Prevotella intermedia produces two proteins homologous to *Porphyromonas gingivalis* HmuY but with different heme coordination mode

Marcin Bielecki[1,#], Svetlana Antonyuk[2,#], Richard W. Strange[3,#], Klaudia Siemińska[1,#], John W. Smalley[4], Paweł Mackiewicz[1], Michał Śmiga[1], Megan Cowan[3], Michael J. Capper[5], Paulina Ślęzak[1], Mariusz Olczak[1], Teresa Olczak[1,*]

P. gingivalis* HmuY homologs produced by *P. intermedia

[1]Faculty of Biotechnology, University of Wrocław, 14A F. Joliot-Curie St., 50-383 Wrocław, Poland;

[2]Institute of Integrative Biology, University of Liverpool, Crown St., Liverpool L69 7ZB, UK;

[3]School of Life Sciences, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, UK;

[4]School of Dentistry, Institute of Clinical Sciences, University of Liverpool, Daulby St., Liverpool L69 3GN, UK;

[5]Department of Pharmacological Sciences, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA

[#]These authors contributed equally to this study and share the first authorship.

*Corresponding author: Teresa Olczak: Faculty of Biotechnology, University of Wrocław, F. Joliot-Curie 14A St., 50-383 Wrocław, Poland; e-mail: teresa.olczak@uwr.edu.pl; Tel. (+48) 71 3752 612.

Abstract

As part of the infective process, *Porphyromonas gingivalis* must acquire heme which is indispensable for life and enables the microorganism to survive and multiply at the infection site. This oral pathogenic bacterium uses a newly discovered novel hmu heme uptake system with a leading role played by the HmuY hemophore-like protein, responsible for acquiring heme and increasing virulence of this periodontopathogen. We demonstrated that *Prevotella intermedia* produces two HmuY homologs, termed PinO and PinA. Both proteins were produced at higher mRNA and protein levels when the bacterium grew under low-iron/heme conditions. PinO and PinA bound heme, but preferentially under reducing conditions, and in a manner different to that of the *P. gingivalis* HmuY. The analysis of the three-dimensional structures confirmed differences between apo-PinO and apo-HmuY, mainly in the fold forming the heme-binding pocket. Instead of two histidine residues coordinating heme iron in *P. gingivalis* HmuY, PinO and PinA could use one methionine residue to fulfil this function, with potential support of additional methionine residue/s. The *P. intermedia* proteins sequestered heme only from the host albumin-heme complex under reducing conditions. Our findings suggest that HmuY-like family might comprise proteins subjected during evolution to significant diversification, resulting in different heme coordination mode. The newer data presented in this manuscript on HmuY homologs produced by *P. intermedia* sheds more light on the novel mechanism of heme uptake, could be helpful in discovering their biological function, and in developing novel therapeutic approaches.

Keywords: *Porphyromonas gingivalis*, *Prevotella intermedia*, HmuY, heme, hemophore, periodontal disease, phylogenetics, evolution, protein structure

Introduction

Periodontal diseases belong to a group of infectious disorders, which are caused by an ecological shift in the microbial composition in the oral cavity and a subsequent exaggerated host immune response [1]. The resulting inflammation leads to bleeding, destruction of tooth-supporting tissues, and tooth loss. The most abundant bacterial species isolated from subgingival samples associated with the clinical features of chronic periodontitis are characterized by the presence of microorganisms belonging to the so-called 'red complex', *i.e.*, *Porphyromonas gingivalis*, *Tannerella forsythia* and *Treponema denticola* [2]. Other bacteria, such as *Prevotella intermedia*, a member of the 'orange complex', serve as early dental plaque colonizers and bridging species with members of the 'red complex' [3,4]. Although *P. gingivalis* is considered the main etiologic agent and a key pathogen responsible for initiation and progression of chronic periodontitis [5-7], an important role has also been ascribed to *P. intermedia*, in that it may aid in creation of an environment which allows *P. gingivalis* to colonize subgingival plaque. A tight association between *P. gingivalis* and *P. intermedia* in the gingival crevice confirms this assumption and suggests some degree of mutualism between these microorganisms [3,8,9].

P. gingivalis and *P. intermedia* belong to anaerobic, black-pigmented, Gram-negative rod-like species, which preferentially utilize proteins and peptides as growth substrates [10]. Both species are unable to synthesize protoporphyrin IX (PPIX) and therefore require heme for survival and ability to establish an infection. Among the best characterized heme acquisition systems of *P. gingivalis* is that encoded by the *hmu* operon, comprising HmuR - a TonB-dependent receptor involved in heme transport through the *P. gingivalis* outer membrane [11-14], HmuY - a heme-binding hemophore-like protein [15-18], and four, so far, uncharacterized proteins. Our

crystallographic studies have revealed a unique protein structure and properties of the *P. gingivalis* HmuY protein [17,18]. HmuY efficiently binds Fe(III)- and Fe(II)heme resulting in iron being in a low-spin Fe(III)/Fe(II) hexa-coordinate environment in the protein, with His134 and His166 coordinating the heme iron [15,16,17]. In contrast to *P. gingivalis*, the heme acquisition mechanisms of *P. intermedia* are less well characterized [19-25]. It has been demonstrated that free heme *in vivo* is not the direct substrate for periodontopathgens; however, it may be derived from hemoglobin by *P. gingivalis* and *P. intermedia* through the actions of hemolysins and proteases, or captured directly by HmuY protein from hemoglobin, as well as from the serum albumin-heme complex and hemopexin [8,18,19,21,25-30].

Our previous analyses identified homologs of *P. gingivalis* HmuY in many other bacteria and demonstrated great variation in their amino acid sequences, which may result from the rapid substitution rate of these sequences and may be associated with various functions fulfilled by these proteins [18,31,32]. Although the amino acid sequences of HmuY homologs identified in phylum Bacteroidetes turned out to be quite similar, significant differences exist in amino acid residues in regions corresponding to the heme-binding pocket of HmuY [18,32]. Importantly, a HmuY homolog produced by *T. forsythia*, termed Tfo, binds heme preferentially under reducing conditions, potentially using methionine residues instead of histidine residues to coordinate heme iron, in contrast to HmuY [17,18].

The presence of two potential homologous genes in *P. intermedia* suggests some functional differentiation of their products, and it is therefore important to characterize the evolution and function of these proteins. We have already overexpressed and purified *P. intermedia* proteins, termed PinO and PinA [32]. The work described here extends our recently published findings on

heme acquisition mechanisms in periodontal pathogens, and presents a detailed characterization of the HmuY homologs produced by *P. intermedia*.

Materials and methods

Bacterial stains and growth conditions

P. gingivalis A7436 and *P. intermedia* 17 were grown anaerobically at 37°C for 5 days on blood agar plates composed of Schaedler broth (containing hemin and L-cysteine), and supplemented with 5% sheep blood and menadione (Biomaxima). These cultures were inoculated into liquid basal medium (BM) prepared of 3% trypticase soy broth (Becton Dickinson), 0.5% yeast extract (Biomaxima), 0.5 mg/l menadione (Fluka), and 0.05% L-cysteine (Sigma). To grow bacteria under high-iron/heme conditions (Hm), this medium was supplemented with 7.7 μM hemin (Fluka), and to grow bacteria under low-iron/heme conditions (DIP), hemin was not added and iron was chelated by addition of 160 μM 2,2-dipyridyl. *Escherichia coli* ER2566 (New England Biolabs) and Rosetta (DE3) (Novagen) strains were cultured under aerobic conditions [18].

Site-directed mutagenesis, overexpression and purification of proteins

The *P. gingivalis* A7436 HmuY protein (NCBI accession no. CAM 31898) was overexpressed using a pHmuY11 plasmid and *E. coli* ER2566 cells (New England Biolabs) and purified from a soluble fraction of the *E. coli* cell lysate as described previously [15,32]. The purified HmuY lacked the signal peptide (MKKIIFSALCALPLIVSLTSC) and additional (GKKK) N-terminal amino acid residues present in the full length protein sequence [17]). *P. intermedia* 17 PinO (NCBI accession no. AFJ07542) and PinA (NCBI accession no. AFJ08449) proteins, lacking the predicted signal peptides (MKTKIFAVACLATLLFTSC and MKFKSFMALSCLTVLLFSSC, respectively) were overexpressed in *E. coli* Rosetta (DE3) cells and purified from a soluble

fraction obtained from *E. coli* cell lysate as reported previously [32]. To crystallize truncated PinO version, the protein lacking the signal peptide (MKTKIFAVACLATLLFTSCS) and additional N-terminal (SKDNNDDPNPKPE) amino acid residues present in the full length protein sequence was overexpressed and purified, using the expression plasmid constructed as described previously for the pHmuY11 plasmid [15] and primers listed in Supplementary Table S1.

Point mutations were introduced using a QuikChange II XL Site-Directed Mutagenesis Kit (Stratagene). Briefly, modified pMAL-c5x_His plasmid constructed in our previous study [33] was used to clone DNA sequences encoding PinO and PinA proteins (all primers are listed in Supplementary Table S1). Selected amino acids with a potential ability to coordinate heme iron were substituted by an alanine, resulting in single or double mutations. Recombinant proteins possessing N-terminal His tag and maltose-binding protein (MBP) were purified using affinity chromatography with amylose-agarose resin, and after Factor Xa cleavage, affinity chromatography with nickel-agarose resin as reported previously [33].

Concentrations of apo- and holo-HmuY were determined spectrophotometrically using the empirical molar absorption coefficients (ε_{280}) 36.86 and 59.26 $mM^{-1}cm^{-1}$, respectively [18]. Empirical molar absorption coefficients determined for PinO was 23.81, for truncated PinO 23.87, and for PinA 45.81 $mM^{-1}cm^{-1}$.

Heme-protein complex formation

Heme (hemin chloride; ICN Biomedicals) solutions were prepared as reported previously [16,34]. Formation of the heme-protein complexes was examined in 100 mM Tris–HCl buffer, pH 7.5, containing 140 mM NaCl (TBS) or in 20 mM sodium phosphate buffer, pH 7.4 or 7.6, containing 140 mM NaCl (PBS). UV-visible spectra were recorded in the range 250-700 nm

with a single beam Ultrospec 2000 spectrophotometer (Biochrom Ltd) or a double beam Jasco V-650 spectrophotometer using cuvettes with 10 or 2 mm path length, respectively. Titration curves were analyzed using the equation for a one-site binding model, and dissociation constant (K_d) values determined as reported earlier [18] using OriginPro 8 software (OriginPro Corporation). To analyze the redox properties of the heme iron, 10 mM sodium dithionite was used as the reductant with or without mineral oil overlay, and potassium ferricyanide as the oxidant [16,35].

Circular dichroism (CD) and magnetic circular dichroism (MCD) spectroscopies

For CD and MCD analyses, heme-protein complexes were formed as described previously [18]. Briefly, the protein concentration was adjusted to 10 μM (for far-UV CD), 100 μM (for CD in the visible region) or 40 μM (for MCD in the visible region). CD spectra were recorded at 200-260 nm (far-UV CD) or 340-660 nm (CD in the visible region) at 25^{o}C using a Jasco J-715 or J-810 spectropolarimeter with a scan speed 50 nm/min, response time 2 s and a slit width of 1.0 nm. MCD spectra were recorded in the visible region at 25^{o}C using a Jasco J-715 spectropolarimeter equipped with an electromagnet generating a magnetic field of 1.46 T, with a scan speed 200 nm/min, response time 2 s and a slit width of 1.0 nm. Measurements were made using a quartz cell with a 2 mm path length. Mean spectra were calculated from five independently recorded data sets. UV-visible absorbance spectra were recorded before and after CD and MCD measurements to verify protein integrity.

Temperature-induced unfolding experiments of PinO protein were performed using Jasco J-1500 spectropolarimeter. Protein samples (2 μM) were prepared in 10 mM sodium phosphate buffer, pH 7.6, and examined in a 10-mm quartz cuvette with magnetic stirrer bar. Thermal

denaturation was carried out between 25 and 90°C at intervals of 5°C, and CD spectra were recorded between 300 and 210 nm.

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and Western blotting

For SDS-PAGE, samples were prepared and analyzed as reported previously [18]. For immunoblotting, samples were separated by SDS-PAGE, transferred onto nitrocellulose membrane (Millipore), probed with rabbit anti-PinO or anti-PinA antibodies, and complexes formed were detected using HRP-conjugated anti-rabbit IgG antibodies (Sigma) and chemiluminescence staining (Perkin Elmer) [32].

Proteolytic digestion

HmuY, PinO and PinA were subjected to proteolysis by trypsin digestion [17]. Briefly, two reactions with proteins in 100 mM Tris-HCl, 20 mM $CaCl_2$, pH 8.0 at a 1:50 protease:substrate molar ratio were conducted in the presence (1:1 molar ratio) or absence of heme. Fresh portions of trypsin were added every 1 h during the first 4 h, and then after 12 h. Protein samples previously subjected to thermal denaturation (95°C, 10 min) were also assayed for their proteolytic susceptibility.

To examine the susceptibility of the PinO and PinA to *P. gingivalis* or *P. intermedia* proteases, *P. gingivalis* and *P. intermedia* cells were grown under rich, high-iron/heme conditions, ensuring proper cell viability and efficient proteolytic activity [15,32], in the presence of added purified HmuY, PinO or PinA proteins at final 1 μM concentration [18]. All cultures (10 ml) were started at OD_{600}=0.2 (time 0), grown and collected at 6 and 24 h (*P. gingivalis*) or 24, 48 and 72 h (*P. intermedia*). The number of bacterial cells at the starting point was ~2×10^8per ml of the culture medium, and increasing during cultivation. As controls, *P.*

gingivalis or *P. intermedia* cultures without addition of the purified proteins were analyzed. At the indicated time points, aliquots of samples were examined by SDS-PAGE and CBB staining or by Western blotting using anti-PinO or anti-PinA antibodies [32].

Cross-linking and analytical size-exclusion chromatography

Cross-linking of proteins in apo- and holo-forms was carried out using 0.1, 0.2, 0.5 and 1.0% glutaraldehyde (Sigma) or 0.1, 1.0, 2.0, and 5.0% formaldehyde (Sigma) for 1 h at 37°C. In addition, 3-maleimidobenzoic acid *N*-hydroxysuccinimide ester (Sigma), which is a heterobifunctional cross-linking reagent reactive toward primary amine and sulfhydryl groups and dimethyl suberimidate dihydrochloride (Fluka), which is a lysine-specific cross-linker for the modification of proteins *via* amidation were used according to the respective protocols. Cross-linking products were examined by SDS-PAGE and CBB staining [18,32].

To analyze the oligomeric state of the native proteins, samples of HmuY, PinO, and PinA (~0.22 mg; 100 µl) in PBS or 200 mM Tris-HCl buffer, containing 140 mM NaCl, pH 8 (under oxidizing or reducing conditions, respectively) were applied in apo- or holo-form onto Superdex 75 Increase 10/300 GL (GE-Healthcare) or a ProteoSEC 11/30 3-70 HR (Protein Ark) columns, respectively, connected to an AKTA Pure FPLC system (GE Healthcare). To analyze proteins under reducing conditions, 30 mM sodium dithionite was added to the separating buffer. Chromatography was carried out with 0.8 ml/min flow rate. Both columns were calibrated using Gel Filtration Markers Kit for Protein Molecular Weights 6,500-66,000 Da (Sigma-Aldrich)

Crystallization

The P212121 crystal form of PinO was obtained by the hanging drop method, with 2 µl of 7 mg/ml protein solution in 50 mM Tris-HCl (pH 8), 50 mM NaCl and 2 µl of reservoir solution containing 25% PEG MED SMEAR (Molecular dimensions) equilibrated over 400 µl of

reservoir solution. Long needle-like crystals grew at room temperature within 3 weeks. Crystals were flash frozen in the reservoir solution supplemented with 20% glycerol, as a cryo-protectant.

X-ray data collection, processing and structure determination

X-ray data were collected at the IO4-1 beamline (Diamond Synchrotron, UK) using a Pilatus 6M-F detector and 0.91587 Å wavelength, to 2.46 Å resolution at 100 K. Data were integrated in XDS [36] and scaled in AutoPROC [37]. The structure was solved by Molecular replacement using MOLREP [38] with apo-Tfo monomer (PDB ID: 6EU8) as the starting model and automatically built by ArpWarp [39], followed by refinement by REFMAC5 [40]. Refinement was iterated with manual model building in COOT [41], NCS restrains were used during refinement. The quality of the model was assessed in MOLPROBITY [42]. Data collection and refinement statistics are shown in Supplementary Table S2.

MD simulations

Molecular dynamics (MD) simulations were performed with Gromacs 2018 [43] using the Gromos54A7 force field [44] on (i) the apo-PinO crystal structure (PDB ID: 6R2H), (ii) two models with Fe-heme added manually to Met119 (Met150; numbering of amino acid residues according to the protein version crystallized and examined by X-ray analysis or theoretically modelled is shown, along with numbering corresponding to the full length protein sequence) of apo-PinO, differing in their initial heme orientations, (iii) a snapshot taken from the MD trajectory from (i) with Fe-heme added manually to Met119. The structures were prepared for MD runs in each case by adding hydrogen atoms and assigning partial charges to protein residues. The total charge of the Fe-heme group was -2. Following solvation with SPC water [45] and charge neutralization with Cl^- ions, the simulation box contained ~25,000 atoms. Following energy minimization employing protein restraints, the system was equilibrated under

microcanonical ensemble at a temperature of 310 K for 200 ps and then switched to the isothermal-isobaric ensemble using the Parrinello-Rahman barostat [46] at 1 atm pressure. The system was further equilibrated for 200 ps. Production runs using a time-step of 2 fs, with several replicates for each model, were then made for 100 ns. The Particle-Mesh-Ewald (PME) sum method [47] was used for all electrostatic calculations with a cutoff distance of 1.0 nm. MD trajectories were examined using the VMD program [48].

Heme sequestration experiments

The albumin-heme complex was prepared by incubating a 120 μM stock solution of human albumin (Sigma; A-8763) in TBS at 37°C with heme at a 1:0.9 protein to heme molar ratio to ensure that no uncomplexed heme remained [8]. Human hemopexin (Sigma; H-9291) was solubilized in TBS and incubated with heme as described above. Bovine methemoglobin (MP Biomedicals; 151234) was solubilized in TBS. The hemoproteins were incubated with HmuY, PinO or PinA as described previously [34].

Co-incubation of HmuY with PinO or PinA was carried out in PBS at 37°C and monitored by UV-visible spectroscopy under oxidizing (pH 7.6 and 6) or reducing (pH 7.6) conditions using each protein at 5 μM concentration [16].

Bacterial cell fractionation

Portions of bacterial cultures were centrifuged at 20,000×g for 30 min at 4°C and supernatants were filtered using sterile 0.22 μm filters (Roth) to separate the cell-free culture supernatant and cells. The cell pellets were washed twice with PBS and used to analyze the whole cell fraction. To separate outer-membrane vesicles (OMVs), the filtered culture supernatant was centrifuged at 100,000×g for 2 h at 4°C using a Beckman fixed-angle rotor (Type 70 Ti), and pelleted

membrane fractions were re-suspended in PBS. After ultracentrifugation, the supernatant was concentrated 25-fold using Amicon Ultra-4 Centrifugal Ultracel-10K filter units (Millipore).

Quantitative reverse transcriptase-polymerase chain reaction (RT-qPCR)

RNA was purified from 0.5×10^8-4×10^8 *P. gingivalis* or *P. intermedia* cells as described previously [18]. Reverse transcription was carried out using 1 μg of RNA using a SensiFAST cDNA Synthesis Kit (Bioline). PCR was preformed using SensiFAST SYBR No-ROX Kit (Bioline) and the LightCycler 96 System (Roche). The amplification reaction was carried out as follows: an initial denaturation at 95°C for 2 min, 40 cycles of denaturation at 95°C for 5 s, primer annealing at 60°C for 10 s, and extension at 72°C for 20 s. The melting curves were analyzed to monitor the quality of PCR products. Relative quantification of *hmuY*, *pinO*, and *pinA* genes was determined in comparison to the *16S rRNA* gene of *P. gingivalis* (gene ID: 2552647) and *P. intermedia* (locus tag: PIN17_RS05870) as references, using the $\Delta\Delta C_t$ method. All samples and controls were run in triplicate in three independent experiments for the target and reference genes. All primers used in this study are listed in Supplementary Table S1.

Selection of homologs and phylogenetic analyses

PinA and PinO amino acid sequences from *P. intermedia* were compared with other homologs of the previously functionally characterized HmuY protein from *P. gingivalis*. The similar sequences were found in GenBank database using sensitive searches based on PSI-BLAST [49] in three iterations assuming *E-value* < 0.005. From this set of 3540 potential homologs, we selected sequences that revealed the presence of HmuY domains (316577, 213031 and 213030) with *E-value* < 0.01 in searches of Conserved Domain Database [50] using rpsBLAST. After elimination of redundant and fragmentary sequences, the multiple sequence alignment was performed in MAFFT using slow and accurate algorithm L-INS-i with 1,000 cycles of iterative

refinement [51]. The alignment was edited manually in JalView [52]. The final global set consisted of 1292 sequences representing various prokaryotic phylogenetic lineages. Moreover, we extracted all 369 sequences assigned to Bacteroidia group, to which *Porphyromonas* and *Prevotella* belong. These sequences were aligned using accurate algorithm combining sequence information with protein structures and profiles [53].

Phylogenetic trees were reconstructed using Bayesian method in MrBayes [54] and PhyloBayes [55], as well as maximum likelihood approach in IQ-TREE [56] and morePhyML [57] based on PhyML [58].

In MrBayes analyses, we applied mixed+I+Γ(5) models rather than fixed ones to specify appropriate substitution models across the large parameter space [59]. Two independent runs starting from random trees were applied, each using 72 or 8 Markov chains for the global and Bacteroidia sets, respectively. The trees were sampled every 100 generations for 20,000,000 generations. After reaching the stationary phase and convergence, i.e. when the standard deviation of split frequencies stabilized and was much below the assumed threshold 0.01, the last 110,421 or 45,431 trees, depending on the data set, were selected to create the final posterior consensus trees.

The PhyloBayes analysis was also carried out for the two alignment sets but convergence was reached only for the Bacteroidia set, i.e. the largest discrepancy observed across all bipartitions (maxdiff) was below the recommended threshold 0.1. Therefore, results only for this set were described here. In this case, the WAG+Γ(5) model was applied because it was proposed in ProtTest [60]. Two independent Markov chains were run for 100,000 generations with one tree sampled for each generation. The last 25,000 trees from each chain were collected to compute the posterior consensus tree.

The tree calculated with (more)PhyML was based on the LG+Γ(5) or WAG+I+Γ(5) models, as proposed by ProtTest, for the global and Bacteroidia sets, respectively. We applied the best heuristic search algorithm, nearest neighbour interchanges (NNI) and subtree pruning and re-grafting (SPR). In the case of IQ-TREE, we used LG+R8 or WAG+R7 models for the global and Bacteroidia sets, respectively, as found by ModelFinder [61]. To assess a significance of particular branches in these maximum likelihood approaches, we carried out a non-parametric bootstrap analysis on 100 or 1000 replicates for the global and Bacteroidia sets, respectively. Moreover, we applied the approximate likelihood ratio test (aLRT) based on a Shimodaira-Hasegawa-like procedure [62], assuming 1000 or 10,000 replicates in IQ-TREE for the global and Bacteroidia sets, respectively. Phylogenetic trees were edited in FigTree [63] and TreeGraph [64].

Results

Phylogenetic analysis of HmuY family proteins

Sensitive homology searches enabled us to identify distant homologs of HmuY and infer their phylogenetic relationships (Supplementary Figure S1). The majority of these homologs belong to the phylum Bacteroidetes and related phyla, *i.e.*, Chlorobi, Ignavibacteriae and Balneolaeota. Among the homologs, there are also representatives of Fibrobacteres and Gemmatimonadetes, which are classified to the FCB group (consisting Bacteroidetes among other bacteria), together with the three previously mentioned phyla. Quite numerous are sequences belonging to α-, β-, γ- and δ-Proteobacteria as well as Spirochaetes. Many sequences assigned to given phyla or classes are not always clustered together into monophyletic clades (Supplementary Figure S1). Also various groups of Bacteroidetes are separated into various clades. Such a distribution suggests

that ancestral *hmuY*-like genes were subjected to ancient duplications before divergence of the main Bacteroidetes groups. Alternatively, horizontal gene transfers could occur between these lineages. Sequences assigned to *Prevotella* are distributed into six main clades among other Bacteroidia representatives (Supplementary Figure S1).

In order to analyze in detail the phylogenetic relationships between Bacteroidia, we constructed trees including only their sequences (Figure 1, Supplementary Figure S2). Sequences assigned to various genera were clustered in the obtained phylogenetic trees in many small monophyletic clades, which were separated from each other. Such distribution suggests that numerous duplications occurred before divergence of these genera. We identified in total 150 sequences from *Prevotella* among which 131 were unique. These sequences were distributed into eight well-supported clades. Eight *Prevotella* species (*P. denticola*, *P. enoeca*, *P. histicola*, *P. ihumii*, *P. micans*, *P. multisaccharivorax*, *P. scopos*, *P. timonensis*) had sequences separated into two clades, five species into three clades (*P. buccalis*, *P. maculosa*, *P. marshii*, *P. melaninogenica*, *P. oralis*) and *P. oris* into four clades. Many of the *Prevotella* species showed additional duplications of *hmuY* homologs within a given clade. One of those is *P. intermedia*, which was separated into two well-supported groups designated Pi1 and Pi2, within the clade 5 (Figures 1 and 2, Supplementary Figure S2). The position of such sequences among other *Prevotella* species indicates a quite recent duplication of these genes. Interestingly, the duplication was associated with the translocation of one copy into the other chromosome because these genes are located on different chromosomes. In the case of the strain 17, one protein, PinO (PIN17_RS00035), is encoded within the *hmu*-like operon on the large chromosome (accession no. NC_017861), while the second protein, PinA (PIN17_RS05355), is encoded separately on the small chromosome (accession no. NC_017860) (Figure 3).

PinO and PinA bind heme but in a manner different to HmuY

We hypothesize that, like the HmuY hemophore-like protein of *P. gingivalis*, its homologs identified in *P. intermedia* may also be engaged in heme acquisition. In contrast to HmuY, which exhibited a Soret λ_{max} at 411 nm, the Soret maxima determined for PinO and PinA were at 406 nm (Figure 4A). Compared to the HmuY Q band maxima (528 and 558 nm), those for PinO were at 499, 530, and 607 nm, and those for PinA were at 495, 530 and 607 nm. The difference spectrum analysis confirmed these findings, although slightly different values of absorbance maxima were observed (Figure 4B). In addition, in the Q band range, an additional maximum was present at 566 nm for the PinA-heme complex. Compared to the HmuY protein, which after overexpression, purification and concentration existed under air (oxidizing) conditions as a red-colored complex, both PinO and PinA heme complexes showed a green color, which were visible as brown-colored complexes after saturation with heme (Figure 5A).

The chemical reduction resulted in red-colored complexes of the PinO- and PinA-heme solutions, very similar to the HmuY-heme solution (Figure 5A). After reduction, the Soret peak maxima of PinO and PinA red shifted and single peaks emerged at 428 and 427, respectively, compared to 424 nm for *P. gingivalis* HmuY (Figure 5B). Moreover, the reduction produced more robust and well-resolved Q bands, which were almost identical to those observed for HmuY. The heme bound to PinO and PinA could be further re-oxidized, resulting in the Soret band shift back to 404 and 406 nm, compared to 411 for *P. gingivalis* HmuY.

The binding stoichiometry of PinO and PinA was 1:1 (Figure 4C). Compared to *P. gingivalis* HmuY and PinO, PinA bound heme with significantly lower ability under oxidizing conditions (Figure 4C). However, reduction resulted in a significantly higher ability of heme binding, with K_d values comparable (PinA) or even lower (PinO) to those observed for HmuY (Figure 4C).

Similar to *P. gingivalis* HmuY, heme binding by these proteins did not cause significant changes in the secondary structure of the proteins, as determined by far-UV CD analysis (Supplementary Figure S3).

The CD spectra of PinO and PinA proteins complexed with heme determined in the visible region were similar to those obtained for Tfo [18], but differed significantly from those observed for *P. gingivalis* HmuY (Figure 6A). The main feature was the lack of a negative Cotton effect in the ferric heme form in the case of both proteins observed under air conditions. Only the reduction of PinA resulted in the minimum similar to that of *P. gingivalis* HmuY. The CD pattern of PinO analyzed under oxidizing or reducing conditions resembled that observed for the HmuY His134Ala variant [16]. The spectrum recorded for PinA under oxidizing conditions was partly similar to the spectrum obtained for the HmuY His166Ala variant [16]. Also, MCD spectra of PinO and PinA proteins complexed with heme where similar to spectra of Tfo, but differed when compared to HmuY spectra (Figures 6B and 6C). However, the spectra were more similar when recorded under reducing conditions, especially in the visible region. Based on the UV-visible, CD, and MCD spectroscopic results, we were not able to define the heme iron coordination mode to PinO and PinA.

Sequence comparison of HmuY homologs

In order to confirm if amino acid residues engaged in heme binding in PinO and PinA are in sites homologous to those in HmuY and Tfo, alignment of their amino acid sequences was performed (Figure 7). The sites with the His134 and His166 residues in HmuY of *P. gingivalis* corresponded to gaps in Tfo and *P. intermedia* homologs. In contrast to that, the two Met sites in Tfo were well conserved because homologous sites in all *P. intermedia* sequences were also occupied by such residues (Met116 and Met145 in the PinO sequence of the solved protein

structure; correspond to Met147 and Met176 in the full length protein sequence). The first Met was also aligned with Met in HmuY, while the second with Ala. Another Met residue potentially coordinating heme iron in PinO (Met119 in the PinO sequence of the solved protein structure; corresponds to Met150 in the full length protein sequence) appeared also to be conserved because it corresponded to Met in PinA homologs and all *T. forsythia* sequences, although it was not proved to participate in heme iron coordination in Tfo. On the other hand, two remaining Met residues potentially binding heme in PinO (Met45 and Met46 in the PinO sequence of the solved protein structure; correspond to Met76 and Met77 in the full length protein sequence) were found only in PinO and its homologs in *P. intermedia* from the group Pi2, while PinA and *P. intermedia* from the group Pi1 had in the homologous sites Ala and Gly, respectively. These residues are a part of an insertion not present in *Porphyromonas* and *Tannerella* amino acid sequences. Their poor conservation suggests that they could be not essential in binding heme.

In order to study how these sites could evolve, we aligned all sequences from the group including *T. forsythia* and *P. intermedia* sequences (Supplementary Figure S4), as well as mapped the relevant sites on the phylogenetic tree (Figure 2). This group consists of two subgroups. One includes ten *Tannerella*, two *Porphyromonas macacae*, seven *Bacteroides*, one *Alloprevotella tannerae*, one *Prevotella* sp., and one unidentified *Prevotellaceae* sequences. The second subgroup clusters only *Prevotella* sequences, including *P. intermedia*. In the first subgroup, only four *T. forsythia* sequences include two Met residues which could be responsible for heme iron coordination. In other sequences at least one of these residues is replaced by another so their proteins most likely do not have abilities to coordinate heme iron. In the second subgroup, 23 out of 31 sequences contain Met residues which are homologous to those in Tfo, and an additional Met residue corresponding to Met119 (Met150) in the PinO sequence of the

412 solved protein structure. Fourteen of these sequences are assigned to taxa other than *P.*
413 *intermedia*. It is possible that they could also coordinate heme iron. However, only six *P.*
414 *intermedia* from the clade Pi2 have additional two Met residues corresponding to Met45 and
415 Met46 (Met76 and Met77) in the PinO sequence of the solved structure. Nevertheless, the
416 presence of Met residues in the homologous positions in many sequences suggests possibility of
417 heme iron coordination. Moreover, many sequences contain, in appropriate positions, amino acid
418 residues (Ile, Leu, Lys and Thr) that can be replaced by Met residue due only to a single
419 mutation at the nucleotide level. Therefore, the proteins clustered in this group are preadapted to
420 bind heme.

421 In a similar way, we compared these sites in other *Prevotella* sequences (Supplementary
422 Figure S5). We found that 11 additional sequences contain Met residues in the homologous
423 positions as the Tfo sequence. Among them, 7 sequences also have an additional Met residue
424 corresponding to Met119 (Met150) in the PinO sequence of the solved structure. These
425 sequences belong to different separated clades. This grouping suggests that the heme-binding
426 properties could evolve independently in various lineages.

427 **Three-dimensional structure of PinO**

428 The results of our phylogenetic and spectroscopic studies did not allow for definite identification
429 of amino acids coordinating iron in heme bound to PinO and PinA, although we demonstrated
430 that both proteins, similar to *T. forsythia* Tfo [18], preferentially bound heme in the ferrous form.
431 Similar to *P. gingivalis* HmuY (Figure 8A), to obtain structural data and identify amino acid
432 residues engaged in heme binding to PinO, we applied crystallographic analysis. The PinO
433 structure in apo-form (PDB ID: 6R2H) was solved to 2.46 Å resolution (Supplementary Table
434 S1, Figure 8B, Supplementary Figure S6). Structurally, the protein is closer to *T. forsythia* apo-

Tfo (PDB ID: 6EU8) (Figure 8C) than to *P. gingivalis* apo-HmuY (PDB ID: 6EWM) and holo-HmuY (PDB ID: 3H8T) (Figure 8A). As in apo-HmuY (Figure 8A), the main part of PinO forms a β barrel composed by two β sheets, one with six antiparallel β strands and another with four antiparallel β strands flanked by two short α helixes (Figure 8B). The possible heme binding cleft is composed of three hairpins, two of which make a β sheet at one side of the heme binding pocket, while the third comprising very short β strands covers the binding cleft. A short helix is located between β-strands 10 and 11. The main difference with the apo-HmuY structure is in the arrangement of its heme binding cleft. Apo-HmuY has only two hairpins, one with two long antiparallel β-strands connected by a short loop, and a second hairpin with strands connected by a longer and flexible loop. The differences between apo-PinO and apo-Tfo are much smaller and mainly occur in the heme binding region, where apo-Tfo has a four stranded antiparallel β sheet with a helical insert involved in heme binding on one side and a hairpin with two short antiparallel β strands on the other (Figure 8C). The overall average r.m.s.d. between main-chain atoms of PinO and Tfo is 0.8 Å (for 121 residues) while it is 1.6 Å when aligned with HmuY (for 96 residues). Unfortunately, we failed to crystallize PinO in its holo-form. So far, we were also not able to crystallize the PinA protein. Therefore, a homology model of apo-PinA is shown in Figure 8D.

Identification of heme-coordinating ligands

To study the effects of specific amino acids on heme iron coordination in PinO and PinA, we systematically replaced several Met residues by an Ala residue and analyzed the ability of the protein variants to bind heme. Such experimental approach was based on amino acid sequence comparisons, available crystallographic data, and theoretical modeling. UV-visible absorbance analysis (Figure 9) and determination of heme dissociation constants (Supplementary Figure S7)

458 demonstrated that Met119 (Met150) in PinO and Met129 (Met162) in PinA could be engaged
459 directly in heme iron coordination. Interestingly, Met45 and Met145 (Met76 and Met176) in
460 PinO, as well as Met158 (Met191) in PinA, may also participate in heme iron coordination.
461 Although differences in K_d values determined under oxidizing conditions were less obvious
462 between wild type proteins (PinO = $9.2\pm3.2\times10^{-8}$ M; PinA = $5.7\pm0.3\times10^{-6}$ M) and site-directed
463 mutagenesis variants (PinO Met119Ala = $1.9\pm0.2\times10^{-6}$ M; PinO Met45Ala/Met145Ala =
464 $3.2\pm0.7\times10^{-7}$ M, PinA Met129Ala = $4.9\pm1.4\times10^{-6}$ M, PinA Met158Ala = $4.0\pm1.7\times10^{-7}$ M),
465 analyses performed under reducing conditions suggested involvement of these amino acids in
466 heme iron coordination (wild type PinO = $3.4\pm1.6\times10^{-9}$ M; PinO Met119Ala = $1.0\pm0.4\times10^{-5}$ M;
467 PinO Met45Ala/Met145Ala = $3.1\pm0.8\times10^{-8}$ M; wild type PinA = $2.0\pm0.5\times10^{-8}$ M; PinA
468 Met129Ala = $1.0\pm0.2\times10^{-5}$ M; PinA Met158Ala = $1.4\pm0.1\times10^{-6}$ M).

469 To verify engagement of amino acids in heme iron coordination in PinO, MD simulations
470 were carried out using two models with heme inserted into the apo-crystal structure at Met119
471 (Met150) (Figure 8E**)**. The models differed in that the heme was rotated about the Fe-Met119
472 bond so that propionate groups were oriented in two different starting positions. For long periods
473 during the resulting MD trajectories, the conformations adopted by the loops containing Met45
474 and Met145 (Met76 and Met176) consistently placed them at the open sixth coordinating
475 position of the heme group. The simulations demonstrated that when the Fe-heme is coordinated
476 by Met119, both Met45 and Met145 have frequent and sufficiently close contact to the Fe ion to
477 be potential ligands (Figure 10A**).** When either Met45 or Met145 was substituted by Ala in
478 simulations the Fe made close contact with the remaining Met45 or Met145 residue (Figure
479 10B). Interestingly, MD simulations of heme binding to the double Met45Ala/Met145Ala variant
480 showed similar behaviour, with Ala45 and Ala145 also occupying positions close to the Fe-heme

(Supplementary Figure S8). The simulations demonstrated that close contact between Fe and Met46 (Met77) is an unlikely event (Supplementary Figure S9). Other amino acids that were found in close neighbourhood to the heme include Lys149 (Lys180) and Asn134 (Asn165), whose sidechains potentially participate in stabilization of the heme orientation through interactions with the heme O-atoms on propionate groups (data not shown).

We also examined a more speculative heme-bound MD starting model, where heme was inserted at Met119 (Met150) to a snapshot from the apo-PinO MD trajectory, selected at an open configuration where Met119 was exposed to the protein surface (Figure 8F). During MD simulations using this model, the Met119 containing loop moved back ~20 Å towards the proposed heme binding pocket loops, enabling Fe-heme to again make contact with the Met45 and Met145 (Met76 and Met176) residues (Supplementary Figure S10). Interestingly, while the mechanism of heme incorporation in PinO is not known, the dynamic properties of the protein revealed by the MD simulations suggest a possible means of binding and assimilating the heme during opening and closing of the pocket loops. These heme-bound PinO MD models thus gave the same consistent result over a number of simulation replicates and for different starting conformations. The final state from one of these simulations is shown in Figure 10C. We previously reported that MD of the apo-Tfo crystal structure resulted in an opening of the heme pocket loops [18]. Here we showed that MD simulations, using a heme-Met149 (Met169) Tfo model constructed from this open configuration (Figure 8G), confirms that closure of the pocket loops occurs also in the case of Tfo.

PinO and PinA exist as monomers and are more susceptible to proteolytic digestion than HmuY

In contrast to HmuY [18] and PinA, where dimers were observed even after SDS-PAGE, PinO usually migrated during electrophoresis as a single band (Supplementary Figure S11A). Cross-linking studies demonstrated that oligomer formation was observed only in the case of formaldehyde and glutaraldehyde cross-linking, with a similar pattern observed for the proteins examined (Supplementary Figure S11A). In contrast, size-exclusion chromatography showed that both PinO and PinA, regardless heme bound and redox conditions, are present in solution in the form of monomers (Supplementary Figure S11B).

Previously, we demonstrated that *P. gingivalis* HmuY is completely resistant to several proteases [8,17,65]. Compared to HmuY, PinO was also quite resistant to trypsin digestion (Figure 11A), even after thermal denaturation, suggesting a fast refolding ability. This finding was confirmed by quantitative analysis of thermal denaturation of the protein (Figure 11B), demonstrating that PinO is very resistant in regard to thermal denaturation as compared with other hemophore-like proteins so far examined. In contrast to HmuY and PinO, PinA was significantly more susceptible to trypsin digestion (Figure 11A). Our observations were further corroborated by experiments demonstrating *P. gingivalis* growth in the presence of the purified proteins. In contrast to HmuY, which was completely resistant during *P. gingivalis* growth, PinA was quickly digested, whereas PinO exhibited a moderate resistance to active *P. gingivalis* proteases (Figure 11C). However, all the proteins examined were not digested when added to *P. intermedia* cultures (Figure 11D).

PinO and PinA sequester heme from the albumin-heme complex under reducing conditions

Our recent studies clearly demonstrated that *P. gingivalis* HmuY efficiently sequestered heme from methemoglobin, heme(FeIII)-albumin [8,34], and heme(FeIII)-hemopexin [18]. Here we showed that PinO and PinA were not able to sequester heme present in methemoglobin or heme

bound to albumin, when it was in the ferric form (data not shown). However, we observed that, similar to *T. forsythia* Tfo [18], both proteins were able to complex heme from heme-albumin under reducing conditions (Figure 12).

To assess any possible syntrophy between *P. gingivalis* HmuY and its homologs from *P. intermedia*, we examined the interactions between apo-HmuY and holo-forms of PinO and PinA. As shown in Figure 13A and 13B, HmuY efficiently sequestered Fe(III)heme which had been complexed to both homologous proteins under oxidizing conditions. A similar effect was observed in the case of *T. forsythia* Tfo [18]. When such interaction was examined under reducing conditions, PinO was able to capture heme bound to HmuY (Figure 13C).

Production of transcripts and proteins of PinO and PinA is increased under low-iron/heme conditions

Similar to *hmuY* mRNA and the HmuY protein, both transcripts (Figure 14A) and proteins (Figure 14B) of PinO and PinA were produced at higher levels by bacteria grown under low-iron/heme conditions (the lack of heme and iron chelation) as compared with heme/iron-rich conditions. The distribution of PinO and PinA proteins between whole *P. intermedia* cells and outer-membrane vesicles was also similar, as compared to *P. gingivalis* HmuY (Figure 14B) and *T. forsythia* Tfo [18]. The only difference was very low level of PinO and the lack of PinA in the form of soluble proteins shed from the outer-membrane into the culture medium (Figure 14B).

Discussion

Our analyses have led us to the finding that the black pigmented anaerobes, *P. gingivalis* and *P. intermedia*, display a novel heme acquisition mechanism, whereby oxyhemoglobin is firstly oxidized to methemoglobin and heme is released due to the relaxation of the Fe(III)heme binding

affinity of globin [28,66]. In the case of *P. gingivalis*, generation of methemoglobin involves the arginine-specific gingipain protease A (HRgpA) [28,34,66], and the cysteine protease interpain A (InpA) in the case of *P. intermedia* [8,29]. Our previous work on *P. gingivalis* has also resulted in extensive characterization of a novel, unique hmu heme acquisition mechanism, with a leading role played by a HmuY protein. We demonstrated that *P. gingivalis* HmuY not only binds free heme, but can capture heme from methemoglobin directly [34] and thus functions similarly to classical secreted hemophores, proteins involved in heme transfer from the host hemoproteins to the bacterial outer-membrane receptors [67,68]. This process is facilitated by *P. gingivalis* and *P. intermedia* proteases [8,34]. Importantly, HmuY is also able to compete with albumin, which is the normal front-line heme scavenger *in vivo* [34], as well as acquiring heme from serum hemopexin [18]. Our data suggest that HmuY may be important for effective heme acquisition, especially in iron/heme-limited environments of oral microbiome and within host cells [15,69,70].

Recently, we demonstrated that HmuY homolog from *T. forsythia*, Tfo, also binds heme but in a manner different to the *P. gingivalis* protein. The main differences were observed in preferential heme binding by Tfo under reducing conditions, which could utilize methionine residues, instead of histidine residues, to coordinate heme iron [18]. In this study, we characterized two HmuY homologs from *P. intermedia* to reveal if PinO and PinA would play a similar function to HmuY and Tfo, thus participating in a synergistic mechanism of heme acquisition by oral pathogens. We found that PinO and PinA could use one Met residue to coordinate heme iron. Interestingly, in PinO two additional Met residues, namely Met45 and Met145 (Met76 and Met176), may interchangeably participate in heme iron coordination, whereas in the case of PinA this function could be played by Met158 (Met191). Some PinO or

PinA protein variants with singly or dually substituted Met residues were still able to bind heme with low efficiency, but one may assume that this binding occurs between other amino acids of PinO or PinA and PPIX ring. Similar properties were reported for *P. gingivalis* HmuY [16]. The replicated MD simulations showed that for PinO significant close contacts - close enough for chemical coordination in a real system - occur between the heme iron and the Met ligands in the heme binding pocket. We assume that these residues could also play supporting role in heme binding, similarly to Met136 in HmuY [17] or His83 in HasA hemophores [71-73].

We suggest that heme bound to PinO and PinA, similar to Tfo, might represent a heme reservoir for *P. gingivalis*. The heme might be accessed by the action of HmuY mainly during phases of colonization when *P. intermedia* dominates over *P. gingivalis* and when bacteria are exposed to oxygen. In chronic periodontitis, formation of deep periodontal pockets is associated with a decrease in oxidation-reduction potential and a reducing environment preferred by anaerobic bacteria [74,75]. In the early stages of this disease, bacteria encounter reducing conditions without bleeding, where the main heme source could be host albumin present in the serum-like exudate of gingival crevicular fluid. If *P. intermedia* can reside in this early environment, when *P. gingivalis* is not yet dominant, then any heme captured from reduced albumin-heme complexes becomes a more important part of a bacterial heme pool for future access by other hemophore-like proteins, such as HmuY. It is noteworthy in this context that the affinity of albumin for Fe(II)heme is lower than for heme in the Fe(III) [76,77], which may facilitate heme capture by PinO and PinA. Moreover, the reducing conditions would influence the properties of iron coordination by Met residues more effectively than by His residues. This may allow for efficient heme binding to PinO and PinA and heme sequestration from the albumin-Fe(II)heme complex by these proteins, due to methionine-heme ligation in PinO and

PinA, which results in stabilization of the reduced state as compared to bis-His ligation in HmuY [78,79]. Such effect could be also explained by theory of hard and soft acids and bases [80,81], demonstrating that Met-ligand binding would be destabilized under oxidizing conditions. Therefore, one may assume that in the early stage of colonization occurring under reducing conditions, when *P. intermedia* dominates over *P. gingivalis*, the former bacterium may take advantage of higher ability of heme binding by PinO and even use the heme pool bound to HmuY. Taken together, we found here that PinO and PinA exhibit properties more similar to *T. forsythia* Tfo [18] than to *P. gingivalis* HmuY [16,17].

Previously, we demonstrated that *P. gingivalis* produces higher levels of HmuY, when the bacteria grew under low-iron/heme conditions or as a biofilm constituent [15,69], as well as intracellularly in host cells [70]. It has been shown by others that levels of PinO associated with the outer-membrane of *P. intermedia* increased under low-iron conditions [82]. Recently, we also showed that *T. forsythia* produces higher levels of Tfo when grown under low-iron/heme conditions [18]. Moreover, the HmuY and Tfo proteins are associated with both the bacterial outer membrane and outer-membrane vesicles through their lipid anchors [18,69,83,84] and can also be shed as intact, soluble proteins [17,18,69,83]. We demonstrated similar localization of *P. intermedia* proteins with the exception that PinO was released in the soluble form at very low level, similar to Tfo [18], and PinA was not processed into soluble protein. This might suggest that *P. intermedia* does not produce protease/s capable of shedding the protein from the bacterial cell surface or there is no specific cleavage sequence in PinA protein recognized by *P. intermedia* proteolytic enzymes. Although one may not exclude the possibility that soluble PinO and PinA are degraded in the external environment by *P. intermedia* proteases, we demonstrated resistance of PinO and PinA to proteolytic activity expressed by *P. intermedia*. Importantly,

therefore, from the pathogenesis point of view, PinO can be accumulated in the environment and used by bacteria more efficiently in the early stage of chronic periodontitis, where *P. intermedia* may dominate over *P. gingivalis*. However, when *P. gingivalis* dominates over *P. intermedia*, the former bacterium can acquire heme bound to these proteins, both by heme liberation due to production of active proteases and sequestration mechanism, the latter being more effective under air (oxidizing) conditions.

The new data presented here on *P. intermedia* PinO and PinA significantly broaden our knowledge about the novel family of hemophore-like proteins. Our analyses indicate that various members of this family may have developed a specific heme-binding pocket and the ability to efficiently sequester heme from host hemoproteins, similar to classical hemophores, or even to acquire heme bound to heme-binding proteins produced by cohabitating bacteria. The independent evolution of these properties has resulted in different mechanisms of heme coordination in HmuY *versus* Tfo, PinO and PinA. The former binds heme very effectively both in oxidizing and reducing environment, while the three other proteins appear to prefer reducing conditions. However, some convergent features could have also evolved. Our ongoing studies are focused on structure-function relationship of HmuY homologs produced by pathogens from phylum Bacteroidetes, composing important human microbiomes of oral cavity and intestine.

Author contributions

T.O., J.W.S., R.W.S., S.A., P.M., M.B., M.O. designed the study; R.W.S., S.A., J.W.S., M.B., M.Ś., P.M., T.O. analyzed the data; M.B., S.A., R.W.S., J.W.S., P.M., M.Ś., P.Ś., M.Co., M.Ca., K.S., M.O., T.O. performed experiments; T.O., J.W.S., R.W.S., S.A., P.M. wrote the original

draft; R.W.S., S.A., J.W.S., P.M., T.O., M.Ś., M.B., P.Ś., K.S., M.O. reviewed and edited the manuscript.

Funding

This work was supported by grant no. 2015/17/B/NZ6/01969 from the National Science Center (NCN, Krakow, Poland) (TO). Preliminary experiments were financed by the Wroclaw Research Center EIT+ under the project "Biotechnologies and advanced medical technologies – BioMed' (POIG 01.01.02-02-003/08/00) from the European Regional Development Fund (Operational Program Innovative Economy, 1.1.2) and Wroclaw Center of Biotechnology, The Leading National Research Center (KNOW) program for years 2014-2018 (TO). We thank Diamond Light Source for access to beamlines IO4-1 (proposal number 15991) that contributed to the results presented here (SA). Some computations were carried out at the Wrocław Center for Networking and Supercomputing under the grant no. 307 (PM). The funders had no role in study design, data collection, analysis, interpretation, and decision to publish the manuscript.

Competing interests

The authors declare that they have no conflicts of interest with the contents of this article.

References

1. Deng, Z.L., Szafranski, S.P., Jarek, M., Bhuju, S. and Wagner-Dobler, I. (2017) Dysbiosis in chronic periodontitis: Key microbial players and interactions with the human host. *Sci. Rep.* **7**(1):3703

663 2. Socransky, S.S., Haffajee, A.D., Cugini, M.A., Smith, C. and Kent, R.L. Jr. (1998) Microbial
664 complexes in subgingival plaque. *J. Clin. Periodontol.* **25**, 134-144
665 3. Kolenbrander, P.E., Andersen, R.N., Blehert, D.S., Egland, P.G., Foster, J.S. and Palmer, R.J.
666 Jr. (2002) Communication among oral bacteria. *Microbiol. Mol. Biol. Rev.* **66**, 486-505
667 4. Kamaguchi, A., Ohyama, T., Sakai, E., Nakamura, R., Watanabe, T., Baba, H. et al. (2003)
668 Adhesins encoded by the gingipain genes of *Porphyromonas gingivalis* are responsible for co-
669 aggregation with *Prevotella intermedia*. *Microbiology* **149**, 1257-1264
670 5. Hajishengallis, G. (2011) Immune evasion strategies of *Porphyromonas gingivalis*. *J. Oral*
671 *Biosci.* **53**, 233-240
672 6. Bostanci, N. and Belibasakis, G.N. (2012) *Porphyromonas gingivalis*: an invasive and evasive
673 opportunistic pathogen. *FEMS Microbiol. Lett.* **333**, 1-9
674 7. Darveau, R.P., Hajishengallis, G. and Curtis, M.A. (2012) *Porphyromonas gingivalis* as a
675 potential community activist for disease. *J. Dent. Res.* **91**, 816-820
676 8. Byrne, D.P., Potempa, J., Olczak, T. and Smalley, J.W. (2013) Evidence of mutualism
677 between two periodontal pathogens: Co-operative haem acquisition by the HmuY
678 haemophore of *Porphyromonas gingivalis* and the cysteine protease interpain A (InpA) of
679 *Prevotella intermedia*. *Mol. Oral Microbiol.* **28**, 219-229
680 9. Haigh, R.D., Crawford, L.A., Ralph, J.D., Wanford, J.J., Vartoukian, S.R., Hijazi, K. et al.
681 (2017) Draft whole-genome sequences of periodontal pathobionts *Porphyromonas gingivalis*,
682 *Prevotella intermedia*, and *Tannerella forsythia* contain phase-variable restriction-
683 modification systems. *Genome Announc.* **5**(46):e01229-17

684 10. Takahashi, N. and Sato, T. (2002) Dipeptide utilization by the periodontal pathogens
685 *Porphyromonas gingivalis*, *Prevotella intermedia*, *Prevotella nigrescens* and *Fusobacterium*
686 *nucleatum*. *Oral Microbiol. Immunol.* **17**, 50-54
687 11. Simpson, W., Olczak, T. and Genco, C.A. (2000) Characterization and expression of HmuR,
688 a TonB-dependent hemoglobin receptor of *Porphyromonas gingivalis*. *J. Bacteriol.* **182**,
689 5737-5748
690 12. Olczak, T., Dixon, D.W. and Genco, C.A. (2001) Binding specificity of the *Porphyromonas*
691 *gingivalis* heme and hemoglobin receptor HmuR, gingipain K, and gingipain R1 for heme,
692 porphyrins, and metalloporphyrins. *J. Bacteriol.* **183**, 5599-5608
693 13. Liu, X., Olczak, T., Guo, H.C., Dixon, D.W. and Genco, C.A. (2006) Identification of amino
694 acid residues involved in heme binding and hemoprotein utilization in the *Porphyromonas*
695 *gingivalis* heme receptor HmuR. *Infect. Immun.* **74**, 1222-1232
696 14. Olczak, T. (2006) Analysis of conserved glutamate residues in *Porphyromonas gingivalis*
697 outer membrane receptor HmuR: toward a further understanding of heme uptake. *Arch.*
698 *Microbiol.* **186**, 393-402
699 15. Olczak, T., Sroka, A., Potempa, J. and Olczak, M. (2008) *Porphyromonas gingivalis* HmuY
700 and HmuR: further characterization of a novel mechanism of heme utilization. *Arch.*
701 *Microbiol.* **189**, 197-210
702 16. Wojtowicz, H., Wojaczynski, J., Olczak, M., Kroliczewski, J., Latos-Grazynski, L. and
703 Olczak, T. (2009) Heme environment of *Porphyromonas gingivalis* HmuY heme-binding
704 protein. *Biochem. Biophys. Res. Commun.* **382**, 178-182

17. Wojtowicz, H., Guevara, T., Tallant, C., Olczak, M., Sroka, A., Potempa, J., Solà, M. et al. (2009) Unique structure and stability of HmuY, a novel heme-binding protein of *Porphyromonas gingivalis*. *PLoS Pathog.* **5**(5):e1000419

18. Bielecki, M., Antonyuk, S., Strange, R.W., Smalley, J.W., Mackiewicz, P., Smiga, M. et al. 2018. *Porphyromonas gingivalis* HmuY and *Tannerella forsythia* Tfo – two homologous proteins with different heme-binding properties. *Biosci. Rep.* **38**(5):BSR20181325

19. Leung, K.P., Subramaniam, P.S., Okamoto, M., Fukushima, H. and Lai, C.H. (1998) The binding and utilization of hemoglobin by *Prevotella intermedia*. *FEMS Microbiol. Lett.* **162**, 227-233

20. Leung, K.P. and Folk, S.P. (2002) Effects of porphyrins and inorganic iron on the growth of *Prevotella intermedia*. *FEMS Microbiol. Lett.* **209**, 15-21

21. Silva, T.A., Rodrigues, P.H., Ribeiro, R.N., Noronha, F.S., Farias, L. de M. and Carvalho, M.A. (2003) Hemolytic activity of *Prevotella intermedia* and *Prevotella nigrescens* strains: influence of abiotic factors in solid and liquid assays. *Res. Microbiol.* **154**, 29-35

22. Suzuki, N., Fukamachi. H., Arimoto, T., Yamamoto, M. and Igarashi, T. (2012) Contribution of *hly* homologs to the hemolytic activity of *Prevotella intermedia*. *Anaerobe* **18**, 350-356

23. Tompkins, G.R., Wood, D.P. and Birchmeier, K.R. (1997) Detection and comparison of specific hemin binding by *Porphyromonas gingivalis* and *Prevotella intermedia*. *J. Bacteriol.* **179**, 620-626

24. Okamoto, M., Maeda, N., Kondo, K. and Leung, K.P. (1999) Hemolytic and hemagglutinating activities of *Prevotella intermedia* and *Prevotella nigrescens*. *FEMS Microbiol. Lett.* **178**, 299-304

25. Takada, K., Fukatsu, A., Otake, S. and Hirasawa, M. (2003) Isolation and characterization of hemolysin activated by reductant from *Prevotella intermedia*. *FEMS Immunol. Med. Microbiol.* **35**, 43-47

26. Smalley, J.W. and Olczak, T. (2017) Heme acquisition mechanisms of *Porphyromonas gingivalis* - strategies used in polymicrobial community in a heme-limited host environment. *Mol. Oral Microbiol.* 32, 1-23

27. Smalley, J.W., Silver, J., Birss, A.J., Withnall, R. and Titler, P.J. (2003) The haem pigment of the oral anaerobes *Prevotella nigrescens* and *Prevotella intermedia* is composed of iron(III) protoporphyrin IX in the monomeric form. *Microbiology* **149**, 1711-1718

28. Smalley, J.W., Birss, A.J., Szmigielski, B. and Potempa, J. (2008) Mechanism of methaemoglobin breakdown by the lysine-specific gingipain of the periodontal pathogen *Porphyromonas gingivalis*. *Biol. Chem.* **389**, 1235-1238

29. Byrne, D.P., Wawrzonek, K., Jaworska, A., Birss, A.J., Potempa, J. and Smalley, J.W. (2009) Role of the cysteine protease interpain A of *Prevotella intermedia* in breakdown and release of haem from haemoglobin. *Biochem. J.* **425**, 257-264

30. Byrne, D.P., Manandhar, S.P., Potempa, J. and Smalley, J.W. (2015) Breakdown of albumin and haemalbumin by the cysteine protease interpain A, an albuminase of *Prevotella intermedia*. *BMC Microbiol.* **15**:185

31. Gmiterek, A., Wojtowicz, H., Mackiewicz, P., Radwan-Oczko, M., Kantorowicz, M., Chomyszyn-Gajewska, M. et al. (2013) The unique *hmuY* gene sequence as a specific marker of *Porphyromonas gingivalis* infection. *PLoS One* **8**(7):e67719

748 32. Smiga, M., Bielecki, M., Olczak, M., Smalley, J.W. and Olczak, T. (2015) Anti-HmuY
749 antibodies specifically recognize *Porphyromonas gingivalis* HmuY protein but not
750 homologous proteins in other periodontopathogens. *PloS One* **10**(2):e0117508
751 33. Smiga, M., Bielecki, M., Olczak, M. and Olczak, T. (2019) *Porphyromonas gingivalis* PgFur
752 is a member of a novel Fur subfamily with non-canonical function. *Front. Cell. Infect.*
753 *Microbiol.* **9**:233
754 34. Smalley, J.W., Byrne, D.P., Birss, A.J., Wojtowicz, H., Sroka, A., Potempa, J. et al. (2011)
755 HmuY haemophore and gingipain proteases constitute a unique synthrophic system of haem
756 acquisition by *Porphyromonas gingivalis*. *PLoS One* **6**(2):e17182
757 35. Pluym, M., Muryoi, N., Heinrichs, D.E. and Stillman, M.J. (2008) Heme binding in the
758 NEAT domains of IsdA and IsdC of *Staphylococcus aureus*. *J. Inorg. Biochem.* **102**, 480-488
759 36. Kabsch, W. (2010) XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**(Pt 2):125-132
760 37. Vonrhein, C., Flensburg, C., Keller, P., Sharff, A., Smart, O., Paciorek, W. et al. (2011) Data
761 processing and analysis with the autoPROC toolbox. *Acta Crystallogr.* **D67**, 293-302
762 38. Vagin, A. and Teplyakov, A. (1997) MOLREP: An Automated Program for Molecular
763 Replacement. *J. Appl. Crystallogr.* **30**, 1022–1025
764 39. Langer, G., Cohen, S.X., Lamzin, V.S. and Perrakis, A. (2008) Automated macromolecular
765 model building for X-ray crystallography using ARP/wARP version 7. *Nat. Protoc.* **3**, 1171-
766 1179
767 40. Murshudov, G.N., Vagin, A.A. and Dodson, E.J. (1997) Refinement of macromolecular
768 structures by the maximum-likelihood method. *Acta Crystallogr. Sect. D Biol. Crystallogr.*
769 **53**, 240–255

41. Emsley, P. and Cowtan, K. (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 2126–2132

42. Davis, IW, Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X. et al. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucl. Acids Res. **35**, W375-W383

43. Abraham, M.J., Murtola, T., Schulz, R., Pall, S., Smith, J.C., Hess, B. et al. (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1**, 19-25

44. Schmid, N., Eichenberger, A., Choutko, A., Riniker, S., Winger, M., Mark, A. et al. (2011) Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur. Biophys. J.* **40**, 843–856

45. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F. and Hermans, J. (1981) Interaction models for water in relation to protein hydration. In *Intermolecular Forces*, edited by B. Pullman (Reidel, Dordrecht), p. 331-342.

46. Parrinello, M. and Rahman, A. (1981) Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190

47. Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H. and Pedersen, L.G. (1995) A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8593

48. Humphrey, W., Dalke, A. and Schulten, K. (1996) VMD - Visual Molecular Dynamics. *J. Mol. Graph.* **14**, 33-38

49. Schaffer, A.A., Aravind, L., Madden, T.L., Shavirin, S., Spouge, J.L., Wolf, Y.I. et al. (2001) Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. *Nucleic Acids Res.* **29**, 2994-3005

50. Marchler-Bauer, A., Derbyshire, M.K., Gonzales, N.R., Lu, S., Chitsaz, F., Geer, L.Y. et al. (2015) CDD: NCBI's conserved domain database. *Nucleic Acids Res.* **43**:D222-226

51. Katoh, K. and Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772-780

52. Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M. and Barton, G.J. (2009) Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189-1191

53. Notredame, C., Higgins, D.G. and Heringa, J. (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **302**, 205-217

54. Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Hohna, S. et al. (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539-542

55. Lartillot, N. and Philippe, H. (2004) A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol. Biol. Evol.* **21**, 1095-1109

56. Nguyen, L.T., Schmidt, H.A., von Haeseler, A. and Minh, B.Q. (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268-274

57. Criscuolo, A. (2011) morePhyML: improving the phylogenetic tree space exploration with PhyML 3. *Mol. Phylogenet. Evol.* **61**, 944-948

58. Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W. and Gascuel, O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307-321

815 59. Huelsenbeck, J.P., Larget, B. and Alfaro, M.E. (2004) Bayesian phylogenetic model selection
816 using reversible jump Markov chain Monte Carlo. *Mol. Biol. Evol.* **21**, 1123-1133
817 60. Darriba, D., Taboada, G.L., Doallo, R. and Posada, D. (2011) ProtTest 3: fast selection of
818 best-fit models of protein evolution. *Bioinformatics* **27**, 1164-1165
819 61. Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A. and Jermiin, L.S. (2017)
820 ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587-
821 589
822 62. Anisimova, M. and Gascuel, O. (2006) Approximate likelihood-ratio test for branches: A
823 fast, accurate, and powerful alternative. *Syst. Biol.* **55**, 539-552
824 63. Rambaut, A. (2012) FigTree v1.4.0. Available from: http://tree.bio.ed.ac.uk/software/figtree/
825 64. Stover, B.C. and Muller, K.F. (2010) TreeGraph 2: combining and visualizing evidence from
826 different phylogenetic analyses. *BMC Bioinformatics* **11**:7
827 65. Benedyk, M., Byrne, D.P., Glowczyk, I., Potempa, J., Olczak, M., Olczak, T. et al. (2015)
828 Pyocyanin, a contributory factor in haem acquisition and virulence enhancement of
829 *Porphyromonas gingivalis* in the lung. *PloS One* **10**(2):e0118319
830 66. Smalley, J.W., Birss, A.J., Szmigielski, B. and Potempa, J. (2007) Sequential action of R and
831 K-specific gingipains of *Porphyromonas gingivalis* in the generation of the haem containing
832 pigment from oxyhaemoglobin. *Arch. Biochem. Biophys.* **465**, 44-49
833 67. Wandersman, C. and Delepelaire, P. (2012) Haemophore functions revisited. *Mol. Microbiol.*
834 **85**, 618–631
835 68. Benson, D.R. and Rivera, M. (2013) Heme uptake and metabolism in bacteria. *Met. Ions Life*
836 *Sci.* **12**, 279-332

837 69. Olczak, T., Wojtowicz, H., Ciuraszkiewicz, J. and Olczak, M. (2010) Species specificity,
838 surface exposure, protein expression, immunogenicity, and participation in biofilm formation
839 of *Porphyromonas gingivalis* HmuY. *BMC Microbiol.* **10**:134
840 70. Olczak, T., Sosicka, P. and Olczak, M. (2015) HmuY is an important virulence factor for
841 *Porphyromonas gingivalis* growth in the heme-limited host environment and infection of
842 macrophages. *Biochem. Biophys. Res. Commun.* **467**, 748-753
843 71. Caillet-Saguy, C., Piccioli, M., Turano, P., Lukat-Rodgers, G., Wolff, N., Rodgers, K.R. et
844 al. (2012) Role of the iron axial ligands of heme carrier HasA in heme uptake and release. *J.*
845 *Biol. Chem.* **287**, 26932-26943
846 72. Ozaki, S.I., Sato, T., Sekine, Y., Migita, C.T., Uchida, T. and Ishimori, K. (2014)
847 Spectroscopic studies on HasA from *Yersinia tuberculosis*. *J. Inorg. Biochem.* **138**, 31-38
848 73. Kumar, R., Matsumura, H., Lovell, S., Yao, H., Rodriguez, J.C., Battaile, K.P. et al. (2014)
849 Replacing the axial ligand tyrosine 75 or its hydrogen bond partner histidine 83 minimally
850 affects hemin acquisition by the hemophore HasAp from *Pseudomonas aeruginosa*.
851 *Biochemistry* **53**, 2112-2125
852 74. Kenney, E.B. and Ash, M.M. (1969) Oxidation reduction potential of developing plaque,
853 periodontal pockets and gingival sulci. *J. Periodontol.* **40**, 630-633
854 75. Leke, N., Grenier, D., Goldner, M. and Mayrand, D. (1999) Effects of hydrogen peroxide on
855 growth and selected properties of *Porphyromonas gingivalis*. *FEMS Microbiol. Lett.* **174**,
856 347-353
857 76. Cao, Y., Nicoletti, F.P., De Sanctis, G., Bocedi, A., Ciaccio, C., Gullotta, F. et al. (2012)
858 Evidence for pH-dependent multiple confomers in iron(II) heme-human serum albumin:

spectroscopic and kinetic investigation of carbon monoxide binding. *J. Biol. Inorg. Chem.* **17**, 133-147

77. Bocedi, A., De Sanctis, G., Ciaccio, C., Tundo, G.R., Di Masi, A., Fanali, G. et al. (2013) Reciprocal allosteric modulation of carbon monoxide and warfarin binding to ferrous human serum heme-albumin. *PLoS One* **8**(3):e58842

78. Moore, G.R. and Pettigrew, G.W. (1990) *Cytochromes c*: Evolutionary, Structural and Physicochemical Aspects. Springer-Verlag, Berlin

79. Schejter, A., Plotkin, B. and Vig, I. (1991) The reactivity of cytochrome *c* with soft ligands. *FEBS Lett.* **280**, 199-201

80. Ayers, P.W., Parr, R.G. and Pearson, R.G. (2006) Elucidating the hard/soft acid/base principle: A perspective based on half-reactions. *J. Chem. Phys.* **124**, 194107

81. Pearson, R.G. (1963) Hard and soft acids and bases. *J. Am. Chem. Soc.* **22**, 3533-3539

82. Yu, F., Anaya, C. and Lewis, J.P. (2007) Outer membrane proteome of *Prevotella intermedia* 17: Identification of thioredoxin and iron-repressible hemin uptake loci. *Proteomics* **7**, 403-412

83. Veith, P.D., Chen, Y.Y., Gorasia, D.G., Chen, D., Glew, M.D., O'Brien-Simpson, N.M. et al. (2014) *Porphyromonas gingivalis* outer membrane vesicles exclusively contain outer membrane and periplasmic proteins and carry a cargo enriched with virulence factors. *J. Proteome Res.* **13**, 2420-2432

84. Veith, P.D., Chen, Y.Y., Chen, D., O'Brien-Simpson, N.M., Cecil, J.D., Holden, J.A. et al. (2015) *Tannerella forsythia* outer membrane vesicles are enriched with substrates of the type IX secretion system and TonB-dependent receptors. *J. Proteome Res.* **14**, 5355-5366

881 85. Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J. and Y Zhang, Y (2015). The I-TASSER Suite:
882 Protein structure and function prediction. *Nat. Methods* **12**, 7-8

883

884 **Legends to figures**

885 **Figure 1. Simplified phylogenetic tree obtained in MrBayes for the HmuY homologs in**
886 **Bacteroidia.**

887 Eight *Prevotella* clades are numbered. Two *P. intermedia* groups, Pi1 and Pi2, are marked. The
888 position of PinO and PinA sequences is indicated. The values at nodes indicate: posterior
889 probabilities found in MrBayes (MB) and PhyloBayes (PB) as well as support values calculated
890 in (more)PhyML and IQ-TREE by approximate likelihood-ratio test (aLRT) based on a
891 Shimodaira-Hasegawa-like procedure (SH-P and SH-I, respectively) and non-parametric
892 bootstrap (BP-P and BP-I, respectively). The posterior probabilities < 0.5 and the percentages <
893 50% are omitted or indicated by a dash "-". The full tree with all taxa names and support values
894 is presented in Supplementary Figure S2. Pg, *Porphyromonas gingivalis*; T f, *Tannerella*
895 *forsythia*; Pi, *Prevotella intermedia.*

896 **Figure 2. The clade 5 of phylogenetic tree obtained in MrBayes for the HmuY homologs in**
897 **Bacteroidia.**

898 The full tree is presented in Figure 1 and S2 Fig. Two *P. intermedia* groups, Pi1 and Pi2, are
899 marked. PinO and PinA sequences are shown in bold. Amino acid residues homologous to sites
900 involved in heme binding in PinO are shown. Residues corresponding to sites potentially
901 coordinating heme iron in Tfo are shown in purple. The values at nodes indicate: posterior
902 probabilities found in MrBayes (MB) and PhyloBayes (PB) as well as support values calculated
903 in (more)PhyML and IQ-TREE by approximate likelihood-ratio test (aLRT) based on a

Shimodaira-Hasegawa-like procedure (SH-P and SH-I, respectively) and non-parametric bootstrap (BP-P and BP-I, respectively). The posterior probabilities < 0.5 and the percentages < 50% are omitted or indicated by a dash "-". The accession numbers to the numbered sequences are included in Supplementary Table S3.

Figure 3. Schematic presentation of genes encoding HmuY in *P. gingivalis* and its homologs in *P. intermedia* (PinO and PinA) shown in red.

Other genes identified in the *P. gingivalis hmu* and *P. intermedia hmu*-like operons are shown in gray. The gene marked in black does not exhibit high homology to the gene encoding putative ATPase in *P. gingivalis* (PGA7_RS02040).

Figure 4. Heme titration experiments of *P. gingivalis* HmuY, *P. intermedia* PinO and PinA.

UV-visible absorption (**A**) and difference (**B**) spectra recorded after titration of proteins (10 μM) with heme (final heme concentration in samples: 0, 2, 4, 6, 8 μM). Various color lines in (**A**) represent increasing concentrations of heme added to the buffer alone (dashed line) and to protein samples (solid lines). (**C**) The curves were generated after titration of 5 μM protein samples with heme by measuring the difference spectra between the protein+heme and heme-only samples. Samples were examined under air (oxidizing) conditions (black) or reduced by sodium dithionite (red). Results are shown as mean ± SD from 3 independent experiments.

Figure 5. UV-visible spectroscopic analysis of heme binding to purified *P. gingivalis* HmuY and *P. intermedia* PinO and PinA under different redox conditions.

(**A**) Colours of 150 μM proteins complexed with heme (protein:heme ratio 1:1) in PBS. (**B**) UV-visible absorption spectra of HmuY-, PinO- and PinA-heme complexes. Proteins (10 μM) in complex with heme were examined under air (oxidizing) conditions and subsequently reduced by sodium dithionite, and then re-oxidized using potassium ferricyanide.

Figure 6. Analysis of heme binding to purified *P. gingivalis* HmuY and *P. intermedia* PinO and PinA under different redox conditions.

Heme binding was monitored in the visible region by CD (**A**) and MCD (**B**, **C**) spectroscopies. Samples were examined under air (oxidizing) conditions (**A**, **B**) and subsequently reduced using sodium dithionite (**A, C**).

Figure 7. Amino acid sequence alignment of *Porphyromonas gingivalis*, *Tannerella forsythia* and *Prevotella intermedia* HmuY-like proteins.

Two *P. intermedia* groups, Pi1 and Pi2, presented in Figures 1 and 2 are marked. Columns with methionine residues potentially coordinating heme iron in *P. intermedia* PinO and PinA, and *T. forsythia* Tfo are marked by numbers, which correspond to positions of amino acid residues in the solved or modeled three-dimensional protein structures: 1 - PinO Met45 (Met76 according to the numbering of amino acids in the full length protein sequence); 2 -PinO Met46 (Met77); 3 - PinO Met116 (Met147), PinA Met126 (Met159); 4 - PinO Met119 (Met150), PinA Met129 (Met162); 5 - PinO Met145 (Met176), PinA Met158 (Met191). Amino acid residues which could participate in heme iron coordination are outlined by red boxes. The accession numbers to the numbered sequences are included in Supplementary Table S3.

Figure 8. The three-dimensional structures of PinO, Tfo and PinA proteins - experimental and MD models.

Cartoon representations are shown of crystal structures of (**A**) *P. gingivalis* holo-HmuY (3H8T), (**B**) *P. intermedia* apo-PinO (6R2H, this work) and (**C**) *T. forsythia* apo-Tfo (6EU8). A predicted 3D structure of *P. intermedia* PinA, generated by the I-TASSER server [85] based on sequence homology to apo-Tfo, is shown in panel (**D**). The Met residues suggested to be important in heme iron coordination are shown as sticks. In the absence of the holo-crystal structures, heme-

bound starting models for MD simulations were constructed using the crystal structures of apo-PinO and apo-Tfo. In each model Fe-heme was incorporated by manual insertion of a Fe-Met ligand bond at biochemically confirmed site, namely Met119 (Met150 according to the numbering of amino acids in the full length protein sequence) in PinO. The constructs shown are: (**E**) heme inserted into the crystal structure of apo-PinO, with its 'closed' pocket loops; (**F**) heme inserted into a conformation with 'open' pocket loops following initial MD of apo-PinO. A snapshot taken at 24 ns was selected, where Met119 is exposed to the surface; and (**G**) heme inserted into a snapshot taken from the MD trajectory at 8 ns of a conformation with open pocket loops in apo-Tfo [18]. The heme-PinO construct shown in (**E**) was also duplicated with a 90° rotation of the heme about the Fe-Met119 bond, to provide an alternative starting position for the heme propionate groups for MD simulations.

Figure 9. UV–visible absorption analysis of *P. intermedia* PinO and PinA site-directed mutagenesis variants complexed with heme.

Spectra were recorded for the protein–Fe(III)heme complexes (solid, black line) and protein–Fe(II)heme complexes (dashed, red line). All spectra were recorded at a 1:1 protein:heme molar ratio for singly or dually replaced methionine residues by alanine residues in PinO (**A**) and PinA (**B**). WT, wild type protein.

Figure 10. Time evolution of heme-Met119 models constructed from the apo-PinO crystal structure.

(**A**) The distances between the Fe atom and the sulphur atoms of Met45 and Met145 (Met76 and Met176 according to the numbering of amino acids in the full length protein sequence) from eight independent MD simulations are shown. There is close (~3 Å) and persistent contact throughout the simulations to allow either Met45 or Met145 to be considered potential ligands to

the Fe-heme. In some of the replicates the trajectories also show close contact is possible between Fe and the O atom of Met145. No other amino acid residues made similar contacts to Fe during these simulations. (**B**) Time evolution of the distance between the Fe atom and sidechain atoms of residues 45, 46 and 145 for Met45Ala and Met145Ala protein variants. The upper panels show trajectories for two independent MD simulations of the Met45Ala variant, while the lower panels show two replicates for the Met145Ala variant. For the majority of the time, the closest approach to Fe (~3 Å) occurs for the sulphur (SD) atom of the unsubstituted Met45 or Met145 residue. This is consistent with the MD simulations results from the native model. (**C**) A close up snapshot of the Fe-heme environment at 100 ns, taken from one of the MD replicates shown in (**A)**. In this configuration Met145SD is poised at the open sixth coordination ligand-binding position of Fe-heme.

Figure 11. Stability of PinO and PinA proteins.

(**A**) Proteins in apo- and holo-forms in their native states (native) and after thermal denaturation (denatured) were subjected to trypsin digestion. (**B**) Thermal denaturation of PinO was monitored using CD spectroscopy. Susceptibility of PinO and PinA to *P. gingivalis* (**C**) or *P. intermedia* (**D**) proteases was examined by growing bacteria under high-iron/heme conditions in the presence of purified HmuY, PinO or PinA proteins (marked with asterisks). Samples collected at indicated time points were separated by SDS-PAGE and proteins stained with Coomassie Brilliant Blue G-250 or examined by Western blotting using anti-PinO or anti-PinA antibodies.

Figure 12. Sequestration of heme by *P. intermedia* PinO and PinA from human albumin complexed with Fe(II)heme.

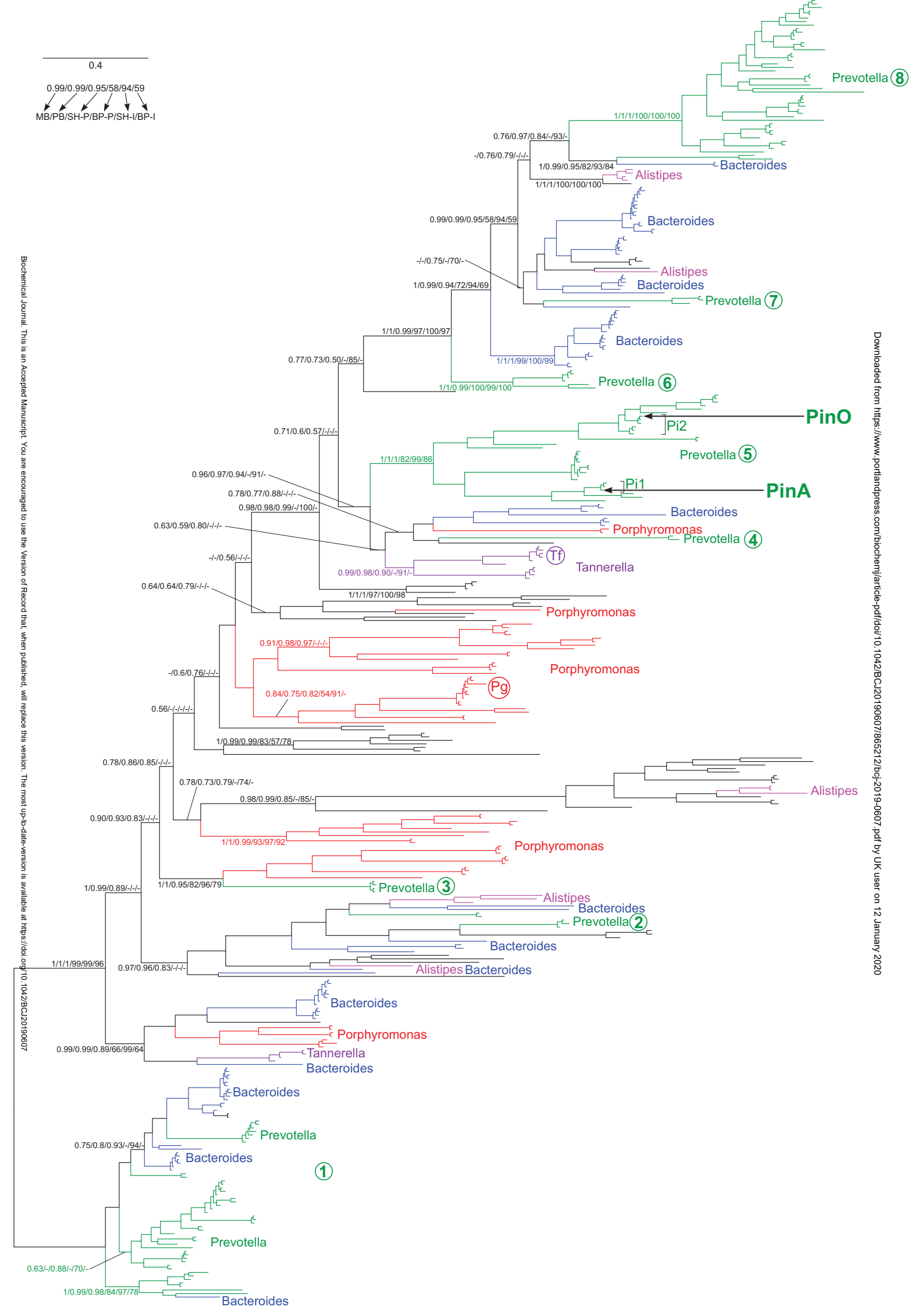
995 Red lines represent the spectra of the PinO- and PinA-Fe(II)heme complexes and black lines
996 represent the spectra of control Fe(II)heme-albumin complex prior to incubating with the
997 hemophore-like proteins. Samples were examined under reducing conditions formed by addition
998 of sodium dithionite.

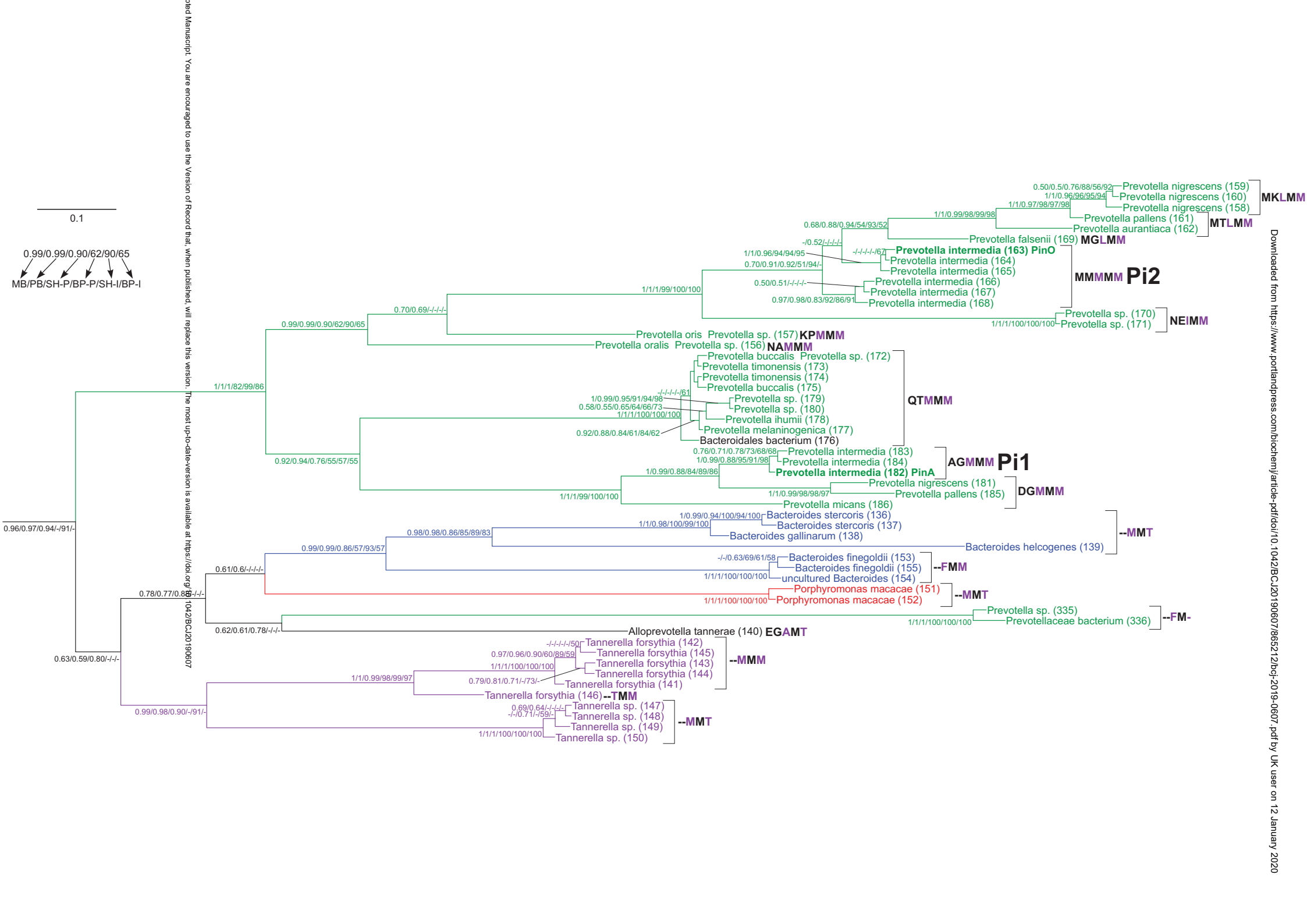
999 **Figure 13. Heme sequestration analysis.**

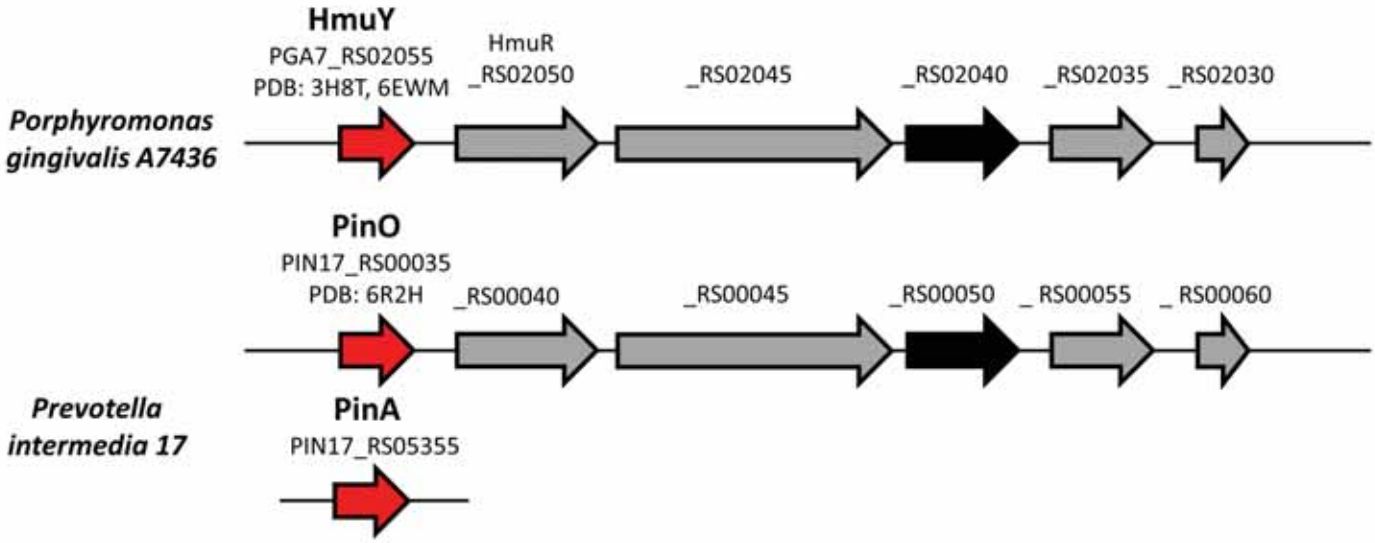
1000 Holo-PinO and holo-PinA (5 μM) were incubated under air (oxidizing) conditions with
1001 equimolar concentration of apo-HmuY at pH 7.6 (**A**) or 6 (**B**). PinO (**C**) and PinA (**D**) in holo- or
1002 apo-form were incubated with apo- or holo-HmuY at pH 7.6 also under reducing conditions.
1003 Changes in absorption spectra analyzed by UV-visible spectroscopy are shown at indicated time
1004 points.

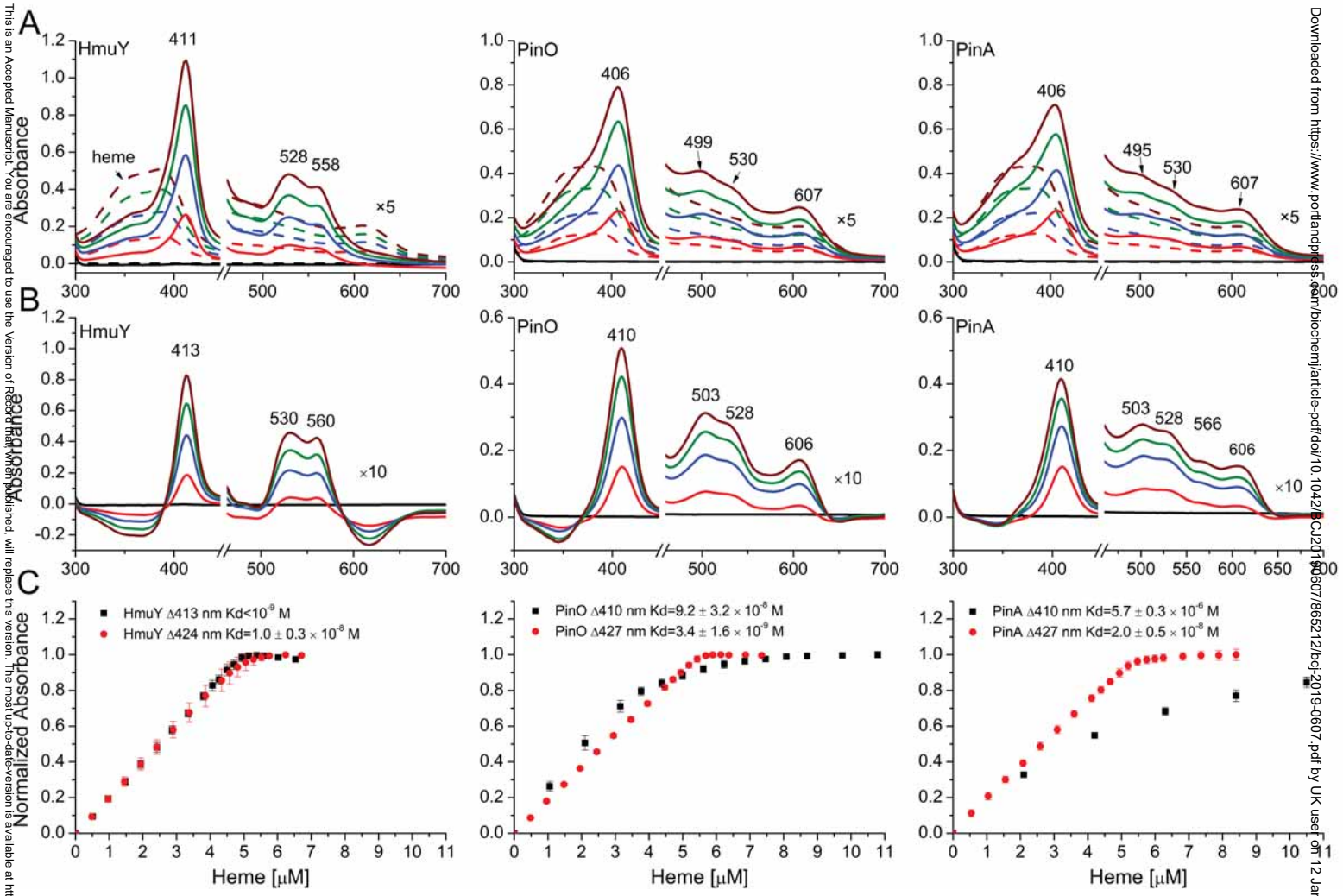
1005 **Figure 14. Expression of *P. gingivalis* HmuY and *P. intermedia* PinO and PinA during**
1006 **bacterial growth.**

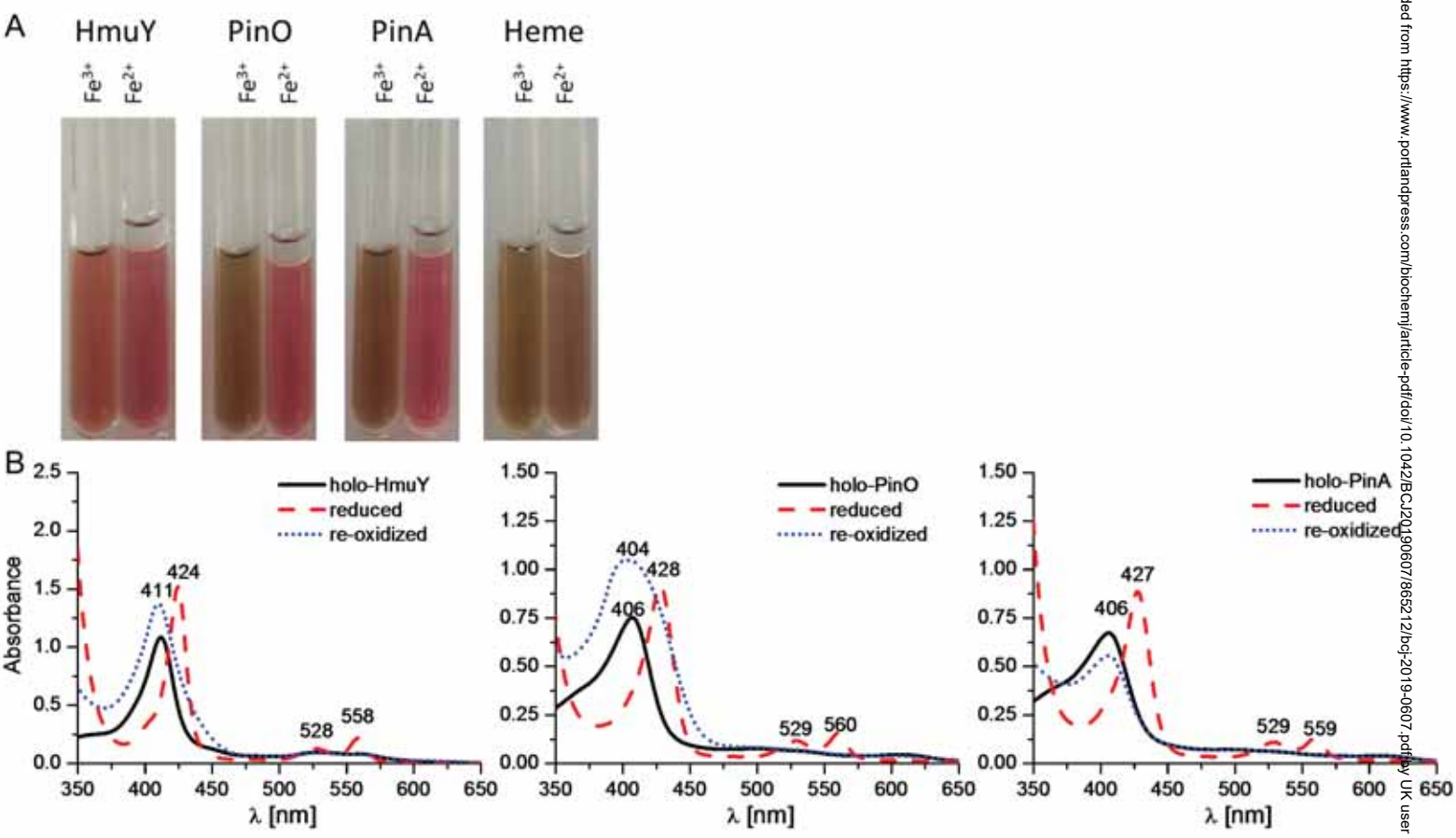
1007 (**A**) Relative changes in levels of transcripts in bacteria grown under low- (DIP) *versus* high-
1008 iron/heme (Hm) conditions at indicated time points at the first (I) and second (II) passage of
1009 bacterial cultures at indicated time points were determined by RT-qPCR. (**B**) Bacteria were
1010 grown in liquid culture media under high- (Hm) or low-iron/heme (DIP) conditions for 24 h.
1011 Samples were analyzed by Western blotting. WC, whole bacterial cells; SP, soluble protein shed
1012 from bacterial outer membrane; OMV, outer-membrane vesicles. To enable visualization of the
1013 soluble PinA and PinO, the culture medium was concentrated 25× by ultrafiltration.

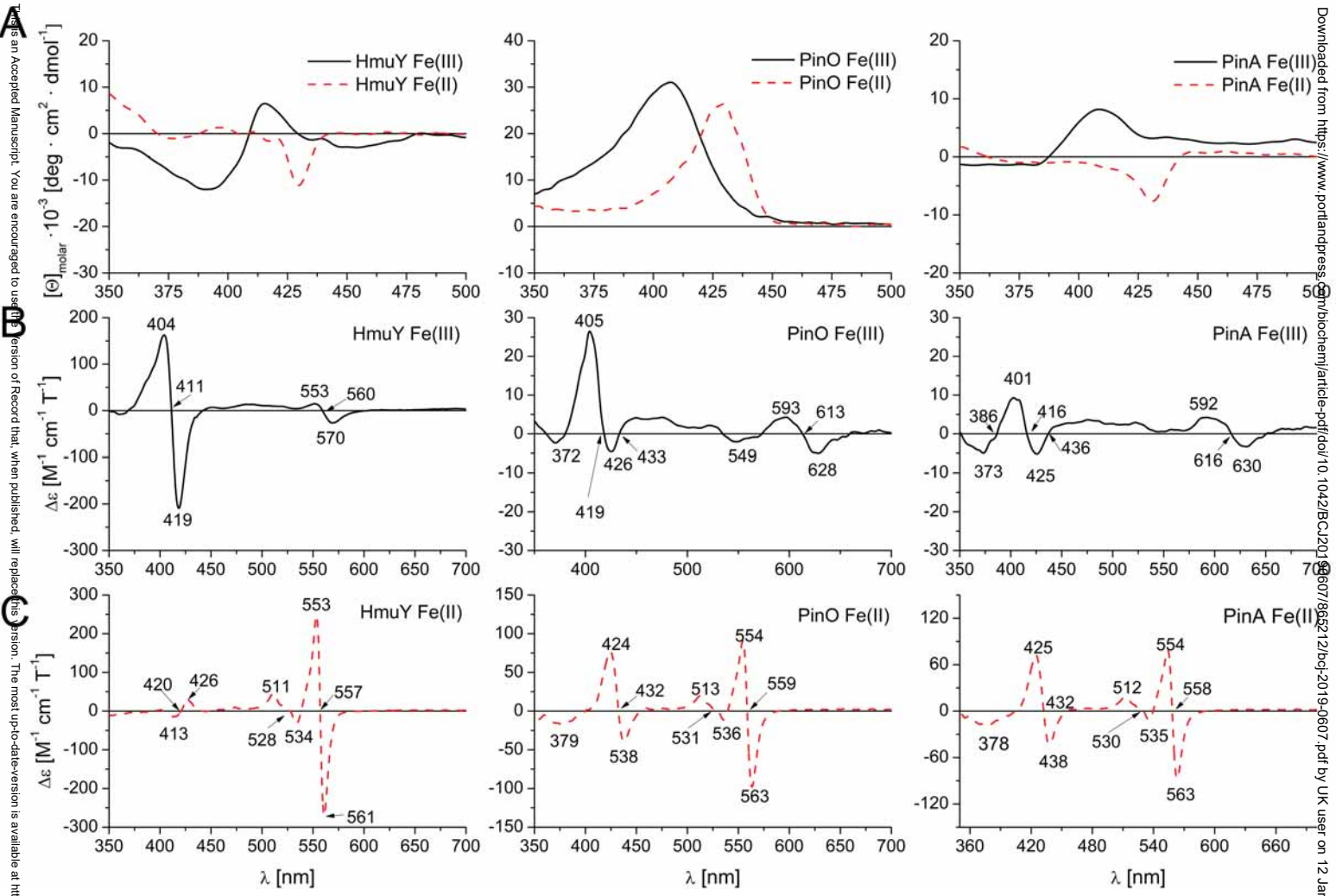


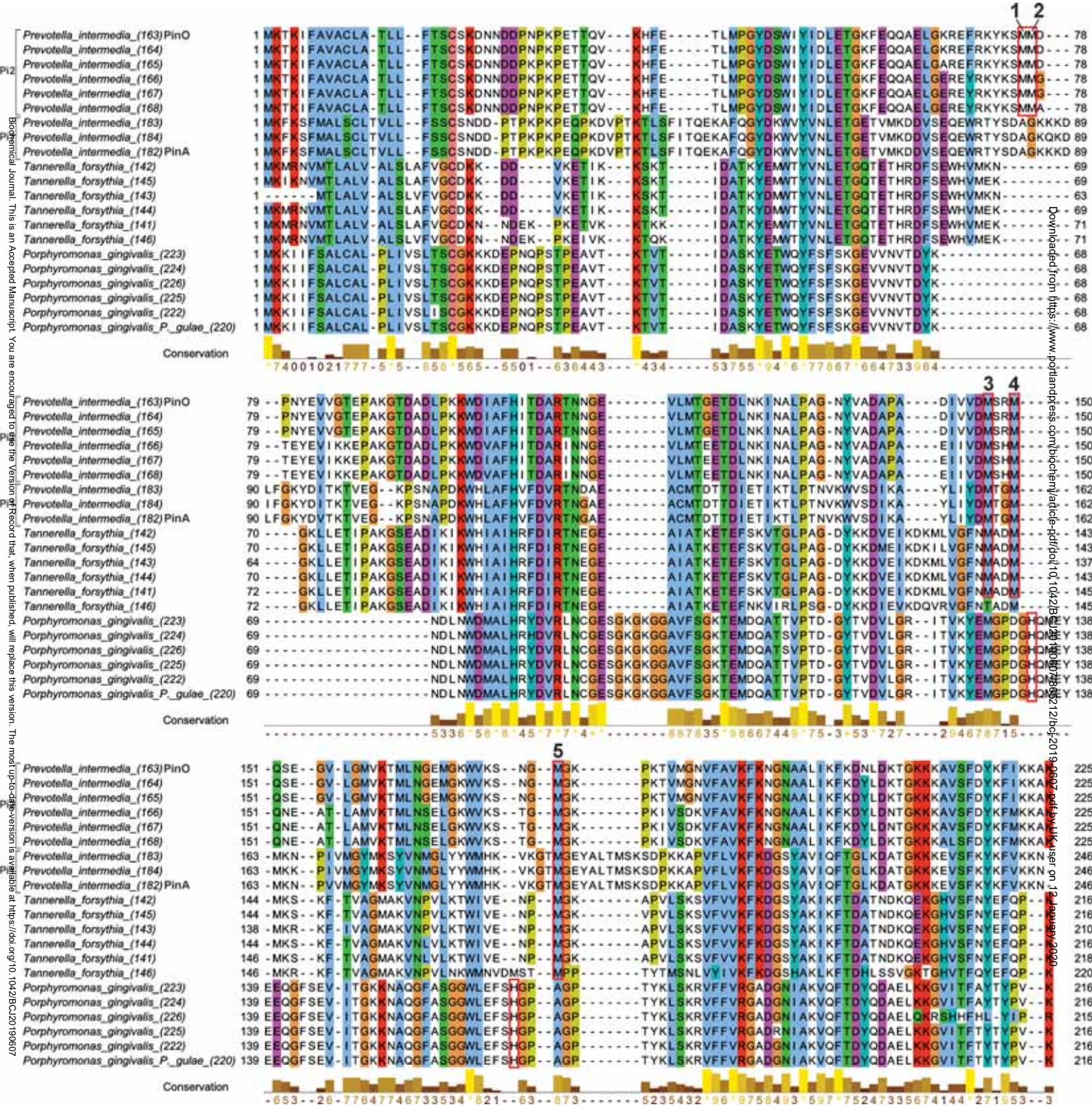


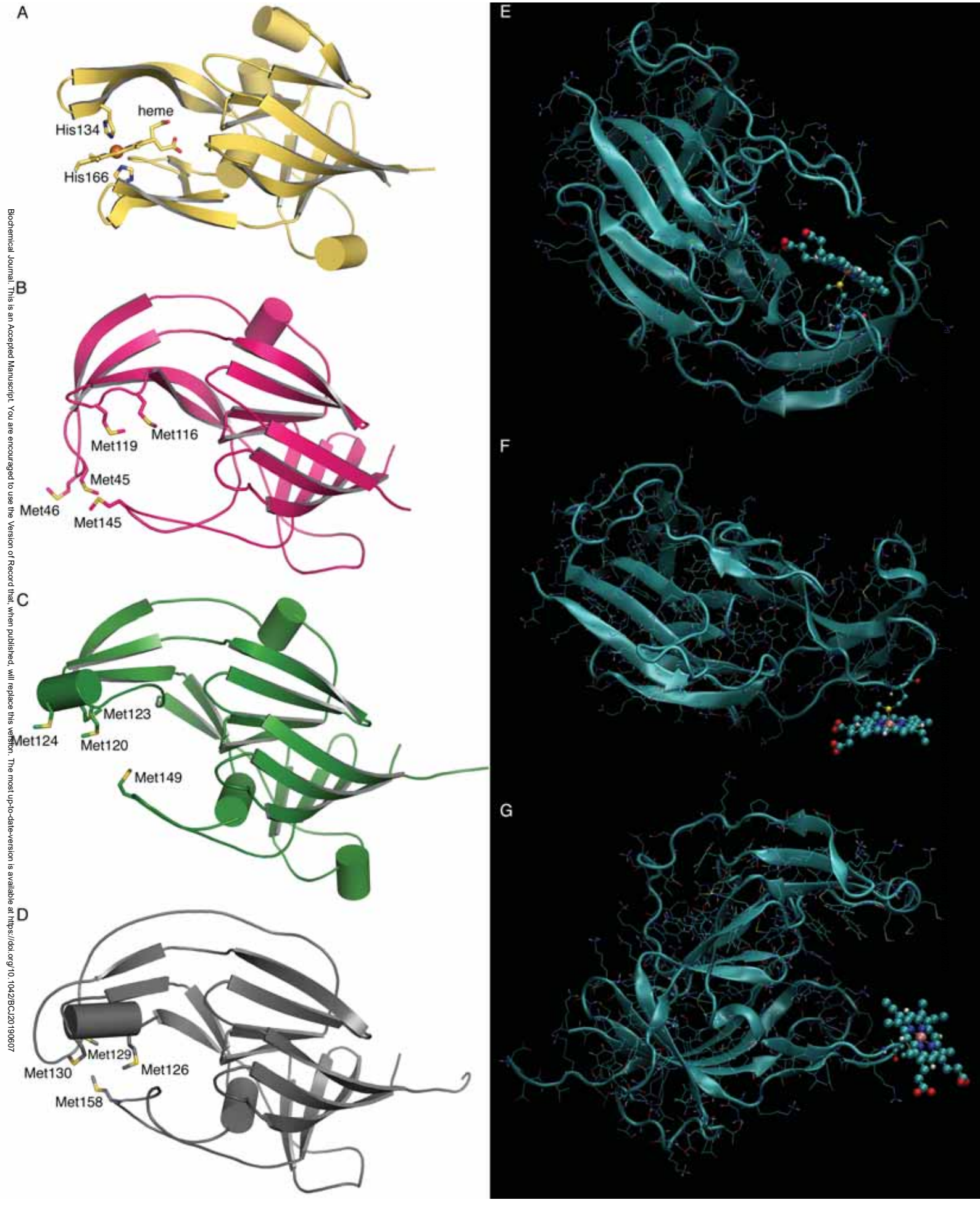


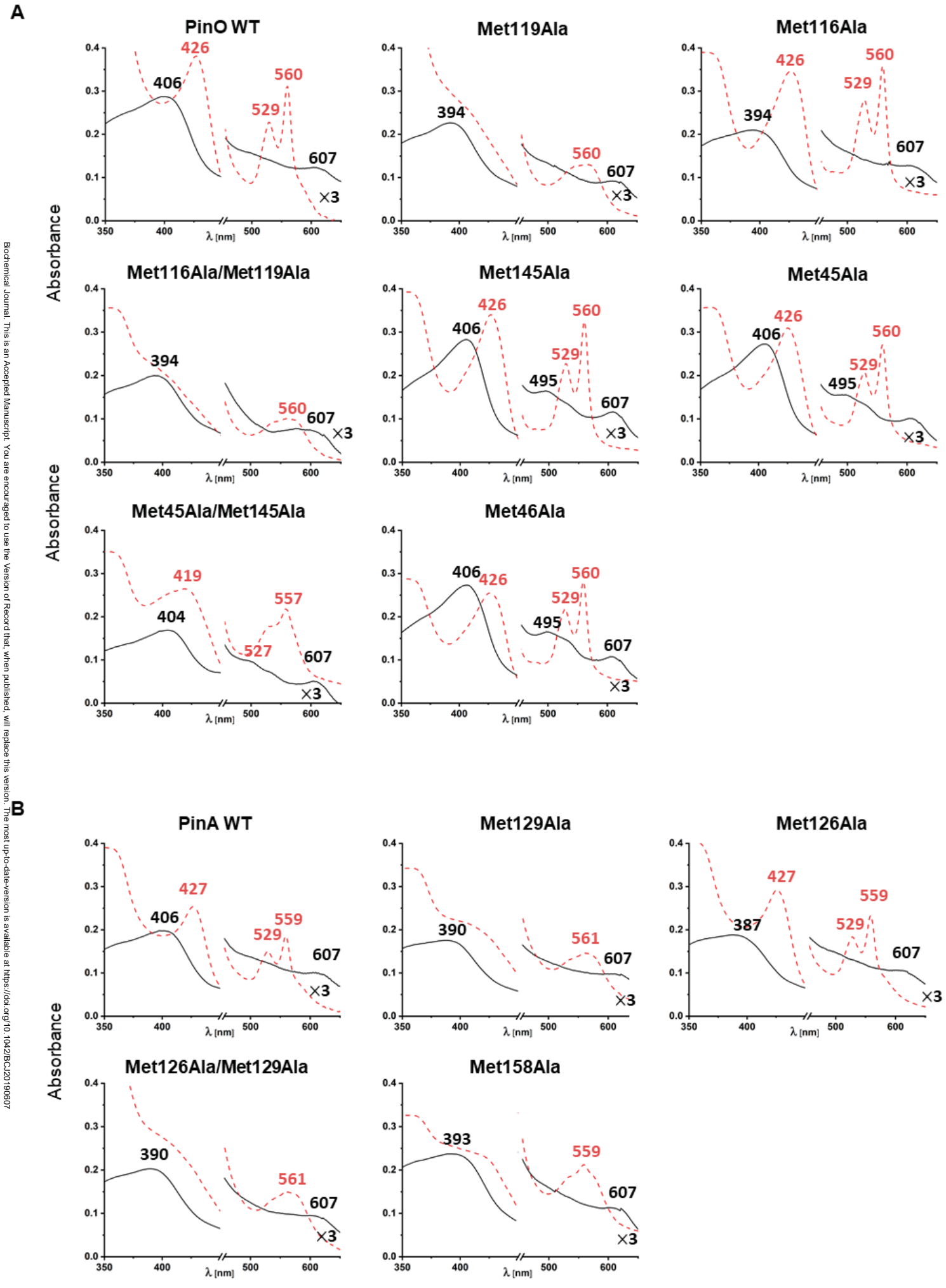


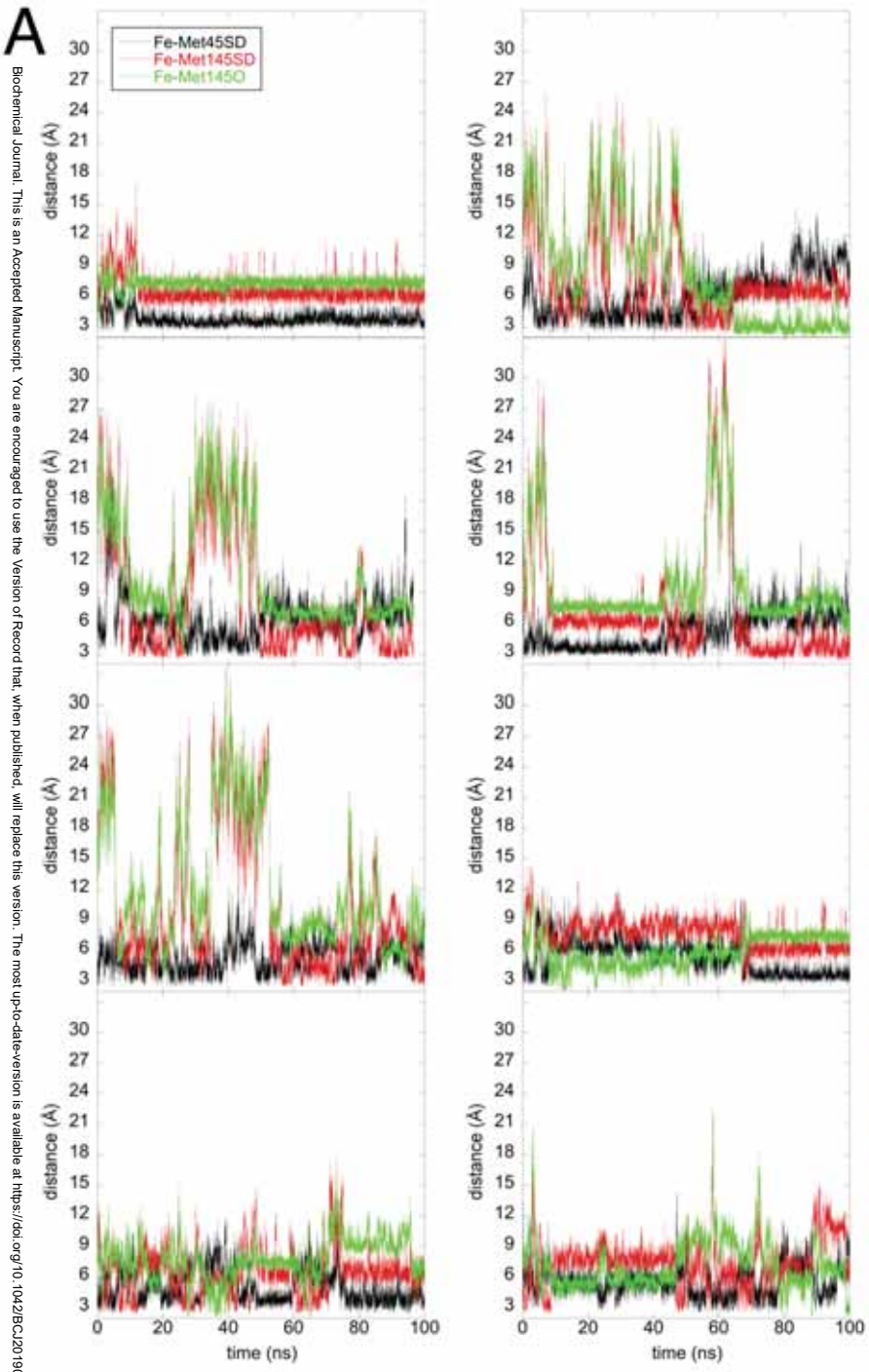


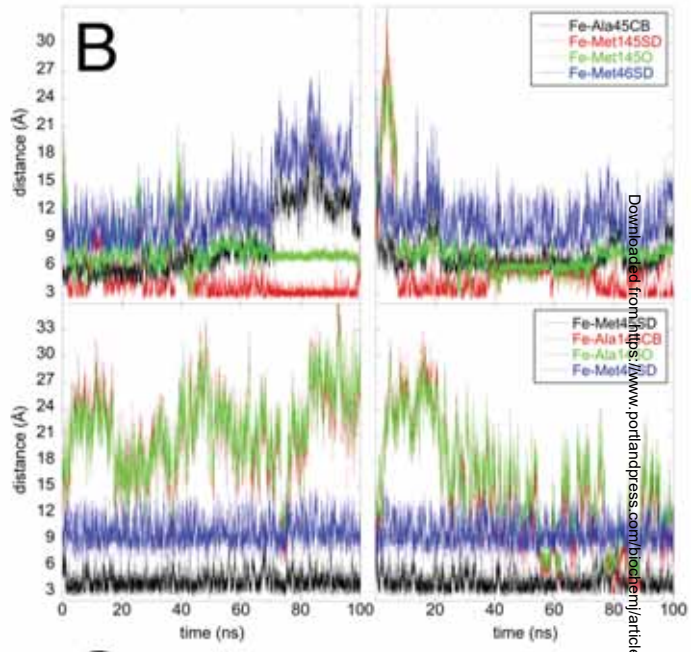


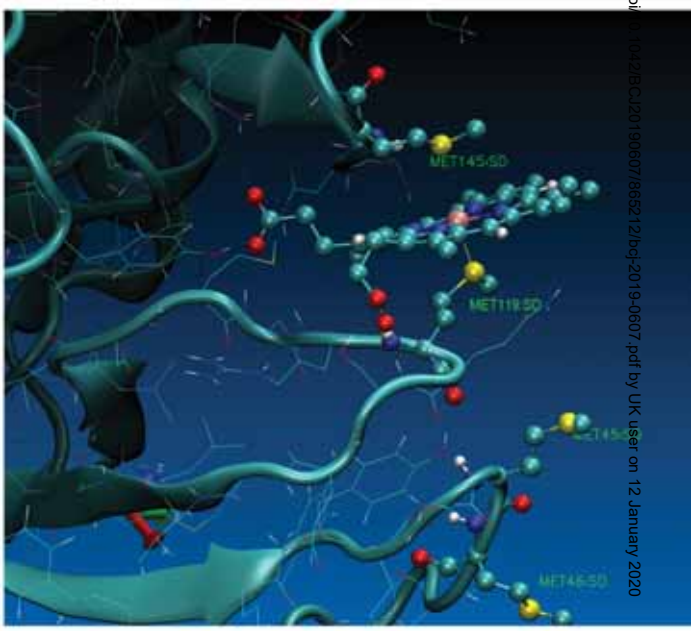


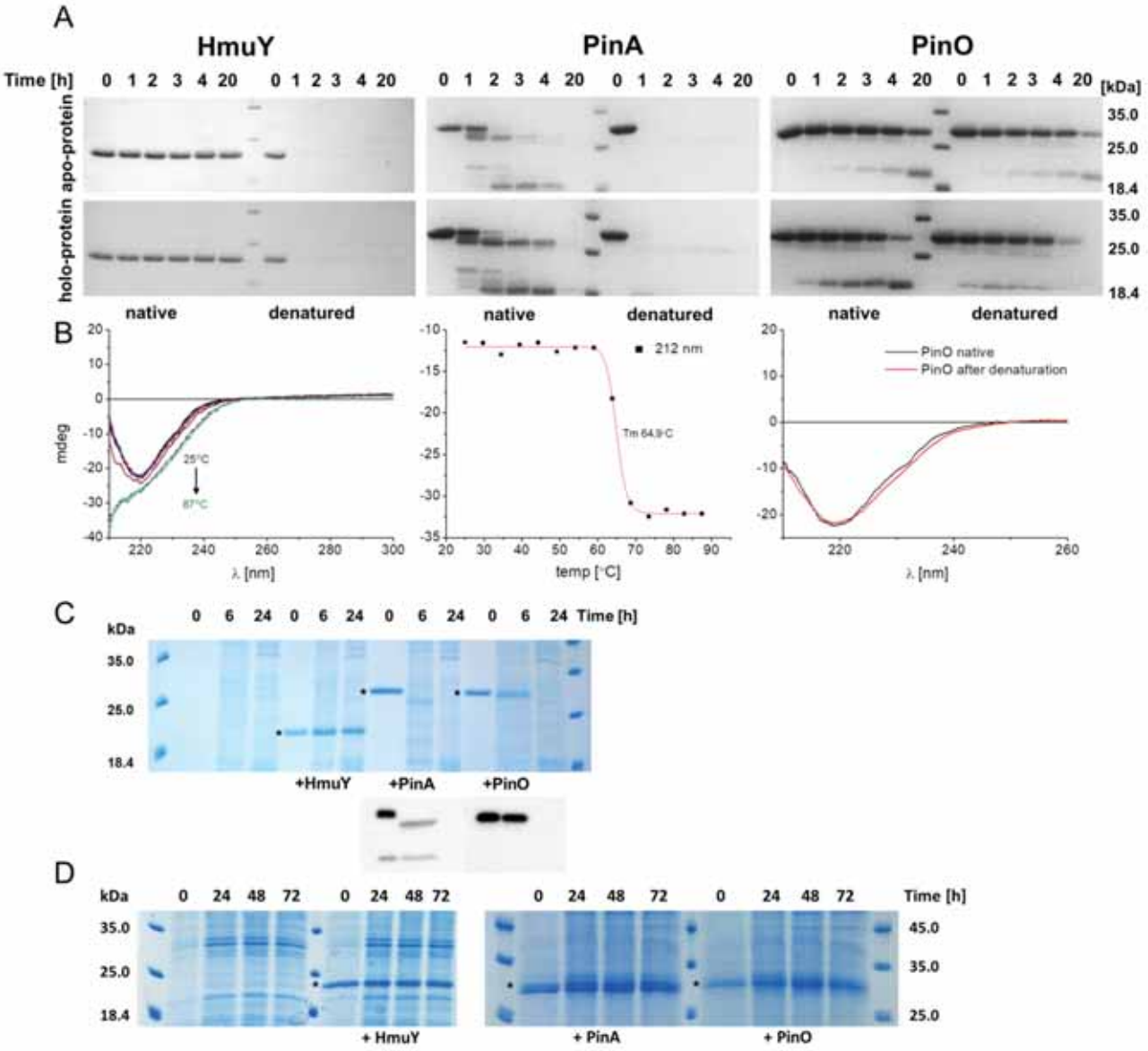


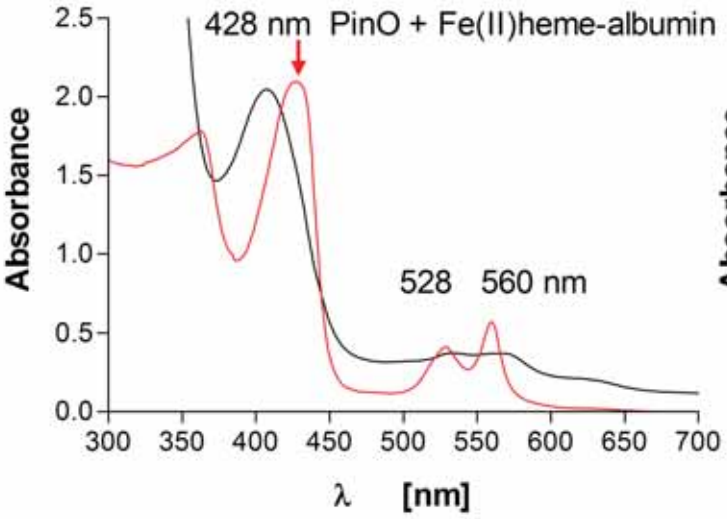


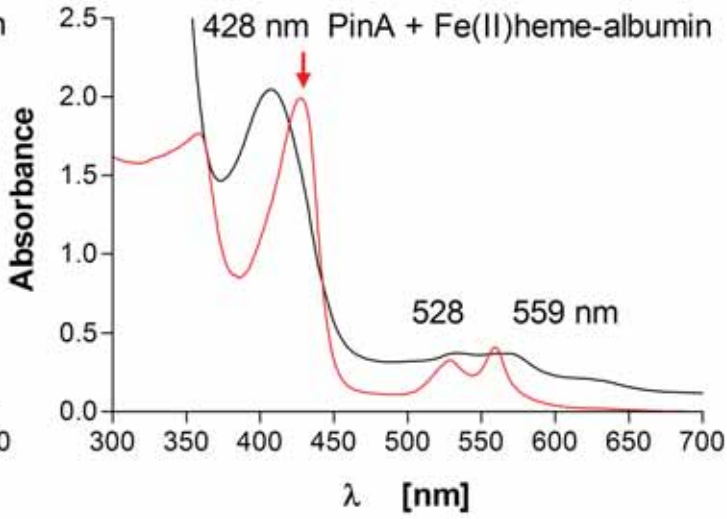


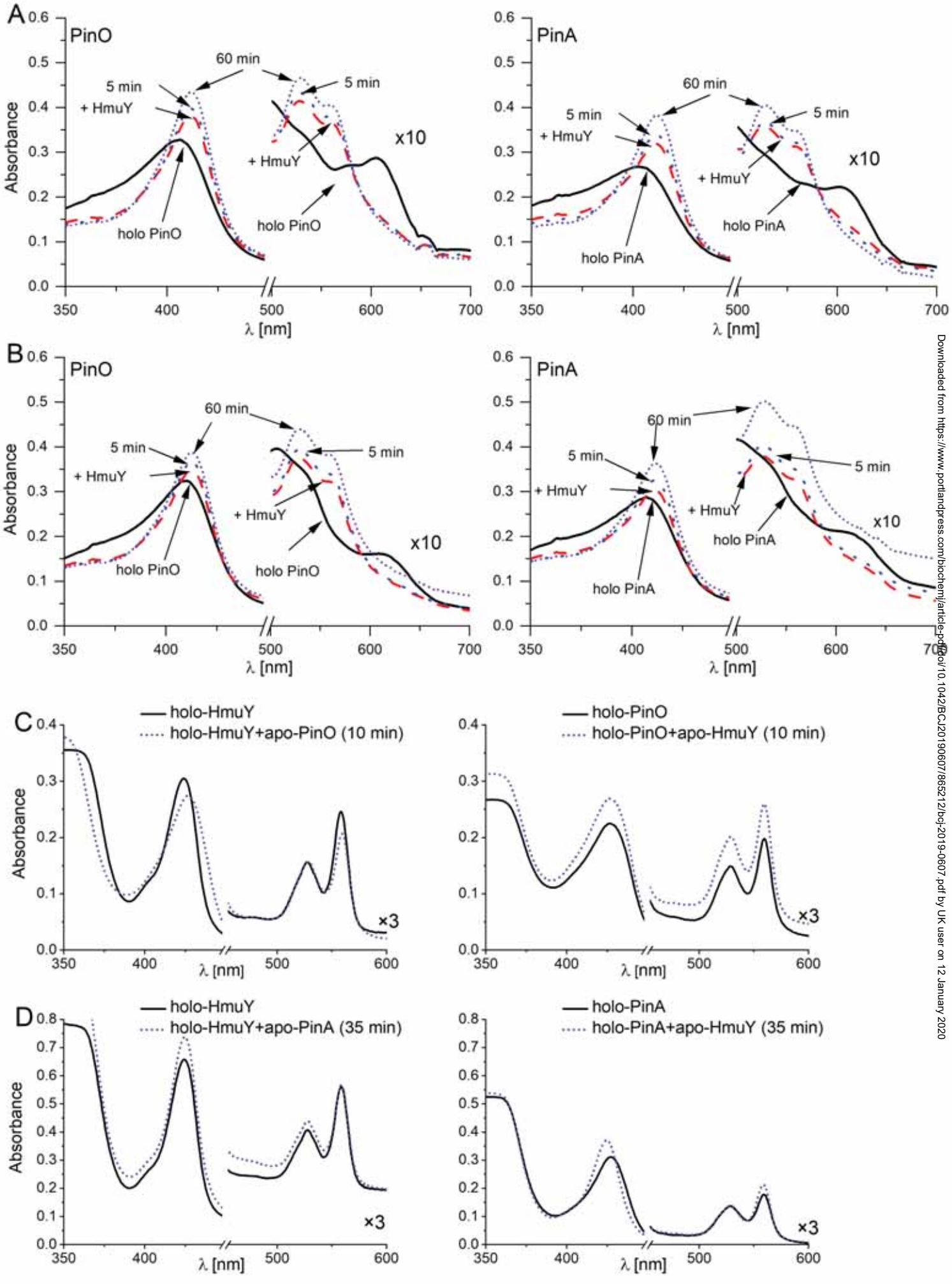


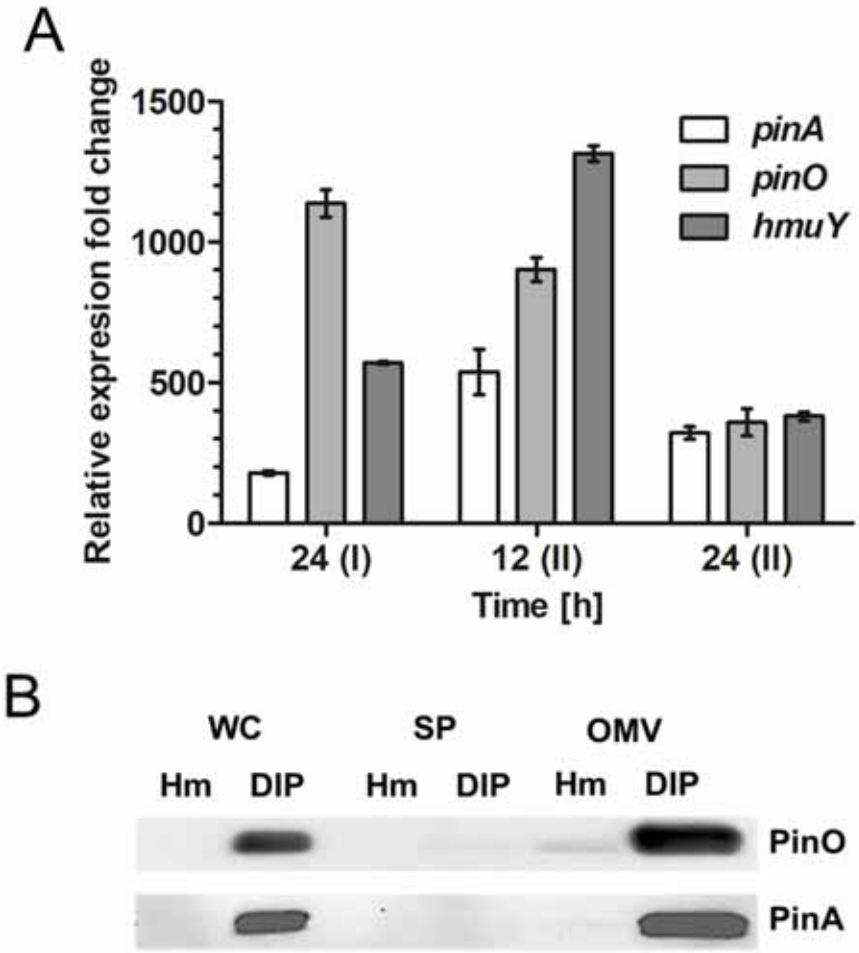












SUPPLEMENTARY TABLES AND FIGURES

Table S1. Primers used in this study.

Name	DNA sequence (5'→3')	Product length (bp)	Locus ID, gene abbreviation	Description
Fq_PinO_1 Rq_PinO_1	GACCCAAACCCCAAACCTGA ACTTCTTTGGCAGGTCAGCA	223	PIN17_RS00035, *pinO*	Amplify fragment of *P. intermedia pinO* gene (qRT-PCR)
FPinO_TTQ_Esp3I RPin0009_STOP(XhoI)	ATCTCGTCTCGCATGACCACGCAA GTAAAGCATTTCG GATCTCGAGTTACTTCGCTTTCTTT ATGAACTTATAG	606	PIN17_RS00035, *pinO*	Amplify truncated version of the gene encoding PinO protein, lacking N-terminal 34 amino-acid residues (for overexpression and purification of the protein for crystallization purposes)
Fq_PinA_1 Rq_PinA_1	TGTGAAATGGGTGAGCGACA AACGGCATAGCTGCCATCTT	229	PIN17_RS05355, *pinA*	Amplify fragment of *P. intermedia pinA* gene (qRT-PCR)
F_pMAL_Hc5x_PinO R_pMAL_Hc5x_PinO	CAACCTGGGATCGAGGGAAGGAT GAGCAAGGACAACAACGA CTTATTTAATTACCTGCAGGGAAT TCGGATCCTTACTTCGCTTTCTTTA TGAACTTATAGTC	678	PIN17_RS00035, *pinO*	Amplify *pinO* gene for overexpression and purification of recombinant N-terminally tagged with 6His and MBP PinO protein lacking signal peptide sequence
F_pMAL_Hc5x_PinA R_pMAL_Hc5x_PinA	CAACCTGGGATCGAGGGAAGGAT GAGCAATGATGACCCAACT TATTTAATTACCTGCAGGGAATTC GGATCCTTAATTCTTCTTGACAAA TTTATACTTGAA	736	PIN17_RS05355, *pinA*	Amplify *pinA* gene for overexpression and purification of recombinant N-terminally tagged with 6His and MBP PinA protein lacking signal peptide sequence
Fq_Pi_16SrRNA Rq_Pi_16SrRNA	GTAGGCGGTCTGTTAAGCGT GCATCCATCGTTTACCGTGC	249	PIN17_RS05870, *16S rRNA*	Amplify fragment of *P. intermedia 16S rRNA* gene (qRT-PCR)
HYq4_F HYq4_R	GCTTCGAAATACGAAACGTG TATATCCGTCTGTCGGAACG	214	PGA7_00004270, *hmuY*	Amplify fragment of *P. gingivalis hmuY* gene (qRT-PCR) (1)
16SrRNA-F 16SrRNA-R	CTTGACTTCAGTGGCGGCAG AGGGAAGACGGTTTTCACCA	378	PGA7_00019410, *16S rRNA*	Amplify fragment of *P. gingivalis 16S rRNA* gene (qRT-PCR) (1)

1. Gmiterek, A., Wojtowicz, H., Mackiewicz, P., Radwan-Oczko, M., Kantorowicz, M., Chomyszyn-Gajewska, M., Fraszczak, M., Bielecki, M., Olczak, M., and Olczak, T. (2013) The unique *hmuY* gene sequence as a specific marker of *Porphyromonas gingivalis* infection. *PLoS One* 8(7):e67719

Table S2. Data collection and refinement statistics for PinO.

Data collection	Value
Temperature	100K
Source	IO4-1 Diamond
Space group	P212121
Cell dimensions	
a, *b*, *c* (Å)	43.23 54.07 161.00
α, β, γ (°)	90 90 90
Resolution (Å)	80.5-2.46 (2.50-2.46)
No. reflections	14071
R_{sym} or R_{merge} [a]	0.135(0.93)
R_{pim}	0.057(0.338)
$I / \sigma I$	11.4(2.0)
CChalf	0.997(0.817)
Completeness (%)	97.5(100)
Redundancy	6.4(6.7)
Wilson B factor (Å^2)	36.6

Refinement	Value
Resolution (Å)	80.5-2.46
No. reflections	13422
R_{work} / R_{free} [b]	19.4/23.3
No. atoms	
Protein	2935
Water	164
Glycerol	18
B-factors	
Protein	42.64
Water	39.14
R.m.s. deviations	
Bond lengths (Å)	0.005
Bond angles (°)	1.223
PDB code	6R2H

[a] $R_{merge} = \Sigma|Ii - Im|/\Sigma Ii$, where Ii is the intensity of the measured reflection and Im is the mean intensity of all symmetry related reflections.
[b] $R_{free} = \Sigma T||Fobs| - |Fcalc||/\Sigma T|Fobs|$, where T is a test data set of about 5% of the total reflections randomly chosen and set aside prior to refinement.

Table S3. Accession numbers to the numbered sequences.

No.	Accession no.	Taxa
1	CCZ99607.1	Alistipes_inops;Alistipes_sp.;Rikenellaceae_(1)
2	CCX54992.1	Bacteroides_sp._(2)
3	WP_068341583.1	Porphyromonadaceae_bacterium_(3)
4	CDN31298.1	Mucinivorans_hirudinis_(4)
5	GAO31972.1	Geofilum_rubicundum_(5)
6	EIY43892.1	Bacteroides_nordii;Bacteroides_sp._(6)
7	OKZ04810.1	Bacteroides_sp._(7)
8	EIY56410.1	Bacteroides_ovatus_(8)
9	EHO68617.1	Prevotella_maculosa_(9)
10	WP_019968907.1	Prevotella_maculosa_(10)
11	WP_010258525.1	Alistipes_timonensis;Alistipes_sp._(11)
12	OUN59306.1	Alistipes_sp._(12)
13	OUN77541.1	Alistipes_sp._(13)
14	CDA95835.1	Bacteroides_sp._(14)
15	OKZ04391.1	Bacteroides_sp._(15)
16	KGF41531.1	Prevotella_buccalis;Prevotella_sp._(16)
17	EFA93136.1	Prevotella_buccalis_(17)
18	KGF34563.1	Prevotella_buccalis_(18)
19	CDA75122.1	Bacteroides_sp._(19)
20	EGG54865.1	Paraprevotella_xylaniphila_(20)
21	EHH01283.1	Paraprevotella_clara_(21)
22	CCZ03267.1	Paraprevotella_clara_(22)
23	KQM09118.1	Candidatus_Bacteroides_(23)
24	EFI36592.1	Bacteroides_fragilis;Bacteroides_ovatus;Bacteroides_xylanisolvens;Bacteroides_sp._(24)
25	EGM96972.1	Bacteroides_ovatus_(25)
26	EDO10376.1	Bacteroides_ovatus_(26)
27	EFS34243.1	Bacteroides_sp._(27)
28	EEO56990.1	Bacteroides_ovatus;Bacteroides_sp._(28)
29	EDM22026.1	Bacteroides_caccae_(29)
30	OKZ19420.1	Bacteroides_sp._(30)
31	CCZ74796.1	Bacteroides_caccae_(31)
32	CDA82979.1	Bacteroides_sp._(32)
33	WP_044653721.1	Bacteroides_acidifaciens_(33)
34	WP_029425604.1	Bacteroides_faecis;Bacteroides_thetaiotaomicron;Bacteroides_sp._(34)
35	AAO75604.1	Bacteroides_thetaiotaomicron;Bacteroides_sp._(35)
36	EFI02812.1	Bacteroides_sp._(36)
37	WP_072066292.1	Bacteroides_thetaiotaomicron_(37)
38	CUP68225.1	Bacteroides_thetaiotaomicron_(38)
39	KXT44720.1	Bacteroides_thetaiotaomicron_(39)
40	WP_048697409.1	Bacteroides_thetaiotaomicron_(40)
41	EFV67477.1	Bacteroides_vulgatus;Bacteroides_sp._(41)
42	EET16222.2	Bacteroides_vulgatus;Bacteroides_sp._(42)
43	CDF17081.1	Bacteroides_vulgatus_(43)
44	EEB24050.1	Bacteroides_dorei;Bacteroides_sp._(44)
45	EIY29775.1	Bacteroides_dorei_(45)

46	3U22_B	Bacteroides_vulgatus_(46)
47	WP_019540215.1	Proteiniphilum_acetatigenes_(47)
48	SFL25844.1	Porphyromonadaceae_bacterium_(48)
49	SCD19201.1	Proteiniphilum_saccharofermentans_(49)
50	EGN07227.1	Bacteroides_cellulosilyticus;Bacteroides_fragilis;Bacteroides_sp._(50)
51	EXY27161.1	Bacteroides_fragilis_(51)
52	EYA61037.1	Bacteroides_fragilis_(52)
53	BAD49437.1	Bacteroides_fragilis;Bacteroides_sp._(53)
54	EXZ62956.1	Bacteroides_fragilis_(54)
55	OCJ75647.1	Bacteroides_fragilis_(55)
56	CDD41062.1	Bacteroides_fragilis_(56)
57	EXY65222.1	Bacteroides_fragilis_(57)
58	EXZ05019.1	Bacteroides_fragilis_(58)
59	EKA91501.1	Bacteroides_fragilis;Bacteroides_sp._(59)
60	KER50729.1	Bacteroides_fragilis_(60)
61	EFR53559.1	Bacteroides_fragilis_(61)
62	EKA82307.1	Bacteroides_fragilis_(62)
63	4GBS_B	Bacteroides_fragilis_(63)
64	OKY88744.1	Bacteroidales_bacterium_(64)
65	OUN58423.1	Alistipes_sp._(65)
66	OUO20712.1	Alistipes_sp._(66)
67	OUN77671.1	Alistipes_sp._(67)
68	OUQ53512.1	Alistipes_sp._(68)
69	OUO36050.1	Muribaculum_sp._(69)
70	CCY36830.1	Alistipes_sp._(70)
71	OKZ02309.1	Bacteroides_sp._(71)
72	CCZ69026.1	Bacteroides_sp._(72)
73	OUN80451.1	Bacteroides_sp._(73)
74	OUO63022.1	Bacteroides_sp._(74)
75	OUP36170.1	Bacteroides_sp._(75)
76	OUO18330.1	Bacteroides_sp._(76)
77	EEF75672.1	Bacteroides_coprophilus_(77)
78	WP_051655754.1	Bacteroides__(78)
79	SDF58790.1	Bacteroidales_bacterium_(79)
80	EKX93307.1	Prevotella_sp._(80)
81	WP_075566499.1	Prevotellaceae_bacterium_(81)
82	SDH14636.1	Prevotella_sp._(82)
83	SDO24210.1	Prevotella_sp._(83)
84	SHL11200.1	Prevotella_ruminicola_(84)
85	ETD16427.1	Prevotella_oralis;Prevotella_sp._(85)
86	EEX17867.1	Prevotella_veroralis_(86)
87	WP_018910386.1	Prevotella_veroralis_(87)
88	EID32800.1	Prevotella_sp._(88)
89	EGC19502.1	Prevotella_multiformis_(89)
90	EGC86550.1	Prevotella_denticola_(90)
91	AEA20432.1	Prevotella_denticola_(91)

92	WP_025067727.1	Prevotella_denticola_(92)
93	EFC72054.1	Prevotella_melaninogenica_(93)
94	WP_036886950.1	Prevotella_melaninogenica_(94)
95	ADK95208.1	Prevotella_melaninogenica_(95)
96	ETS96241.1	Prevotella_sp._(96)
97	WP_025837568.1	Prevotella_scopos_(97)
98	EGW46597.1	Prevotella_sp._(98)
99	ERJ79263.1	Prevotella_sp._(99)
100	AKU69105.1	Prevotella_fusca_(100)
101	EGQ16555.1	Hallella_seregens;Prevotella_dentalis_(101)
102	WP_025002989.1	Prevotella_dentasini_(102)
103	EHG16195.1	Prevotella_histicola_(103)
104	WP_025792651.1	Prevotella_histicola_(104)
105	KGF25502.1	Prevotella_histicola_(105)
106	EHO65992.1	Prevotella_micans_(106)
107	EEX72705.1	Alloprevotella_tannerae_(107)
108	EEX53869.1	Prevotella_sp._(108)
109	WP_018968046.1	Prevotella_loescheii_(109)
110	EFC67243.1	Prevotella_sp._(110)
111	WP_044078386.1	Prevotella_sp._(111)
112	OFQ22241.1	Prevotella_sp._(112)
113	WP_019188856.1	Prevotella_conceptionensis_(113)
114	WP_025817488.1	Prevotella_shahii_(114)
115	WP_036891688.1	Prevotella_corporis_(115)
116	KXA38810.1	Prevotella_corporis_(116)
117	WP_060940719.1	Prevotella_corporis_(117)
118	EHJ41958.1	Prevotella_stercorea_(118)
119	CDE33319.1	Prevotella_stercorea_(119)
120	CDB04641.1	Prevotella_sp._(120)
121	CDA57371.1	Prevotella_sp._(121)
122	CDD05082.1	Prevotella_sp._(122)
123	KIP61358.1	Prevotella_sp._(123)
124	KIP54187.1	Prevotella_sp._(124)
125	KIP59830.1	Prevotella_sp._(125)
126	CDE86579.1	Prevotella_sp._(126)
127	EGN56289.1	Prevotella_multisaccharivorax_(127)
128	WP_028909645.1	Prevotella_sp._(128)
129	EFB31210.1	Prevotella_oris_(129)
130	EHO65065.1	Prevotella_maculosa_(130)
131	WP_019969145.1	Prevotella_maculosa_(131)
132	EFM01222.1	Prevotella_marshii_(132)
133	ERK00973.1	Prevotella_pleuritidis_(133)
134	WP_036888310.1	Prevotella_enoeca_(134)
135	WP_075559049.1	Parabacteroides_sp._(135)
136	EDS13961.1	Bacteroides_stercoris_(136)
137	CDA46594.1	Bacteroides_stercoris_(137)

138	WP_018668019.1	Bacteroides_gallinarum_(138)
139	ADV45155.1	Bacteroides_helcogenes_(139)
140	EEX71812.1	Alloprevotella_tannerae_(140)
141	AEW22472.1	Tannerella_forsythia_(141)
142	CEH11291.1	Tannerella_forsythia_(142)
143	SCQ23850.1	Tannerella_forsythia_(143)
144	SCQ22663.1	Tannerella_forsythia_(144)
145	BAR52082.1	Tannerella_forsythia_(145)
146	BAR49389.1	Tannerella_forsythia_(146)
147	ETK02786.1	Tannerella_sp._(147)
148	AOH40522.1	Tannerella_sp._(148)
149	ETK08917.1	Tannerella_sp._(149)
150	ETK06482.1	Tannerella_sp._(150)
151	WP_018360193.1	Porphyromonas_macacae_(151)
152	WP_052080461.1	Porphyromonas_macacae_(152)
153	EEX45726.1	Bacteroides_finegoldii_(153)
154	SCI24075.1	uncultured_Bacteroides_(154)
155	EKJ91405.1	Bacteroides_finegoldii_(155)
156	EFZ37334.1	Prevotella_oralis;Prevotella_sp._(156)
157	EFB32073.1	Prevotella_oris;Prevotella_sp._(157)
158	EGQ11734.1	Prevotella_nigrescens_(158)
159	ELX67767.1	Prevotella_nigrescens_(159)
160	OWP28965.1	Prevotella_nigrescens_(160)
161	EGQ14244.1	Prevotella_pallens_(161)
162	WP_025000693.1	Prevotella_aurantiaca_(162)
163	AFJ07542.1	Prevotella_intermedia_(163)
164	OWP31585.1	Prevotella_intermedia_(164)
165	KJJ86550.1	Prevotella_intermedia_(165)
166	WP_028905844.1	Prevotella_intermedia_(166)
167	WP_061869484.1	Prevotella_intermedia_(167)
168	BAU18907.1	Prevotella_intermedia_(168)
169	WP_024999526.1	Prevotella_falsenii_(169)
170	KGI61129.1	Prevotella_sp._(170)
171	KXB83406.1	Prevotella_sp._(171)
172	ERT59593.1	Prevotella_buccalis;Prevotella_sp._(172)
173	EFA97464.1	Prevotella_timonensis_(173)
174	WP_025072675.1	Prevotella_timonensis_(174)
175	KGF37557.1	Prevotella_buccalis_(175)
176	KXB38492.1	Bacteroidales_bacterium_(176)
177	KGF50177.1	Prevotella_melaninogenica_(177)
178	WP_077195474.1	Prevotella_ihumii_(178)
179	WP_009229011.1	Prevotella_sp._(179)
180	EFC70142.2	Prevotella_sp._(180)
181	EGQ14446.1	Prevotella_nigrescens_(181)
182	AFJ08449.1	Prevotella_intermedia_(182)
183	BAU17704.1	Prevotella_intermedia_(183)

184	WP_045167649.1	Prevotella_intermedia_(184)
185	EGQ23302.1	Prevotella_pallens_(185)
186	EHO67115.1	Prevotella_micans_(186)
187	EKA89362.1	Bacteroides_fragilis;Bacteroides_sp._(187)
188	EFR51743.1	Bacteroides_fragilis_(188)
189	KER58751.1	Bacteroides_fragilis_(189)
190	KER49675.1	Bacteroides_fragilis_(190)
191	AAD40726.1	Bacteroides_cellulosilyticus,Bacteroides_fragilis;Bacteroides_sp._(191)
192	AAD56752.1	Bacteroides_fragilis_(192)
193	EXY85552.1	Bacteroides_fragilis_(193)
194	CDD40776.1	Bacteroides_fragilis;Bacteroides_sp._(194)
195	EYA40265.1	Bacteroides_fragilis_(195)
196	EXY28431.1	Bacteroides_fragilis_(196)
197	EIY95461.1	Bacteroides_fragilis_(197)
198	EXZ06542.1	Bacteroides_fragilis_(198)
199	OJV83253.1	Bacteroidia_bacterium_(199)
200	KGL51490.1	Porphyromonas_canoris;Porphyromonas_sp._(200)
201	KGN66949.1	Porphyromonas_canoris;Porphyromonas_sp._(201)
202	WP_025837250.1	Porphyromonas_cangingivalis_(202)
203	KGL49707.1	Porphyromonas_cangingivalis_(203)
204	ETK00627.1	Tannerella_sp._(204)
205	ETK05806.1	Tannerella_sp._(205)
206	ETK03428.1	Tannerella_sp._(206)
207	AOH40157.1	Tannerella_sp._(207)
208	KQM09183.1	Candidatus_Bacteroides_(208)
209	EEK17166.1	Porphyromonas_uenonis_(209)
210	EFR35409.1	Porphyromonas_asaccharolytica_(210)
211	KXB33646.1	Bacteroidales_bacterium_(211)
212	WP_026327444.1	Proteiniphilum_acetatigenes;Porphyromonadaceae_bacterium_(212)
213	SDZ73806.1	Proteiniphilum_saccharofermentans;Porphyromonadaceae_bacterium_(213)
214	OJV34985.1	Bacteroidales_bacterium_(214)
215	WP_051290700.1	Dysgonomonas_capnocytophagoides_(215)
216	WP_019541147.1	Proteiniphilum_acetatigenes_(216)
217	SFK49338.1	Porphyromonadaceae_bacterium_(217)
218	KGL49149.1	Porphyromonas_gulae;Porphyromonas_sp._(218)
219	KGO02733.1	Porphyromonas_gulae_(219)
220	ABL74281.1	Porphyromonas_gingivalis;Porphyromonas_gulae;Porphyromonas_sp._(220)
221	3H8T_B	Porphyromonas_gingivalis_(221)
222	SJL32660.1	Porphyromonas_gingivalis_(222)
223	CAM31898.1	Porphyromonas_gingivalis_(223)
224	CAM31897.1	Porphyromonas_gingivalis_(224)
225	SJL33096.1	Porphyromonas_gingivalis_(225)
226	ALJ25004.1	Porphyromonas_gingivalis_(226)
227	EEN82265.1	Porphyromonas_endodontalis_(227)
228	WP_040581261.1	Porphyromonas_endodontalis_(228)
229	KGN97374.1	Porphyromonas_gingivicanis_(229)

230	SJZ84111.1	Porphyromonas_circumdentaria_(230)
231	WP_083377505.1	Millionella_massiliensis_(231)
232	WP_084135281.1	Rikenella_microfusus_(232)
233	EEK16193.1	Porphyromonas_uenonis_(233)
234	EFR34766.1	Porphyromonas_asaccharolytica_(234)
235	AEE13138.1	Porphyromonas_asaccharolytica_(235)
236	WP_025883254.1	Porphyromonas_uenonis_(236)
237	EEK17583.1	Porphyromonas_uenonis_(237)
238	WP_027451380.1	Porphyromonas_uenonis_(238)
239	EFR34959.1	Porphyromonas_asaccharolytica_(239)
240	KXB36609.1	Bacteroidales_bacterium_(240)
241	ERJ70310.1	Porphyromonas_sp._(241)
242	WP_052405356.1	Porphyromonas_sp._(242)
243	WP_052408920.1	Porphyromonas_sp._(243)
244	ADY33429.1	Odoribacter_splanchnicus;Odoribacter_sp._(244)
245	OUN66721.1	Butyricimonas_sp._(245)
246	KKB56668.1	Parabacteroides_gordonii_(246)
247	WP_082211603.1	Parabacteroides_sp._(247)
248	WP_018358608.1	Porphyromonas_levii_(248)
249	SDZ75040.1	Porphyromonadaceae_bacterium_(249)
250	WP_052110816.1	Porphyromonadaceae_bacterium_(250)
251	WP_081961801.1	Porphyromonadaceae_bacterium_(251)
252	WP_018337986.1	Butyricimonas__(252)
253	WP_051465760.1	Butyricimonas_virosa_(253)
254	OKZ18475.1	Butyricimonas_synergistica_(254)
255	OUN66843.1	Butyricimonas_sp._(255)
256	OOD28628.1	Bacteroides_fragilis;Bacteroides_sp._(256)
257	BAD48850.1	Bacteroides_fragilis_(257)
258	WP_080696530.1	Bacteroides_fragilis_(258)
259	WP_083790870.1	Bacteroides_sp._(259)
260	CAH07859.1	Bacteroides_cellulosilyticus;Bacteroides_fragilis;Bacteroides_sp._(260)
261	EIY90254.1	Bacteroides_fragilis;Bacteroides_sp._(261)
262	EYA38788.1	Bacteroides_fragilis_(262)
263	EXY46540.1	Bacteroides_fragilis_(263)
264	EXY51337.1	Bacteroides_fragilis_(264)
265	EEZ27613.1	Bacteroides_sp._(265)
266	EDV01216.1	Bacteroides_caccae;Bacteroides_coprocola;Bacteroides_fragilis; Bacteroides_ovatus;Bacteroides_thetaiotaomicron;Bacteroides_sp._(266)
267	WP_080783883.1	Bacteroides__(267)
268	AHF11876.1	Barnesiella_viscericola_(268)
269	WP_025277678.1	Barnesiella_viscericola_(269)
270	ALA76171.1	Bacteroides_dorei_(270)
271	EFR52434.1	Bacteroides_fragilis_(271)
272	CCZ39697.1	Bacteroides_fragilis_(272)
273	WP_080715515.1	Bacteroides_fragilis;Bacteroides_sp._(273)
274	EKA89913.1	Bacteroides_fragilis_(274)
275	KER57755.1	Bacteroides_fragilis_(275)

276	WP_080714222.1	Bacteroides_fragilis_(276)
277	ADV43822.1	Bacteroides_helcogenes_(277)
278	WP_042369327.1	Bacteroides_neonati_(278)
279	EFC76042.1	Prevotella_buccae;Prevotella_sp._(279)
280	EFB92697.1	Prevotella_bivia_(280)
281	EFC70885.1	Prevotella_histicola;Prevotella_scopos;Prevotella_sp._(281)
282	WP_025064743.1	Prevotella_disiens_(282)
283	KGF46523.1	Prevotella_disiens_(283)
284	WP_077195481.1	Prevotella_ihumii_(284)
285	KGI59770.1	Prevotella_sp._(285)
286	KXB84501.1	Prevotella_sp._(286)
287	EFL46544.1	Prevotella_disiens_(287)
288	WP_025071823.1	Prevotella_timonensis_(288)
289	KGF33794.1	Prevotella_buccalis_(289)
290	ERK39051.1	Prevotella_baroniae_(290)
291	WP_025880170.1	Prevotella_baroniae_(291)
292	EFB31294.1	Prevotella_oris;Prevotella_sp._(292)
293	EFV03529.1	Prevotella_salivae_(293)
294	EGV34132.1	Prevotella_oulorum_(294)
295	ERK02099.1	Prevotella_salivae_(295)
296	WP_025070447.1	Prevotella_oulorum_(296)
297	WP_028899033.1	Prevotella_sp._(297)
298	EHG23459.1	Alloprevotella_rava_(298)
299	EFN91352.1	Prevotella_amnii_(299)
300	KXB79610.1	Prevotella_amnii_(300)
301	WP_019036574.1	Prevotella_amnii_(301)
302	EFC70735.1	Prevotella_sp._(302)
303	WP_018363165.1	Prevotella_nanceiensis_(303)
304	EFM02307.1	Prevotella_marshii_(304)
305	ERT60382.1	Prevotella_buccalis;Prevotella_sp._(305)
306	KGF37515.1	Prevotella_buccalis_(306)
307	ERT58548.1	Prevotella_buccalis;Prevotella_timonensis;Prevotella_sp._(307)
308	KXB47685.1	Bacteroidales_bacterium_(308)
309	KGF19162.1	Prevotella_sp._(309)
310	WP_025065553.1	Prevotella_enoeca;Prevotella_oralis;Prevotella_salivae;Prevotella_sp._(310)
311	WP_025070206.1	Prevotella_oulorum;Prevotella_sp._(311)
312	EGN57860.1	Prevotella_multisaccharivorax_(312)
313	EKU92319.1	Bacteroides_oleiciplenus_(313)
314	EFA44395.1	Prevotella_bergensis;Prevotella_sp._(314)
315	EGC87335.1	Prevotella_denticola;Prevotella_sp._(315)
316	EFI70933.1	Prevotella_bryantii_(316)
317	CDB05899.1	Prevotella_sp._(317)
318	WP_077154031.1	Bacteroides_sp._(318)
319	KGF43557.1	Prevotella_melaninogenica_(319)
320	WP_068856409.1	Prevotella_sp._(320)
321	CCY01465.1	Prevotella_sp._(321)

322	KGN86564.1	Porphyromonas_sp._(322)
323	WP_036854131.1	Porphyromonas_cangingivalis_(323)
324	WP_081963905.1	Porphyromonas_cangingivalis_(324)
325	WP_081968484.1	Porphyromonas_cangingivalis_(325)
326	GAD06226.1	Porphyromonas_crevioricanis_(326)
327	KGL50276.1	Porphyromonas_cangingivalis_(327)
328	WP_081963718.1	Porphyromonas_cangingivalis_(328)
329	KGN68075.1	Porphyromonas_sp._(329)
330	SJZ33180.1	Porphyromonas_cangingivalis_(330)
331	WP_081968472.1	Porphyromonas_cangingivalis_(331)
332	WP_072517548.1	Porphyromonadaceae_bacterium_(332)
333	WP_052079399.1	Porphyromonas_canoris;Porphyromonas_sp._(333)
334	WP_081966368.1	Porphyromonas_sp._(334)
335	EKX93557.1	Prevotella_sp._(335)
336	WP_075566865.1	Prevotellaceae_bacterium_(336)
337	EFB31221.1	Prevotella_oris;Prevotella_sp._(337)
338	KGI60325.1	Prevotella_sp._(338)
339	EFM01199.1	Prevotella_marshii_(339)
340	EHO65056.1	Prevotella_maculosa_(340)
341	WP_018361046.1	Porphyromonas_macacae_(341)
342	WP_052080488.1	Porphyromonas_macacae_(342)
343	WP_052081901.1	Porphyromonas_macacae_(343)
344	KGL48186.1	Porphyromonas_cangingivalis_(344)
345	KGN78482.1	Porphyromonas_cangingivalis_(345)
346	SJZ46749.1	Porphyromonas_cangingivalis_(346)
347	KGL51180.1	Porphyromonas_canoris_(347)
348	KGN93615.1	Porphyromonas_canoris_(348)
349	KGN68558.1	Porphyromonas_sp._(349)
350	WP_083428898.1	Porphyromonadaceae_bacterium_(350)
351	KXB32812.1	Bacteroidales_bacterium_(351)
352	GAP71374.1	Candidatus_Symbiothrix_(352)
353	CDN31206.1	Mucinivorans_hirudinis_(353)
354	CDB07867.1	Odoribacter_splanchnicus;Odoribacter_sp._(354)
355	EDS02374.1	Alistipes_putredinis_(355)
356	CDE63967.1	Alistipes_putredinis_(356)
357	CVI67324.1	Alistipes_sp._(357)
358	EGJ99335.1	Dysgonomonas_gadei_(358)
359	SBW06587.1	uncultured_Dysgonomonas_(359)
360	WP_026626689.1	Dysgonomonas_capnocytophagoides_(360)
361	WP_062178748.1	Dysgonomonas_macrotermitis_(361)
362	WP_050708967.1	Dysgonomonas_sp._(362)
363	OJX91808.1	Paludibacter_sp._(363)
364	WP_088656163.1	Geofilum_sp._(364)
365	OAV63999.1	Bacteroidales_bacterium_(365)
366	OAV65255.1	Bacteroidales_bacterium_(366)
367	WP_066188549.1	Bacteroidales_bacterium_(367)

368	WP_026474971.1	Alkaliflexus_imshenetskii_(368)
369	GAP43010.1	Lentimicrobium_saccharophilum_(369)

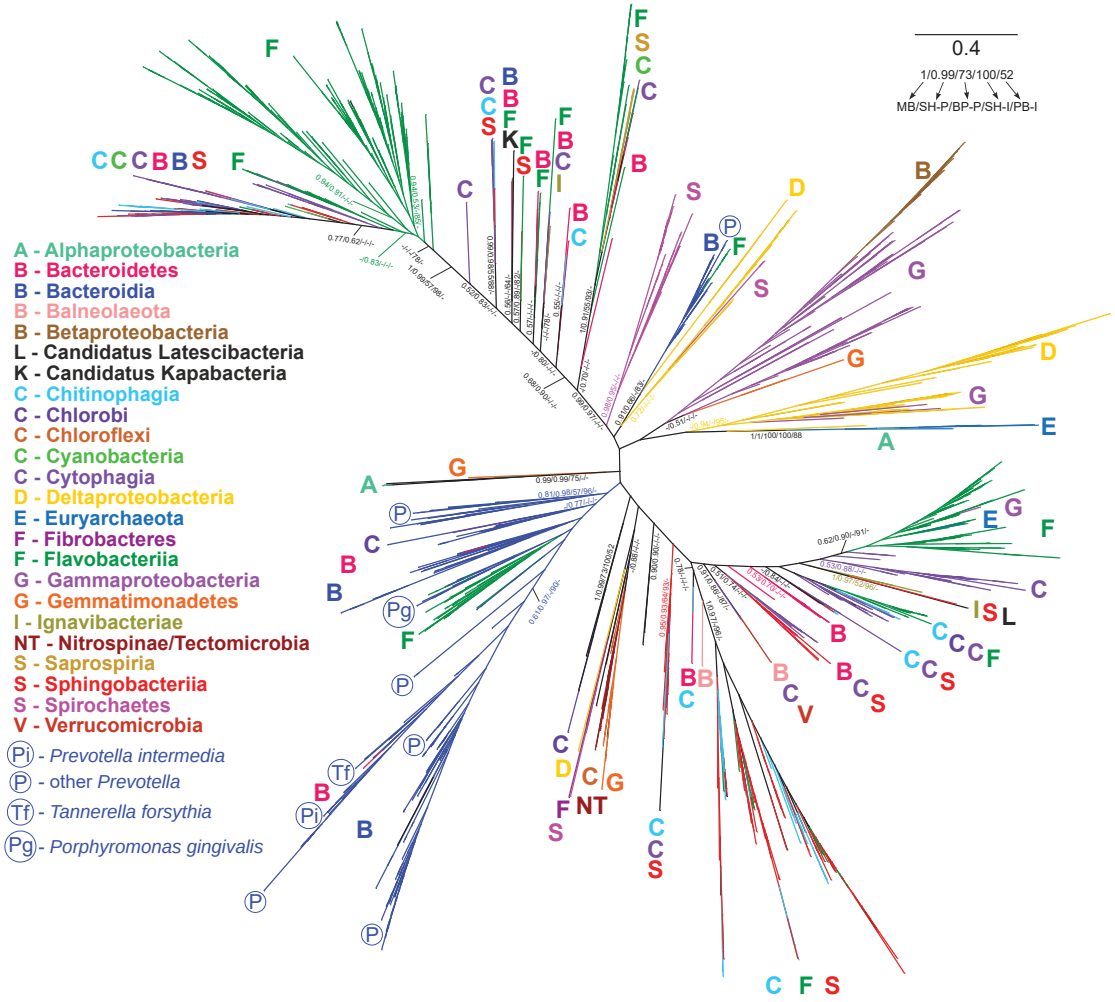


Figure S1. Phylogenetic tree obtained in MrBayes for the HmuY-like family. Main bacterial lineages are marked in different colors and indicated by the first letter of their name. The values at nodes indicate: posterior probabilities found in MrBayes (MB) as well as support values calculated in (more)PhyML and IQ-TREE by approximate likelihood-ratio test (aLRT) based on a Shimodaira-Hasegawa-like procedure (SH-P and SH-I, respectively) and non-parametric bootstrap (BP-P and BP-I, respectively). The posterior probabilities < 0.5 and the percentages < 50% are omitted or indicated by a dash "-". Pg, *Porphyromonas gingivalis*; Tf, *Tannerella forsythia*; Pi, *Prevotella intermedia*, P, *Prevotella*.

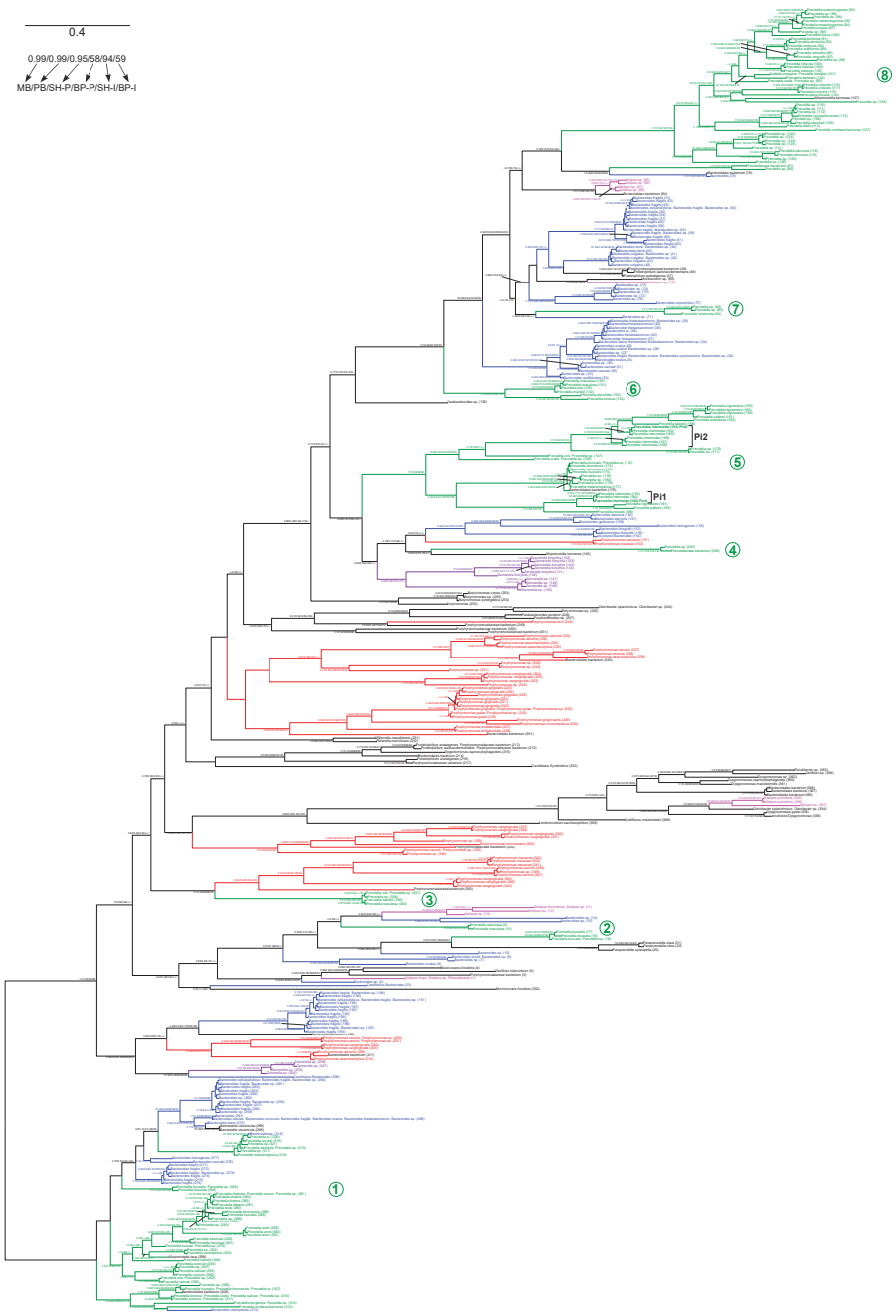


Figure S2. Phylogenetic tree obtained in MrBayes for the HmuY homologs in Bacteroidia. Eight *Prevotella* clades are numbered. Two *P. intermedia* groups, Pi1 and Pi2, are marked. The position of PinA and PinO sequences is indicated. The values at nodes indicate: posterior probabilities found in MrBayes (MB) and PhyloBayes (PB) as well as support values calculated in (more)PhyML and IQ-TREE by approximate likelihood-ratio test (aLRT) based on a Shimodaira-Hasegawa-like procedure (SH-P and SH-I, respectively) and non-parametric bootstrap (BP-P and BP-I, respectively). The posterior probabilities < 0.5 and the percentages < 50% are omitted or indicated by a dash "-".The accession numbers to the numbered sequences are included in Table S3.

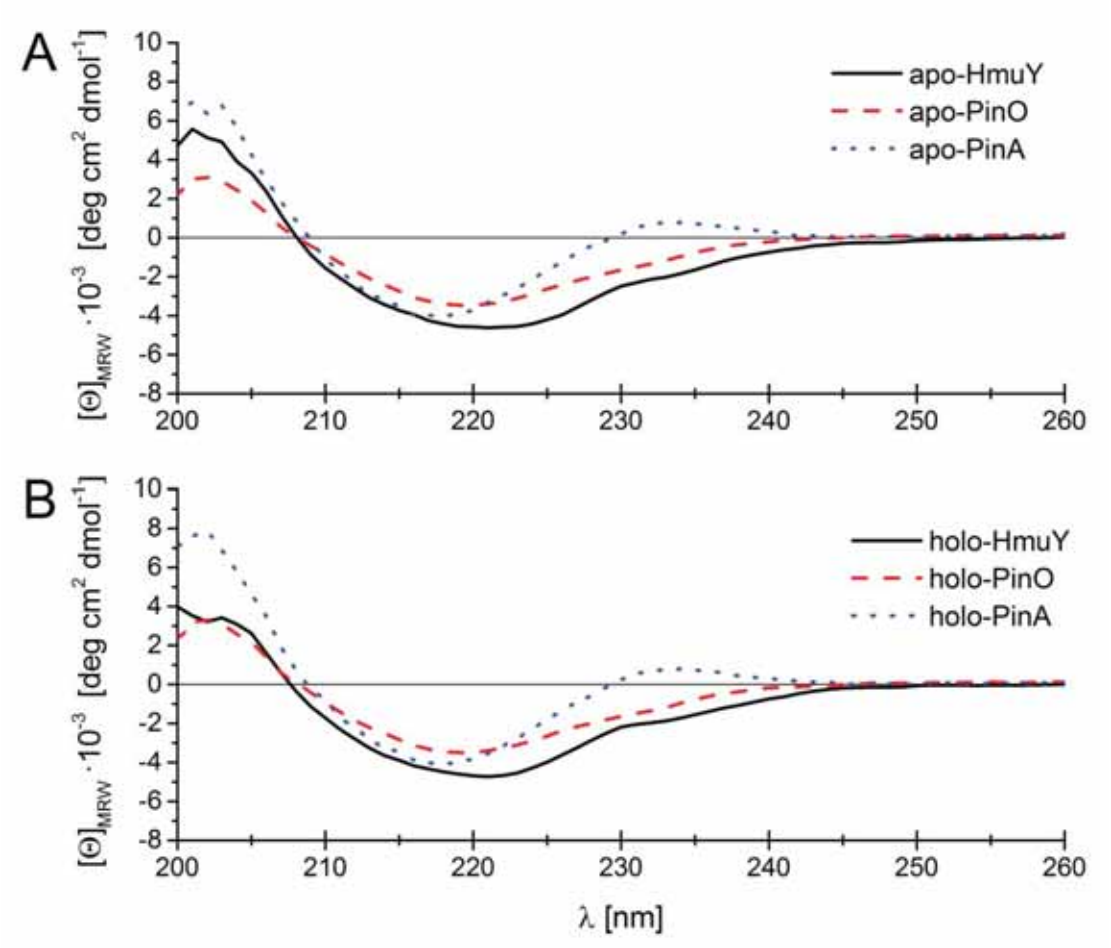


Figure S3. Analysis of secondary structure. Far-UV CD spectra of *P. gingivalis* HmuY, *P. intermedia* PinO and PinA in apo- (A) and holo-forms (B).

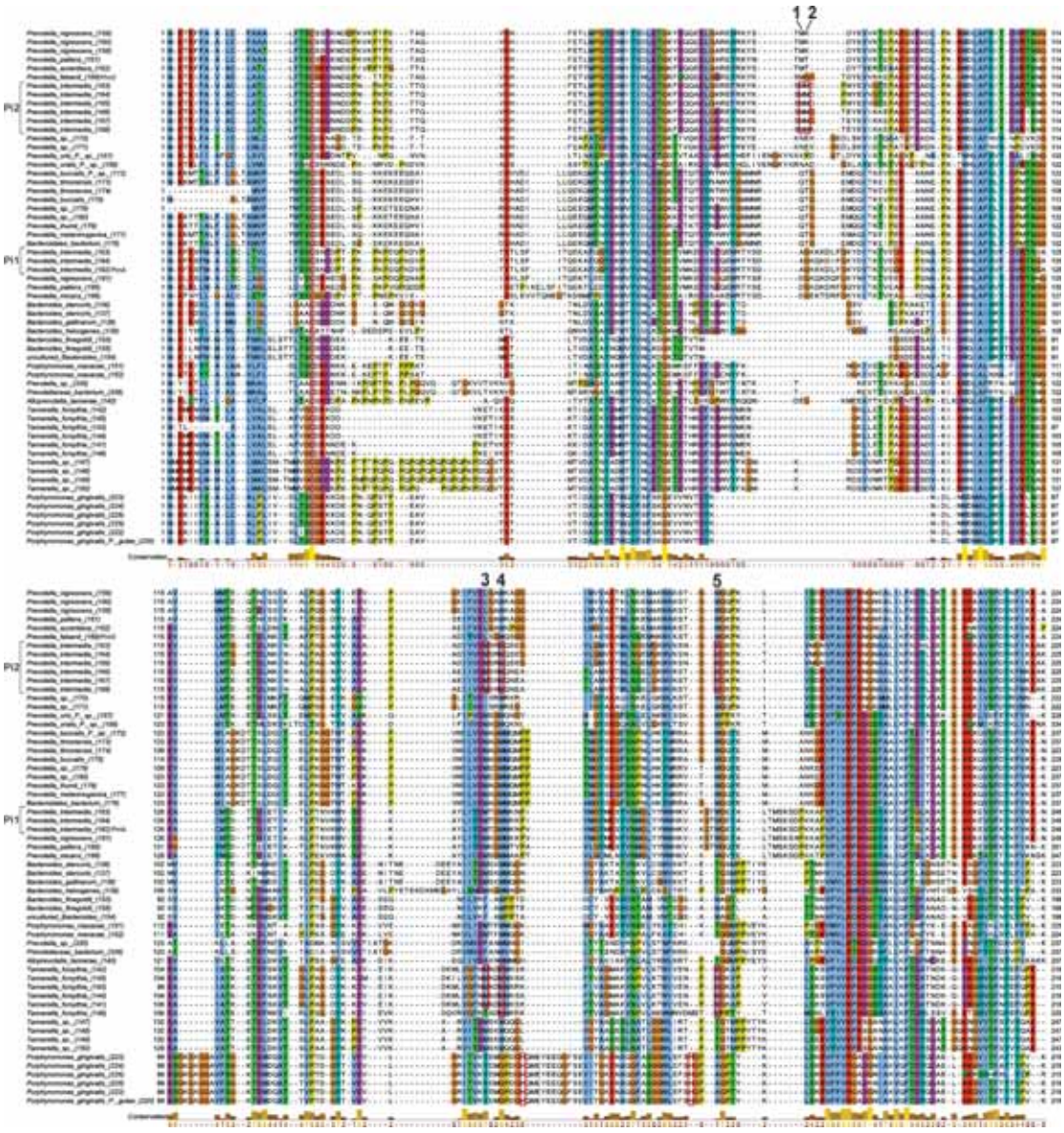


Figure S4. Amino-acid sequence alignment of Bacteroidia from clade 5 presented in Figure 2. Two *P. intermedia* groups, Pi1 and Pi2 are marked. Columns with methionine residues potentially coordinating heme iron in *T. forsythia* Tfo and *P. intermedia* PinO are shown by numbers and correspond to positions of amino acid residues in solved or modeled three-dimensional protein structures (numbering of amino acid residues in the full length protein sequences is shown in parentheses): 1 - PinO Met45 (Met76); 2 -PinO Met46 (Met77); 3 - PinO Met116 (Met147), PinA Met126 (Met159); 4 - PinO Met119 (Met150), PinA Met129 (Met162); 5 - PinO Met145 (Met176), PinA Met158 (Met191). Amino acid residues which coordinate heme iron in *P. gingivalis* HmuY and potentially coordinate heme iron in PinO and PinA are outlined by red boxes. The accession numbers to the numbered sequences are included in Table S3.

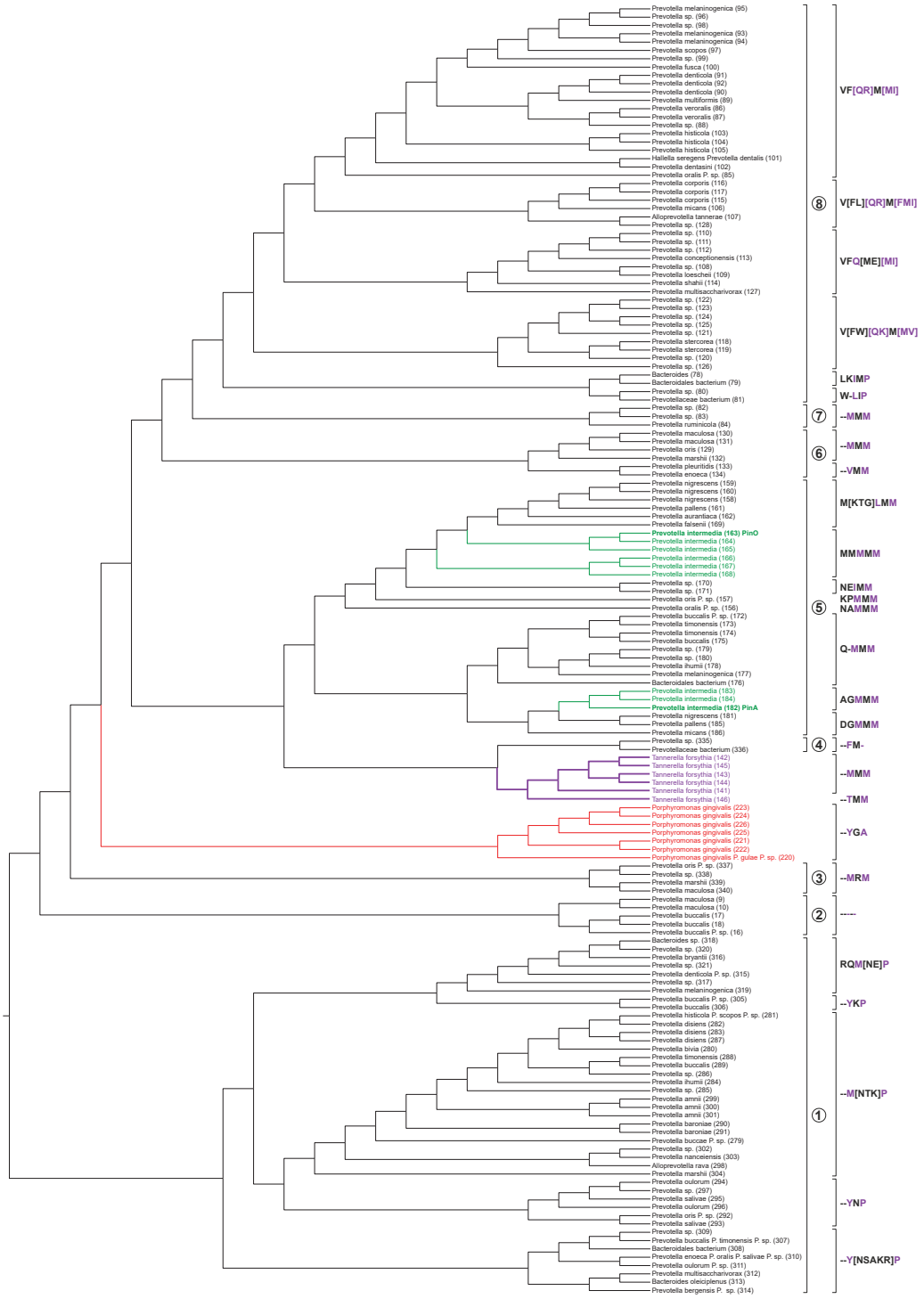


Figure S5. Cladogram of clades containing *Prevotella* sequences. The sequences of *P. gingivalis* and *T. forsythia* are included for comparison. Eight *Prevotella* clades are numbered. Amino acid residues homologous to sites involved in heme binding in PinO are shown. Residues corresponding to sites coordinating heme in Tfo are in purple. The accession numbers to the numbered sequences are included in Table S3.

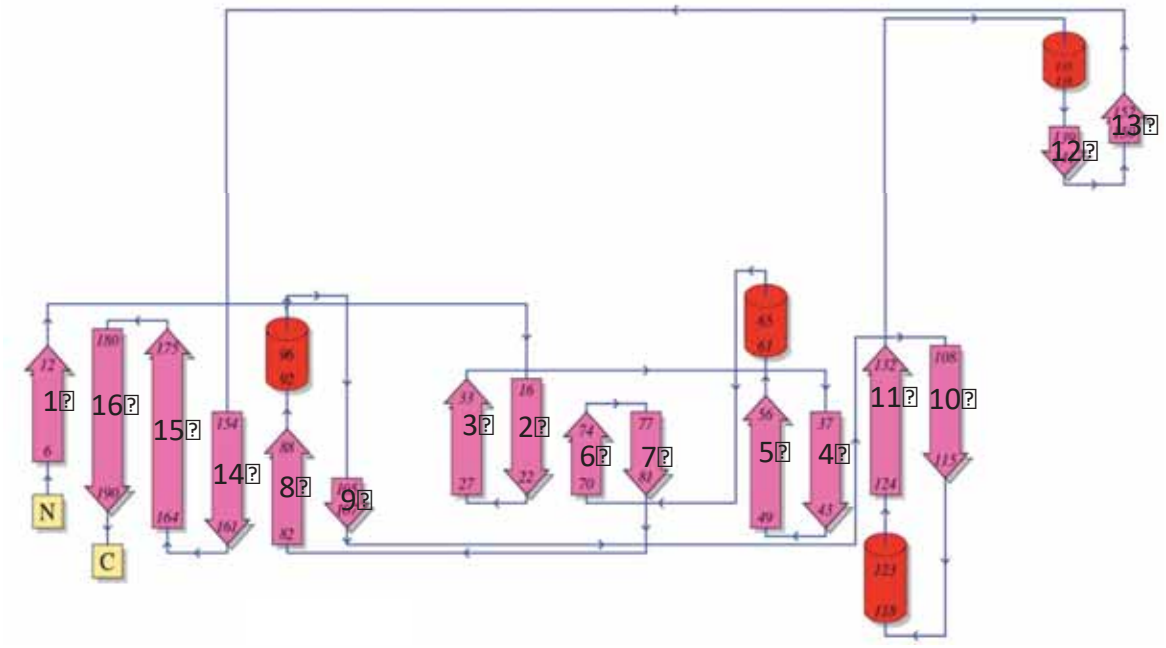


Figure S6. Topological scheme for *P. intermedia* PinO protein structure. Schematic presentation is shown based on crystal structure of apo-PinO (PDB ID: 6R2H).

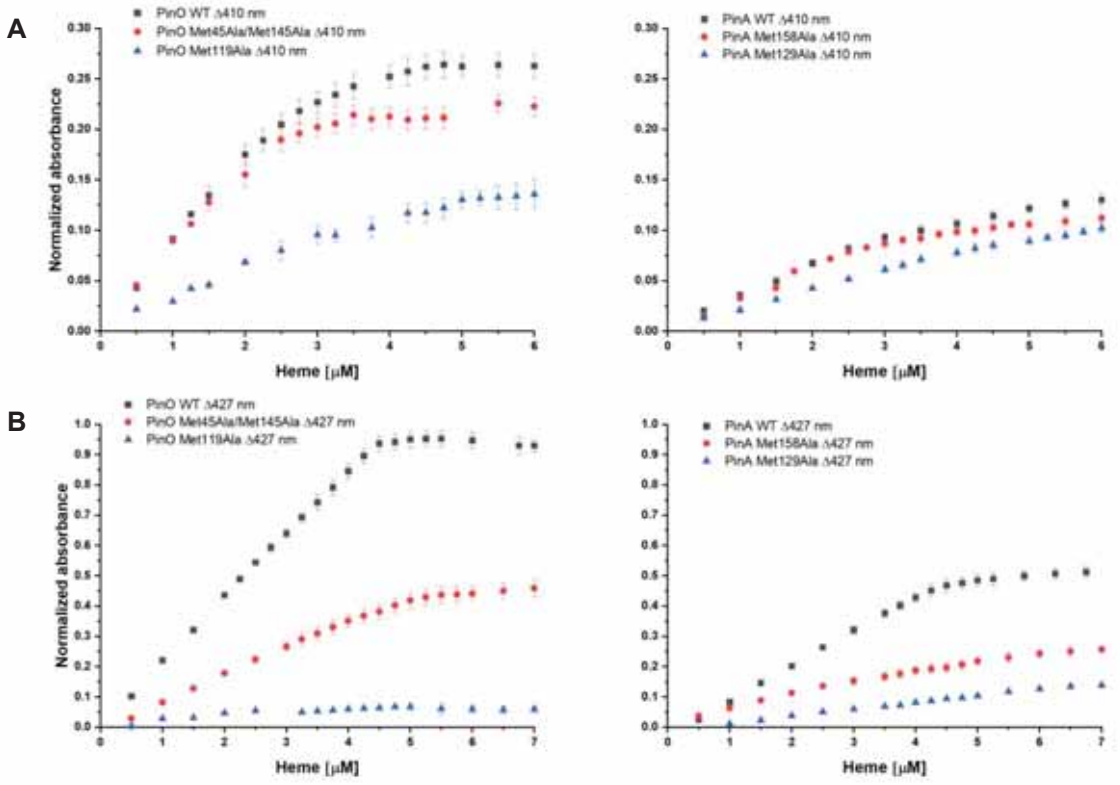


Figure S7. Heme titration experiments of selected *P. intermedia* PinO and PinA site-directed mutagenesis variants. The curves were generated after titration of 5 μM protein samples with heme by measuring the difference spectra between the protein+heme and heme-only samples under oxidizing (**A**) or reducing conditions formed by sodium dithionite (**B**). Results are shown as mean ± SD from 3 independent experiments. WT, wild type proteins.

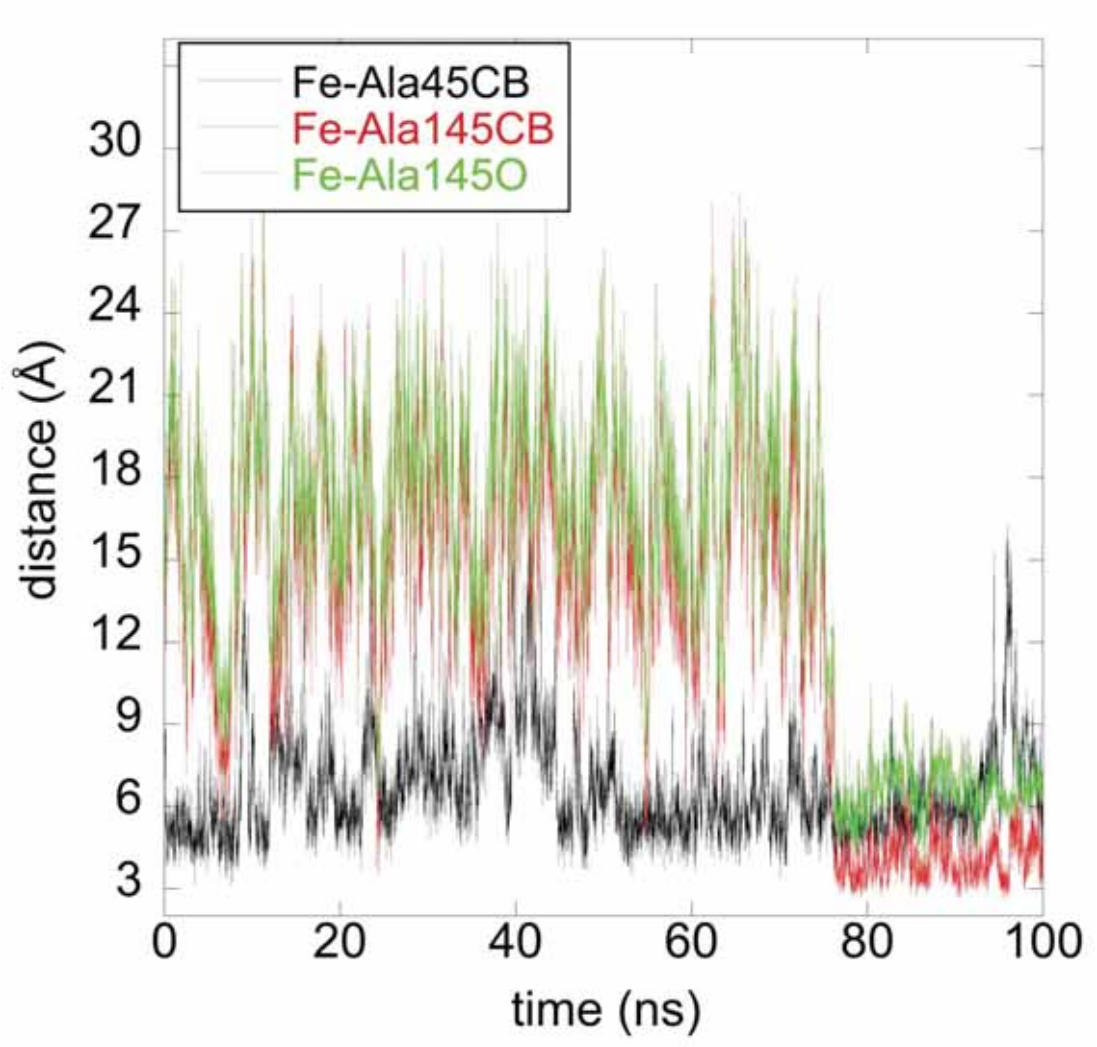


Figure S8. **Time evolution of the distance between the Fe atom and sidechain atoms of the double mutant Met45Ala/Met145 of heme-Met119 models based on the apo-PinO crystal structure.** One simulation is shown to demonstrate the presence of these loops in the vicinity of the heme.

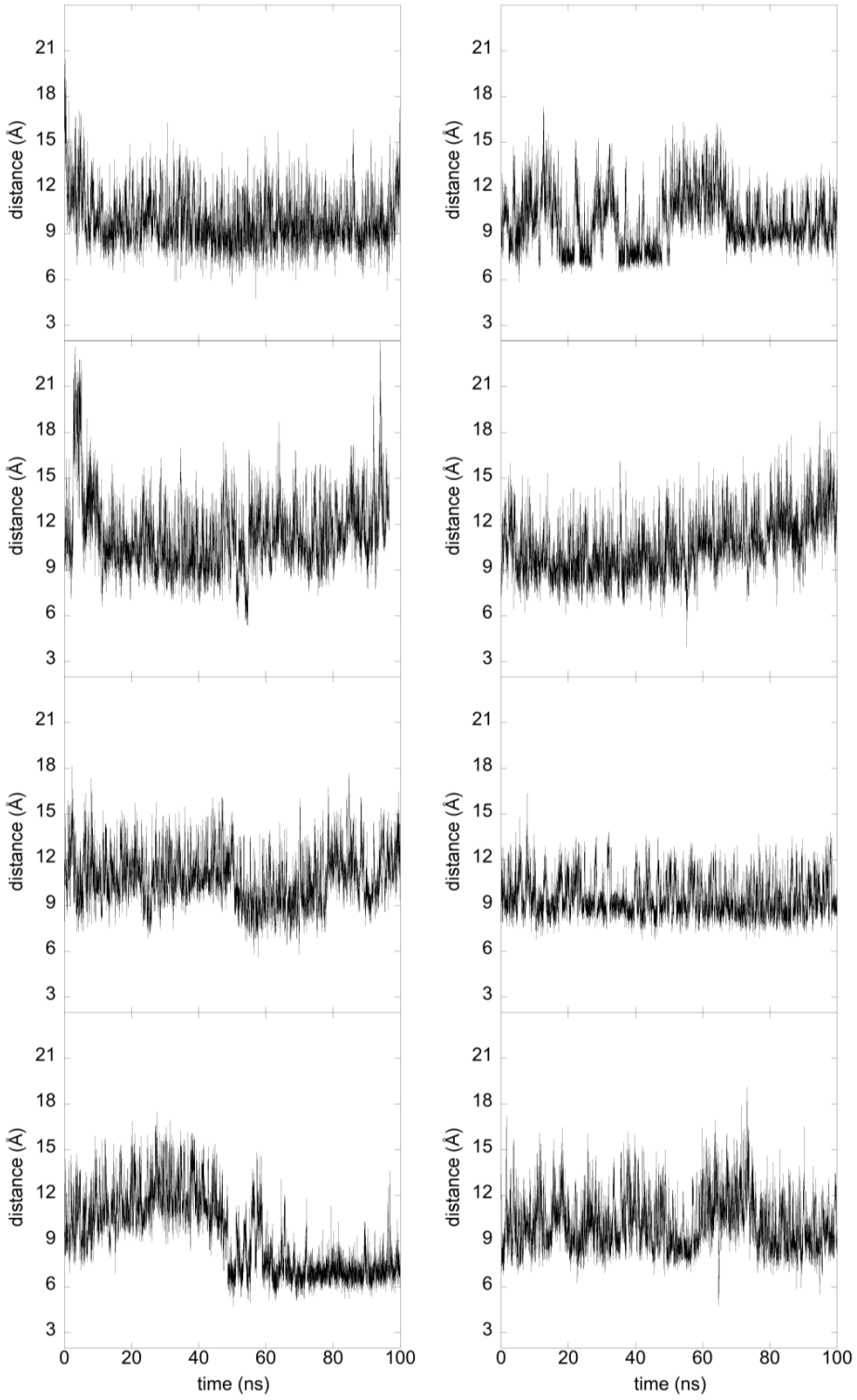


Figure S9. **Time evolution of the distance between the Fe atom and the sulphur atom of Met46 using the heme-Met119 constructs of PinO shown in Figure 10.** Eight independent MD simulations to 100 ns are presented and show that the Met46 sidechain sulphur atom is located at an average distance of ~10 Å from the Fe atom, with its closest approach ~5Å.

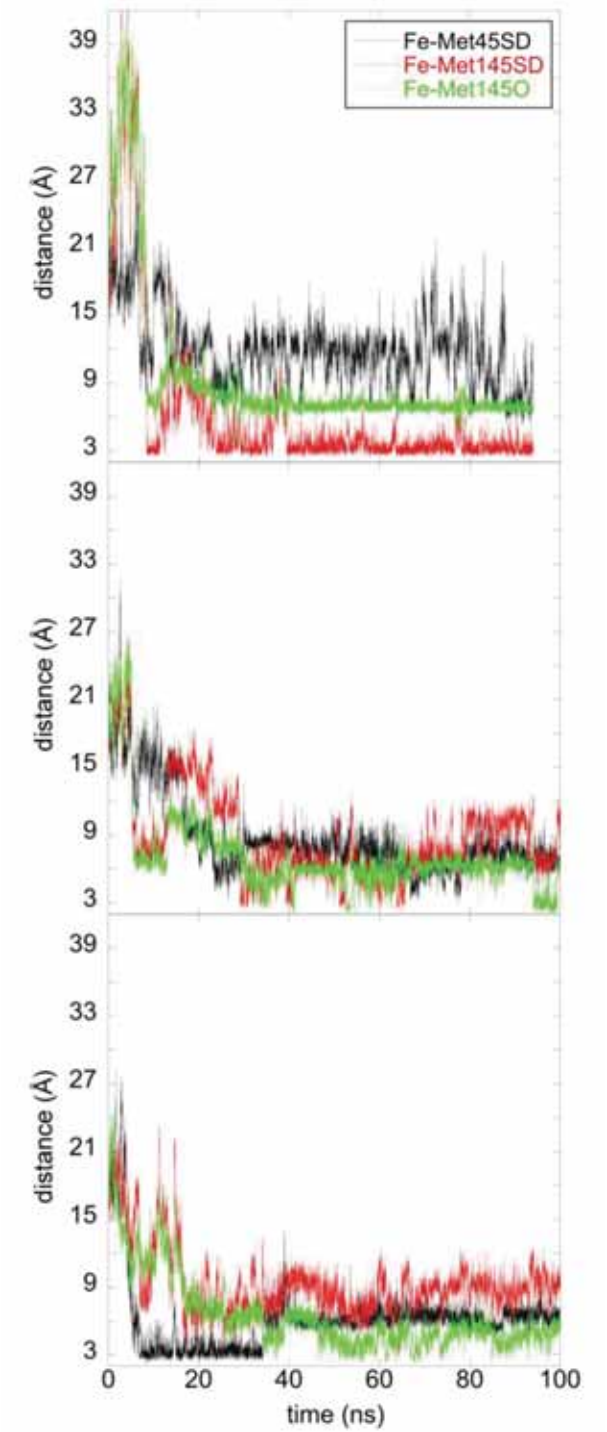


Figure S10. Time evolution of the loops containing Met45, Met46 and Met145 residues using the construction shown in Figure 8F. In this initial configuration, where heme is inserted at Met119 to a structure arising from a 24 ns snapshot of apo-PinO MD, the loops are located ~20 Å from the Fe-heme. During MD simulations the loop containing Met119-heme 'flips' back into the main body of the protein and makes closer contact with the Met-containing loops. Three independent MD replicates are shown.

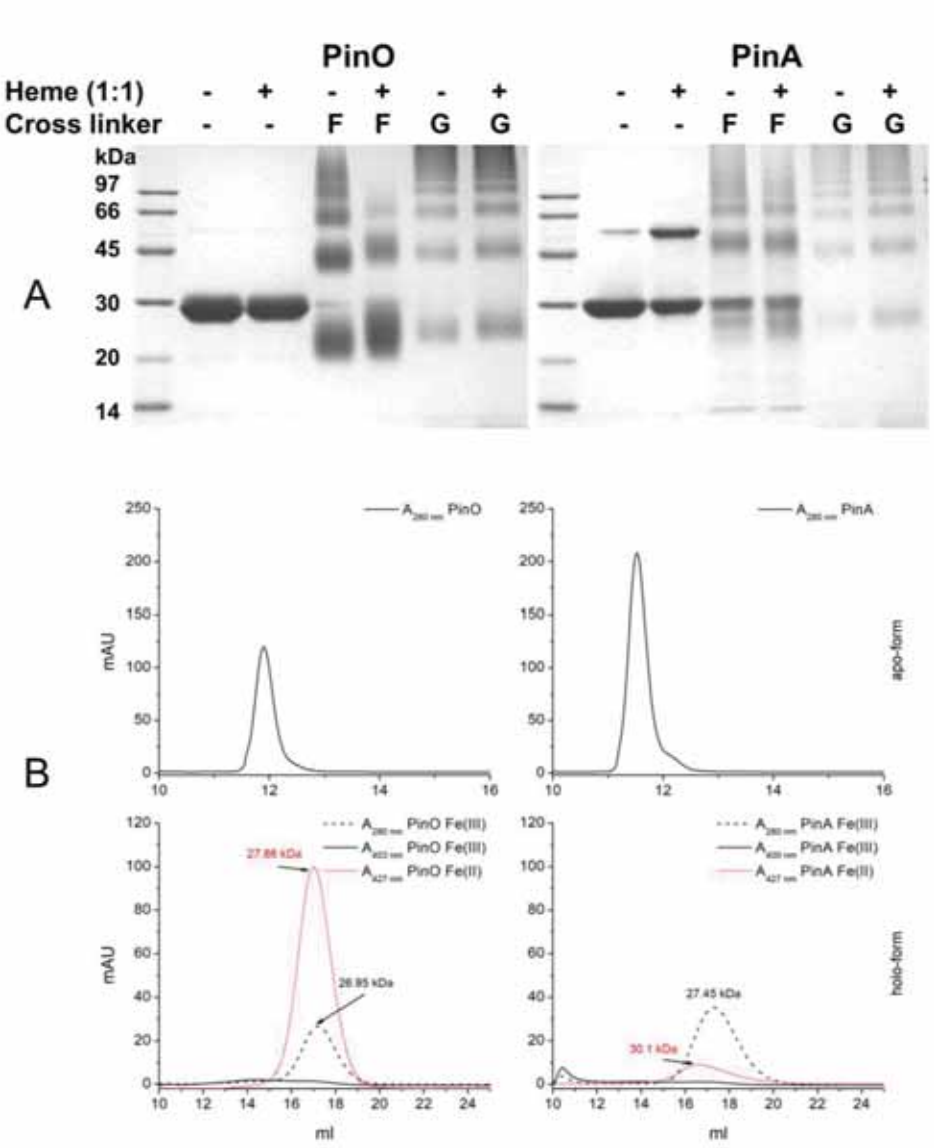


Figure S11. Analysis of oligomer formation. (**A**) Cross-linking analysis and (**B**) size exclusion chromatography of apo- and holo-forms of *P. intermedia* PinO and PinA. Cross linking of proteins in apo- and holo-forms was carried out using 1% formaldehyde (F) or 0.1% glutaraldehyde (G) for 1 h at 37°C. Chromatography was carried out under air (oxidizing conditions; black solid and dashed lines) and reducing conditions (addition of sodium dithionite; red line).