the PilY1 complete sequence of *At. thiooxidans* clusters with other iron/sulfur oxidizing microorganisms [20]. In this work, the phylogenetic analysis of the vWA showed that the acidophiles form a separate and cohesive group with the *Acidithiobacillus* genus. group, except for the thermoacidophilic Archaea *Sulfolobus acidocaldarious* (Fig. 5). Also, the phylogenetic tree reveals that the vWA domain of PilY1 of some acidophiles and non-acidophiles are indeed homologues.

3.5. The PilY1 vWA domain has intrinsic disorder and displays pH resilience

Intrinsic disorder has been identified in extracellular pilins and PilY1

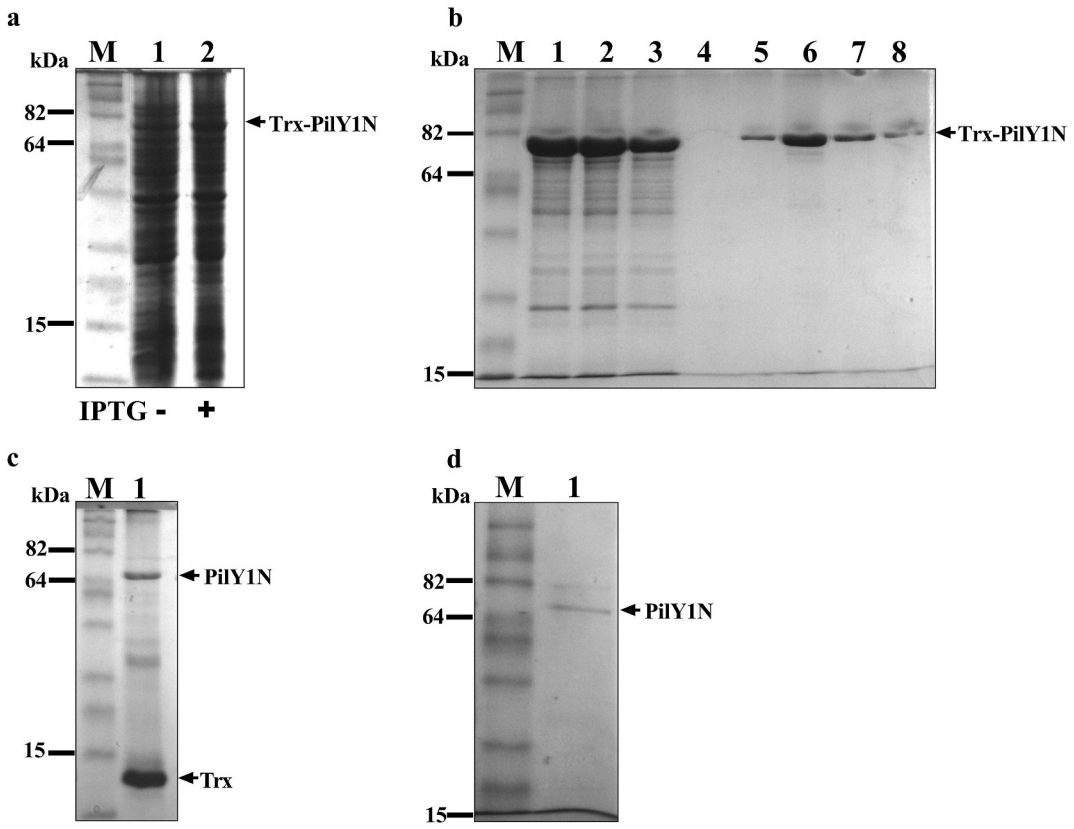

Fig. 8. Expression and purification of PilY1N of *Acidithiobacillus thiooxidans*. (a) Heterologous expression of PilY1N in *E. coli* BL21-CodonPlus (DE3)-RIPL with the pET32-*pilY1N* vector. M: molecular weight. Lane 1: protein profile before induction. Lane 2: protein profile after induction (Trx-PilY1N: 76.52 kDa). (b) Refolding of PilY1N from inclusion bodies with 2 % (w/v) Sarkosyl. Lane 1: solubilized fraction. Lane 2: unbound fraction. Line 3 and 4: washed fractions. Lane 5–8: eluted fractions with 300 mM imidazole. (c) Cleavage of the Trx-PilY1N fusion protein with thrombin protease. Lane 1: cleaved fusion protein after incubation with thrombin protease (Trx: 13.9 kDa and PilY1N: 62.6 kDa). (d) PilY1N obtained after purification by anion exchange chromatography. Lane 1: purified PilY1N.

of *At. thiooxidans* that could confer an advantage in acidic environments [8]. Assuming that the structure of the vWA domain must be maintained under harsh conditions, we decided to evaluate the intrinsic disorder and other features of the PilY1 vWA domain in more detail. Our results show a segment between Q42 and I171 that is predicted to transition from disorder to order (Fig. 6a) probably due to the presence of a ligand or cofactors and includes a portion of the MIDAS motif and the α1-helix of vWA (Fig. 4a).

Focusing on the vWA domain and investigating whether this disorder could be important in other PilY1 vWA domains, we perform a disorder analysis between the vWA domains of acidophiles and non-acidophiles. The data on disorder showed statistically significant differences in acidophiles versus non acidophiles (Fig. 6b). We applied a one-way ANOVA with the mean PONDR score of these domains. We found that in general, the structural disorder of the vWA domain in acidophiles is significantly higher ($p < 0.0001$) compared to their non-acidophilic counterparts. For the non-acidophiles, the vWA domain of the sulfur-oxidizing *Sulfuriferula* sp. is also significantly higher compared to the PilY1 vWA of *P. aeruginosa*.

The vWA in PilY1 of *At. thiooxidans* must be a pH-resistant protein. Three characteristics have been proposed to explain why the structure and function of extracellular proteins are maintained in acidophiles: presence of structural disorder, the enrichment of hydroxy and amide residues, and low surface charge [8]. Full-length PilY1 sequences keep these characteristics, but the stability of its vWA exposed to the acidic environment has not been evaluated. If the vWA domain is key in sensing surfaces for *At. thiooxidans*, an unchanged structure is expected.

The stability of the vWA domain at different pH (1–8) was evaluated through MD simulations. The results from the simple root mean square deviation (RMSD) and root mean square fluctuation (RMSF) exhibit closely aligned values, with a difference of scarcely <0.15 nm between successive pH values. This suggests that the globular structure of the vWA maintains stability across different pH. To confirm this, we conducted predicted CD analysis from MD trajectories. The CD analysis showed characteristic peaks of α/β proteins at approximately 210 and 222 nm across different pH (Fig. 7a). Notably, at pH 1 and 2, there was a slight increase in α-helix peaks and a decrease in β-sheet peaks. Interestingly, disorder persisted across all pH, with the lowest values observed at pH 2 and 8 (Fig. 7b).

The MD for the PilY1 vWA domain of *At. thiooxidans* showed that the tertiary and secondary structures remain stable through different pH levels. This stability is likely due to its tendency to disorder, as observed in other acidophiles [8].

The MD simulations conducted on the vWA domain of PilY1 in *At. thiooxidans* revealed remarkable stability in both tertiary and secondary structures across a spectrum of pH levels. This robustness can likely be attributed to the inherent tendency of the protein towards disorder, a characteristic shared with other acidophiles as highlighted in previous work [8]. These findings underscore the resilience of the vWA domain in maintaining its structural integrity, further reinforcing our earlier conclusions regarding the adaptive mechanisms of acidophiles to thrive in extreme environments.

3.6. Purification and acid stability evaluation of the PilY1N

To date, only the recombinant expression of the C-terminal and the vWA domains of *P. aeruginosa* PilY1 has been reported. Our initial attempts to express recombinant full-length PilY1 in *E. coli* resulted in the protein being expressed in a truncated and insoluble form (data not shown). Consequently, we decided to express the N-terminal domain

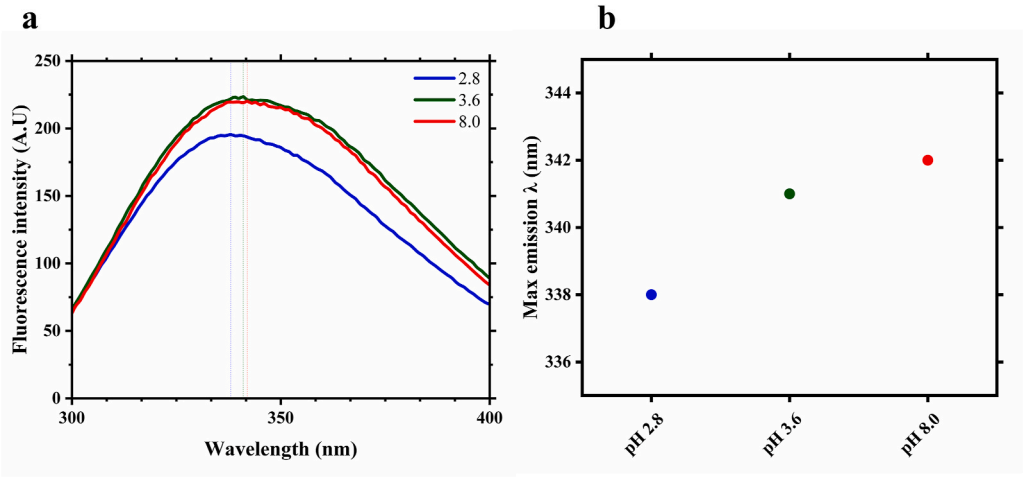

Fig. 9. Intrinsic fluorescence spectra of the PilY1N. (a) Emission spectra of PilY1N under different pH conditions. An excitation wavelength of 280 nm was used, and the emission was monitored in the range of 300–400 nm. (b) Maximum emission wavelength λ plotted as function of pH.

separately. Although we successfully expressed the domain independently as fusion protein with thioredoxin, it was also expressed as inclusion bodies (Fig. 8a). Various strategies were employed to solubilize the domain, including the use of chaotropic agents such as urea and guanidinium chloride, as well as detergents like SDS [72]. However, none of these agents were able to solubilize the fusion protein because the on-column refolding process could not be carried out; the protein eluted from the Ni resin column despite containing a 6x-His tag. After screening various conditions and solubilizing agents, we found that using 2 % sarkosyl [73] could solubilize part of the inclusion bodies of PilY1N (Fig. 8b). Furthermore, with the use of this detergent, the solubilized protein did not elute from the Ni column, allowing for successful on-column refolding. After the refolding process, the fusion protein was cleaved with thrombin (Fig. 8c), followed by an additional purification step using anion-exchange chromatography (Fig. 8d).

Once the purified PilY1N was obtained (Fig. 8d), it was concentrated and quantified to assess its acid stability through intrinsic fluorescence. Tryptophan fluorescence is a widely used technique for assessing protein stability by tracking conformational changes. The fluorescence intensity and maximum emission wavelength of tryptophan are indicators of protein stability. This method is particularly effective for monitoring protein unfolding and can be employed to study the effects of pH or temperature on protein conformational changes.

Changes in the tertiary structure of PilY1N were monitored by analyzing change in the intrinsic fluorescence emission. The change in the emission spectra of the protein at different pH values is shown in Fig. 9a. The emission spectra at pH 8.0 and 3.6 are very similar, with maximum fluorescence intensities at 342 and 341.1 nm, respectively, indicating a characteristic feature of folded proteins despite the transition from basic to acidic pH. Interestingly, when we evaluated the emission spectrum at pH 2.8, we observed a blue shift to 338 nm (Fig. 9b). This indicates that the tryptophan residues have become more buried in a less polar environment, possibly due to the protein adopting a more compact, folded conformation. Regarding the decrease in emission intensity, it may have resulted from the quenching mechanism, the protonation of water molecules, or the presence of acidic amino acids near the intrinsic fluorophores.

4. Conclusions

Even though the acidophilic bacteria *At. thiooxidans* and neutrophilic bacteria live under different environmental conditions, both have the pili tip associated protein PilY1. *In silico* characterization of PilY1 provides crucial information revealing the PilY1 protein structure of *At. thiooxidans* possesses the same representative domains as its non-acidophile homologues, e.g. the N-terminal vWA domain with MIDAS motif and the C-terminal domain with the β-propeller folding, which mainly interact with the minor pilins PilX and PilW. The vWA domain of PilY1 is particularly noteworthy, as it is conserved and comparable to other vWA domains, mainly in its 3D folding and the sequence of the ligand-binding MIDAS motif. The enriched presence of cysteines in vWA suggests its active role in adhesion and sensing. To maintain its structure and function in acidic environments, the vWA domain has intrinsically disordered regions; our results of MD and circular dichroism indicate that the structure of this domain is maintained over a wide pH range (1–8), as a resilient domain. To verify this, we successfully achieved the recombinant expression of PilY1N. Although the protein was initially obtained as inclusion bodies, we were able to solubilize and renature it. This enabled us to partially characterize this domain experimentally in terms of its acid stability, suggesting that the protein remains stable across a wide range of pH levels. Consequently, this provides a guideline for the characterization of more extracellular proteins from extremophilic organisms.

CRediT authorship contribution statement

Araceli Hernández-Sánchez: Writing – review & editing, Visualization, Supervision, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Edgar D. Páez-Pérez:** Writing – review & editing, Visualization, Supervision, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Elvia Alfaro-Saldaña:** Writing – review & editing, Formal analysis. **J. Viridiana García-Meza:** Writing – review & editing, Visualization, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work was supported by the National Council of Humanities Science and Technology (CONAHCyT-CB2017-2018 Project A1-S-

11505). EDPP acknowledges the support of this research through a postdoctoral scholarship from CONAHCyT (I1200/311/2023). The authors acknowledge Vanesa Olivares-Illana for her assistance with fluorescence spectroscopy.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbrep.2024.101797.

References

[1] Z. Yin, S. Feng, Y. Tong, H. Yang, Adaptive mechanism of *Acidithiobacillus thiooxidans* CCTCC M 2012104 under stress during bioleaching of low-grade chalcopyrite based on physiological and comparative transcriptomic analysis, J. Ind. Microbiol. Biotechnol. 46 (2019) 1643–1656, https://doi.org/10.1007/s10295-019-02224-z.

[2] T. Kamizela, A. Grobelak, M. Worwag, Use of *Acidithiobacillus thiooxidans* and *Acidithiobacillus ferrooxidans* in the recovery of heavy metals from landfill leachates, Energies 14 (2021) 3336, https://doi.org/10.3390/en14113336.

[3] D.A. Cowan, S.V. Albers, G. Antranikian, et al., Extremophiles in a changing world, Extremophiles 28 (2024) 26, https://doi.org/10.1007/s00792-024-01341-7.

[4] S. Kaushik, A. Alatawi, S.R. Djiwanti, et al., Potential of extremophiles for bioremediation, in: D.G. Panpatte, Y.K. Jhala (Eds.), Microbial Rejuvenation of Polluted Environment. Microorganisms for Sustainability, Springer, Singapore, 2021, https://doi.org/10.1007/978-981-15-7447-4_12.

[5] J.V. García-Meza, J.J. Fernández, R.H. Lara, I. González, Changes in biofilm structure during the colonization of chalcopyrite by *Acidithiobacillus thiooxidans*, Appl. Microbiol. Biotechnol. 97 (2013) 6065–6075, https://doi.org/10.1007/s00253-012-4420-6.

[6] M. Díaz, M. Castro, S. Copaja, N. Guiliani, Biofilm formation by the acidophile bacterium *Acidithiobacillus thiooxidans* involves c-di-GMP pathway and Pel exopolysaccharide, Genes 9 (2018) 113, https://doi.org/10.3390/genes9020113.

[7] D. Tang, Q. Gao, Y. Zhao, et al., Mg^{2+} reduces biofilm quantity in *Acidithiobacillus ferrooxidans* through inhibiting Type IV pili formation, FEMS Microbiol. Lett. 365 (2018), https://doi.org/10.1093/femsle/fnx266.

[8] E.D. Páez-Pérez, A. Hernández-Sánchez, E. Alfaro-Saldaña, J.V. García-Meza, Disorder and amino acid composition in proteins: their potential role in the adaptation of extracellular pilins to the acidic media, where *Acidithiobacillus thiooxidans* grows, Extremophiles 27 (2023) 31, https://doi.org/10.1007/s00792-023-01317-z.

[9] S. Burdman, O. Bahar, J.K. Parker, L. De La Fuente, Involvement of type IV Pili in pathogenicity of plant pathogenic bacteria, Genes 2 (2011) 706–735, https://doi.org/10.3390/genes2040706.

[10] Y. Luo, K. Zhao, A.E. Baker, et al., A hierarchical cascade of second messengers regulates *Pseudomonas aeruginosa* surface behaviors, mBio 6 (2015), https://doi.org/10.1128/mBio.02456-14.

[11] C.L. Giltner, M. Habash, L.L. Burrows, *Pseudomonas aeruginosa* minor pilins are incorporated into type IV pili, J. Mol. Biol. 398 (2010) 444–461, https://doi.org/10.1016/j.jmb.2010.03.028.

[12] L. Craig, K.T. Forest, B. Maier, Type IV pili: dynamics, biophysics and functional consequences, Nat. Rev. Microbiol. 17 (2019) 429–440, https://doi.org/10.1038/s41579-019-0195-4.

[13] R.A. Alm, J.P. Hallinan, A.A. Watson, et al., Fimbrial biogenesis genes of *Pseudomonas aeruginosa*: *pilW* and *pilX* increase the similarity of type 4 fimbriae to the GSP protein-secretion systems and *pilY1* encodes a gonococcal PilC homologue, Mol. Microbiol. 22 (1996) 161–173, https://doi.org/10.1111/j.1365-2958.1996.tb02665.x.

[14] T. Rudel, I. Scheuerpflug, T.F. Meyer, *Neisseria* PilC protein identified as type-4 pilus tip-located adhesin, Nature 373 (1995) 357–359, https://doi.org/10.1038/373357a0.

[15] P.C. Morand, E. Bille, S. Morelle, et al., Type IV pilus retraction in pathogenic *Neisseria* is regulated by the PilC proteins, EMBO J. 23 (2004) 2009–2017, https://doi.org/10.1038/sj.emboj.7600200.

[16] D.E. Clapham, Calcium signaling, Cell 131 (2007) 1047–1058, https://doi.org/10.1016/j.cell.2007.11.028.

[17] J. Orans, M.D.L. Johnson, K.A. Coggan, et al., Crystal structure analysis reveals *Pseudomonas* PilY1 as an essential calcium-dependent regulator of bacterial surface motility, Proc. Natl. Acad. Sci. U.S.A. 107 (2010) 1065–1070, https://doi.org/10.1073/pnas.0911616107.

[18] A. Treuner-Lange, Y.W. Chang, T. Glatter, et al., PilY1 and minor pilins form a complex priming the type IVa pilus in *Myxococcus xanthus*, Nat. Commun. 11 (2020) 5054, https://doi.org/10.1038/s41467-020-18803-z.

[19] A. Siryaporn, S.L. Kuchma, G.A. O'Toole, Z. Gitai, Surface attachment induces *Pseudomonas aeruginosa* virulence, Proc. Natl. Acad. Sci. U.S.A. 111 (2014) 16860–16865, https://doi.org/10.1073/pnas.1415712111.

[20] E. Alfaro-Saldaña, A. Hernández-Sánchez, O.A. Patrón-Soberano, et al., Sequence analysis and confirmation of the type IV pili-associated proteins PilY1, PilW and PilV in *Acidithiobacillus thiooxidans*, PLoS One 14 (2019) e0199854, https://doi.org/10.1371/journal.pone.0199854.

[21] S.S. Webster, C.K. Lee, W.C. Schmidt, et al., Interaction between the type 4 pili machinery and a diguanylate cyclase fine-tune c-di-GMP levels during early biofilm formation, Proc. Natl. Acad. Sci. U.S.A. 118 (2021) e2105566118, https://doi.org/10.1073/pnas.2105566118.

[22] A.L. Sacharok, E.A. Porsch, JW 3rd St Geme, The *Kingella kingae* PilC1 MIDAS motif is essential for type IV pilus adhesive activity and twitching motility, Infect. Immun. 91 (2023) e0033822, https://doi.org/10.1128/iai.00338-22.

[23] S.S. Webster, G.C.L. Wong, G.A. O'Toole, The power of touch: type 4 pili, the von Willebrand A domain, and surface sensing by *Pseudomonas aeruginosa*, J. Bacteriol. 204 (2022) e0008422, https://doi.org/10.1128/jb.00084-22.

[24] T.A. Springer, Complement and the multifaceted functions of VWA and integrin I domains, Structure 14 (2006) 1611–1616, https://doi.org/10.1016/j.str.2006.10.001.

[25] H.H. Tuson, D.B. Weibel, Bacteria–surface interactions, Soft Matter 9 (2013) 4368, https://doi.org/10.1039/c3sm27705d.

[26] Y. Nguyen, S. Sugiman-Marangos, H. Harvey, et al., *Pseudomonas aeruginosa* minor pilins prime type IVa pilus assembly and promote surface display of the PilY1 adhesin, J. Biol. Chem. 290 (2015) 601–611, https://doi.org/10.1074/jbc.M114.616904.

[27] B.-J. Laventie, U. Jenal, Surface sensing and adaptation in bacteria, Annu. Rev. Microbiol. 74 (2020) 735–760, https://doi.org/10.1146/annurev-micro-012120-063427.

[28] S.L. Kuchma, A.E. Ballok, J.H. Merritt, et al., Cyclic-di-GMP-mediated repression of swarming motility by *Pseudomonas aeruginosa*: the pilY1 gene and its impact on surface-associated behaviors, J. Bacteriol. 192 (2010) 2950–2964, https://doi.org/10.1128/JB.01642-09.

[29] L.F. Cruz, J.K. Parker, P.A. Cobine, L. de la Fuente, Calcium-enhanced twitching motility in *Xylella fastidiosa* is linked to a single PilY1 homolog, Appl. Environ. Microbiol. 80 (2014) 7176–7185, https://doi.org/10.1128/AEM.02153-14.

[30] J. Hoppe, C.M. Ünal, S. Thiem, et al., PilY1 promotes *Legionella pneumophila* infection of human lung tissue explants and contributes to bacterial adhesion, host cell invasion, and twitching motility, Front. Cell. Infect. Microbiol. 7 (2017), https://doi.org/10.3389/fcimb.2017.00063.

[31] V.A. Marko, S.L.N. Kilmury, L.T. MacNeil, L.L. Burrows, *Pseudomonas aeruginosa* type IV minor pilins and PilY1 regulate virulence by modulating FimS-AlgR activity, PLoS Pathog. 14 (2018) e1007074, https://doi.org/10.1371/journal.ppat.1007074.

[32] M. Herfurth, A. Treuner-Lange, T. Glatter, et al., A noncanonical cytochrome c stimulates calcium binding by PilY1 for type IVa pili formation, Proc. Natl. Acad. Sci. U.S.A. 119 (6) (2022), https://doi.org/10.1073/pnas.2115061119.

[33] S. Xue, R. Mercier, A. Guiseppi, et al., The differential expression of PilY1 proteins by the HsfBA phosphorelay allows twitching motility in the absence of exopolysaccharides, PLoS Genet. 18 (2022) e1010188, https://doi.org/10.1371/journal.pgen.1010188.

[34] K.S. Makarova, E.V. Koonin, S.V. Albers, Diversity and evolution of type IV pili systems in Archaea, Front. Microbiol. 7 (2016) 667, https://doi.org/10.3389/fmicb.2016.00667.

[35] E.V. Koonin, K.S. Makarova, Y.I. Wolf, Evolution of microbial genomics: conceptual shifts over a quarter century, Trends Microbiol. 29 (2021) 582–592, https://doi.org/10.1016/j.tim.2021.01.005.

[36] M. Champdoré, M. Staiano, M. Rossi, S. D'Auria, Proteins from extremophiles as stable tools for advanced biotechnological applications of high social interest, J. R. Soc. Interface 4 (2007) 183–191, https://doi.org/10.1098/rsif.2006.0174.

[37] C. Brininger, S. Spradlin, L. Cobani, C. Evilia, The more adaptive to change, the more likely you are to survive: protein adaptation in extremophiles, Semin. Cell Dev. Biol. 84 (2018) 158–169, https://doi.org/10.1016/j.semcdb.2017.12.016.

[38] A. Kumar, A. Alam, D. Tripathi, M. Rani, H. Khatoon, S. Pandey, N.Z. Ehtesham, S. E. Hasnain, Protein adaptations in extremophiles: an insight into extremophilic connection of mycobacterial proteome, Semin. Cell Dev. Biol. 84 (2018) 147–157, https://doi.org/10.1016/j.semcdb.2018.01.003.

[39] A.S. Panja, S. Maiti, B. Bandyopadhyay, Protein stability governed by its structural plasticity is inferred by physicochemical factors and salt bridges, Sci. Rep. 10 (2020) 1–9, https://doi.org/10.1038/s41598-020-58825-7.

[40] D. Piovesan, I. Walsh, G. Minervini, S.C.E. Tosatto, FELLS: fast estimator of latent local structure, Bioinformatics 33 (2017) 1889–1891, https://doi.org/10.1093/bioinformatics/btx085.

[41] B. Xue, R.L. Dunbrack, R.W. Williams, et al., PONDR-FIT: a meta-predictor of intrinsically disordered amino acids, Biochim. Biophys. Acta, Proteins Proteomics 1804 (2010) 996–1010, https://doi.org/10.1016/j.bbapap.2010.01.011.

[42] J. Jumper, R. Evans, A. Pritzel, et al., Highly accurate protein structure prediction with AlphaFold, Nature 596 (2021) 583–589, https://doi.org/10.1038/s41586-021-03819-2.

[43] M. Mirdita, K. Schütze, Y. Moriwaki, et al., ColabFold: making protein folding accessible to all, Nat. Methods 19 (2022) 679–682, https://doi.org/10.1038/s41592-022-01488-1.

[44] E.F. Pettersen, T.D. Goddard, C.C. Huang, et al., UCSF Chimera—a visualization system for exploratory research and analysis, J. Comput. Chem. 25 (2004) 1605–1612, https://doi.org/10.1002/jcc.20084.

[45] M. van Kempen, S.S. Kim, C. Tumescheit, et al., Fast and accurate protein structure search with Foldseek, Nat. Biotechnol. 42 (2023) 243 - 246, doi:10.1038/s41587-023-01773-0.

[46] F. Teufel, J.J. Almagro Armenteros, A.R. Johansen, et al., SignalP 6.0 predicts all five types of signal peptides using protein language models, Nat. Biotechnol. 40 (2022) 1023–1025, https://doi.org/10.1038/s41587-021-01156-3.

[47] C.H. Lu, C.C. Chen, C.S. Yu, et al., MIB2: metal ion-binding site prediction and modeling server, Bioinformatics 38 (2022) 4428–4429, https://doi.org/10.1093/bioinformatics/btac534.

[48] D. Szklarczyk, A.L. Gable, K.C. Nastou, D. Lyon, R. Kirsch, S. Pyysalo, N. T. Doncheva, M. Legeay, T. Fang, P. Bork, L.J. Jensen, C. von Mering, The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets, Nucleic Acids Res. 8 (2021) D605–D612, https://doi.org/10.1093/nar/gkaa1074.
[49] R. Sanchez-Garcia, C.O.S. Sorzano, J.M. Carazo, J. Segura, BIPSPI: a method for the prediction of partner-specific protein-protein interfaces, Bioinformatics 35 (2019) 470–477, https://doi.org/10.1093/bioinformatics/bty647.
[50] K. Tamura, G. Stecher, S. Kumar, MEGA11: molecular evolutionary genetics analysis version 11, Mol. Biol. Evol. 38 (2021) 3022–3027, https://doi.org/10.1093/molbev/msab120.
[51] E. Jurrus, D. Engel, K. Star, et al., Improvements to the APBS biomolecular solvation software suite, Protein Sci. 27 (2018) 112–128, https://doi.org/10.1002/pro.3280.
[52] M.J. Abraham, T. Murtola, R. Schulz, et al., GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers, SoftwareX 1–2 (2015) 19–25, https://doi.org/10.1016/j.softx.2015.06.001.
[53] E.D. Drew, R.W. Janes, PDBMD2CD: providing predicted protein circular dichroism spectra from multiple molecular dynamics-generated protein structures, Nucleic Acids Res. 48 (2020) W17–W24, https://doi.org/10.1093/nar/gkaa296.
[54] N. Oganesyan, S.H. Kim, R. Kim, On-column protein refolding for crystallization, J. Struct. Funct. Genom. 6 (2005) 177–182, https://doi.org/10.1007/s10969-005-2827-3.
[55] E. Gasteiger, C. Hoogland, A. Gattiker, et al., Protein identification and analysis tools on the ExPASy server, in: J.M. Walker (Ed.), The Proteomics Protocols Handbook, Springer, Berlin, 2005, pp. 571–607, https://doi.org/10.1385/1-59259-890-0:571.
[56] I. Chaudhuri, J. Söding, A.N. Lupas, Evolution of the β-propeller fold, Proteins 71 (2008) 795–803, https://doi.org/10.1002/prot.21764.
[57] V.J. Goosens, A. Busch, M. Georgiadou, M. Castagnini, K.T. Forest, G. Waksman, V. Pelicic, Reconstitution of a minimal machinery capable of assembling periplasmic type IV pili, Proc. Natl. Acad. Sci. U.S.A. 114 (2017) E4978–E4986, https://doi.org/10.1073/pnas.1618539114.
[58] J.K. Parker, L.F. Cruz, M.R. Evans, L. De La Fuente, Presence of calcium-binding motifs in PilY1 homologs correlates with Ca-mediated twitching motility and evolutionary history across diverse bacteria, FEMS Microbiol. Lett. 362 (2015) 1–9, https://doi.org/10.1093/femsle/fnu063.
[59] E.A. Porsch, M.D.L. Johnson, A.D. Broadnax, et al., Calcium binding properties of the *Kingella kingae* PilC1 and PilC2 proteins have differential effects on type IV pilus-mediated adherence and twitching motility, J. Bacteriol. 195 (2013) 886–895, https://doi.org/10.1128/jb.02186-12.
[60] S. Lewenza, J.L. Gardy, F.S.L. Brinkman, R.E.W. Hancock, Genome-wide identification of *Pseudomonas aeruginosa* exported proteins using a consensus computational strategy combined with a laboratory-based PhoA fusion screen, Genome Res. 15 (2005) 321–329, https://doi.org/10.1101/gr.3257305.
[61] C.A. Whittaker, R.O. Hynes, Distribution and evolution of von Willebrand/integrin A domains: Widely dispersed domains with roles in cell adhesion and elsewhere, Mol. Biol. Cell 13 (2002) 3369–3387, https://doi.org/10.1091/mbc.e02-05-0259.
[62] Q. Zhang, Y.F. Zhou, C.Z. Zhang, et al., Structural specializations of A2, a force-sensing domain in the ultralarge vascular protein von Willebrand factor, Proc. Natl. Acad. Sci. U.S.A. 106 (2009) 9226–9231, https://doi.org/10.1073/pnas.0903679106.
[63] L. Hoffmann, K. Anders, L.F. Bischof, et al., Structure and interactions of the archaeal motility repression module ArnA-ArnB that modulates archaellum gene expression in *Sulfolobus acidocaldarius*, J. Biol. Chem. 294 (2019) 7460–7471, https://doi.org/10.1074/jbc.RA119.007709.
[64] N. Qin, H. Sun, M. Lu, et al., A single von Willebrand factor C-domain protein acts as an extracellular pattern-recognition receptor in the river prawn *Macrobrachium nipponense*, J. Biol. Chem. 295 (2020) 10468–10477, https://doi.org/10.1074/jbc.RA120.013270.
[65] M. Steinert, I. Ramming, S. Bergmann, Impact of Von Willebrand factor on bacterial pathogenesis, Front. Med. 7 (2020) 543, https://doi.org/10.3389/fmed.2020.00543.
[66] M.G. Rossmann, D. Moras, K.W. Olsen, Chemical and biological evolution of a nucleotide-binding protein, Nature 250 (1974) 194–199, https://doi.org/10.1038/250194a0.
[67] I. Hanukoglu, Proteopedia: rossmann fold: a beta-alpha-beta fold at dinucleotide binding sites: rossmann Fold in FAD, NAD and NADP Binding Domains, Biochem. Mol. Biol. Educ. 43 (2015) 206–209, https://doi.org/10.1002/bmb.20849.
[68] C. Cantí, M. Nieto-Rostro, I. Foucault, et al., The metal-ion-dependent adhesion site in the Von Willebrand factor-A domain of $\alpha_2\delta$ subunits is key to trafficking voltage-gated Ca^{2+} channels, Proc. Natl. Acad. Sci. U.S.A. 102 (2005) 11230–11235, https://doi.org/10.1073/pnas.0504183102.
[69] A.A. Bhattacharya, M.L. Lupher Jr., D.E. Staunton, R.C. Liddington, Crystal structure of the A domain from complement factor B reveals an integrin-like open conformation, Structure 12 (2004) 371–378, https://doi.org/10.1016/j.str.2004.02.012.
[70] Qin Wang, Y. Chen, S. Li, et al., Ca^{2+}-based allosteric switches and shape shifting in RGLG1 VWA domain, Comput. Struct. Biotechnol. J. 18 (2020) 821–833, https://doi.org/10.1016/j.csbj.2020.03.023.
[71] D.T. Jones, W.R. Taylor, J.M. Thornton, The rapid generation of mutation data matrices from protein sequences, Comput. Appl. Biosci. 8 (1992) 275–282, https://doi.org/10.1093/bioinformatics/8.3.275.
[72] C. He, K. Ohnishi, Efficient renaturation of inclusion body proteins denatured by SDS, Biochem. Biophys. Res. Commun. 490 (2017) 1250–1253, https://doi.org/10.1016/j.bbrc.2017.07.003.
[73] H. Tao, W. Liu, B.N. Simmons, et al., Purifying natively folded proteins from inclusion bodies using sarkosyl, Triton X-100, and CHAPS, Biotechniques 48 (2010) 61–64, https://doi.org/10.2144/000113304.