A heme-binding protein produced by *Haemophilus haemolyticus* inhibits non-typeable *Haemophilus influenzae*

Roger D. Latham [1], Mario Torrado [2], Brianna Atto [3], James L. Walshe [2], Richard Wilson [4], J. Mitchell Guss [2], Joel P. Mackay [2], Stephen Tristram [3], David A. Gell [1]

[1] School of Medicine, University of Tasmania, Hobart

[2] School of Life and Environmental Sciences, University of Sydney, Camperdown

[3] School of Health Sciences, University of Tasmania, Launceston

[4] Central Science Laboratory, University of Tasmania, Australia

Correspondence and requests for materials should be addressed to ST (Stephen.Tristram@utas.edu.au) or DAG (david.gell@utas.edu.au)

Keywords: Haemophilus haemolyticus, non-typeable *Haemophilus influenzae* (NTHi), hemophore, hemophilin

Abstract

Haemophilus haemolyticus is a commensal bacterium of humans and occupies the same respiratory tract niche as the pathogenic bacterium, non-typeable *Haemophilus influenzae* (NTHi). Previously we showed that an isolate of *H. haemolyticus* secretes a protein that inhibits the growth of NTHi in culture. Here, we purify the *H. haemolyticus* NTHi-inhibitory protein, identify the gene using mass spectrometry and proteomics, and show that the recombinant protein produced in *E. coli* carrys a bound heme molecule, leading us to name it hemophilin. Recombinant hemophilin has NTHi-inhibitory activity and a gene knockout in *H. haemolyticus* confirms that hemophilin is the source of NTHi-inhibitory activity in the native organism. An x-ray crystal structure shows that hemophilin has a topology similar to bacterial proteins that bind human transferrin, hemoglobin, complement factor H and heparin, but none of these proteins bind heme, and no heme-binding domain with similarity to hemophilin could be detected by structure- or sequence-based searches. The heme occupies a pocket that is protected from solvent, with the heme iron atom coordinated by the side chain of His119. Hemophilin knockout bacteria show a limited capacity to utilise ferric heme to support growth, compared to the parent strain. The data suggest that hemophilin is a hemophore and that inhibition of NTHi growth occurs by heme starvation, raising the possibility that competition for heme between some commensal and pathogenic bacteria might influence bacterial colonisation and therefore disease outcome.

Significance statement

Non-typeable *Haemophilus influenzae* (NTHi) are human pathogens that reside in the upper respiratory tract and require heme for growth. There is currently no effective vaccination against this pathogen, and antibiotic control is becoming less effective. *Haemophilus haemolyticus* is a commensal of the upper respiratory tract, and is also a heme auxotroph. Our study suggests that production of hemophilin might allow some strains of *H. haemolyticus* to compete more effectively with NTHi. Lower colonisation density may result in less frequent opportunistic NTHi infections and provide an alternative means for preventing infection.

Introduction

Non-typeable *Haemophilus influenzae* (NTHi) are Gram-negative bacteria with their reservoir in the upper respiratory tract of humans. Although frequent colonizers of healthy children and adults (1), NTHi are also an important cause of non-invasive disease, and are the leading cause of recurrent otitis media in children (2) and exacerbations of chronic obstructive pulmonary disease in the elderly (3, 4). NTHi have now replaced *H. influenzae* capsular type b (Hib) as the leading cause of invasive infections for this species, and are also an increasingly important cause of conjunctivitis, sinusitis, exacerbations of cystic fibrosis, persistent bacterial bronchitis, and pneumonia (5).

The management of NTHi infections is becoming increasingly difficult. The intrinsic heterogeneity of NTHi has hampered vaccine development, and despite significant effort, an effective vaccine is not currently available (6). There has been an increase in both the incidence and spectrum of antimicrobial resistance in NTHi (7) and the development of new antibiotics for NTHi has recently been listed as a priority by the World Health Organisation, emphasising the need for new approaches to the prevention and management.

As an alternative to antibiotics, bacteriocin-producing strains of upper respiratory tract commensal streptococci have been commercialized as probiotics to prevent and treat *S. pyogenes* infections (8), and, more recently, bacteriocin-producing strains of *Staphylococcus lugdunensis* were shown to be associated with reduced nasal carriage of *S. aureus* (9). In light of these studies, we considered that certain commensal *Haemophilus* spp. might have potential as probiotics to counter NTHi infection. *Haemophilus haemolyticus* are bacteria that are resident in the upper respiratory tract of healthy adults and children at sites that are also colonised by NTHi. *H. haemolyticus* is genetically and phenotypically closely related to NTHi, with both requiring heme and NAD for growth. However, unlike NTHi, *H. haemolyticus* is not considered to be an opportunistic pathogen of the respiratory tract (10, 11).

Recently, with the purpose of developing strains of *H. haemolyticus* as respiratory tract probiotics, we performed a screen of 100 *H. haemolyticus* isolates and found that the isolate BW1 produced a compound capable of inhibiting the growth of NTHi *in vitro* (12). The activity co-purified with a protein of ~30 kDa based on SDS-PAGE and size-exclusion chromatography (SEC), but the identity of the protein and its mechanism of action were not known. Here, we show that the NTHi inhibitory factor is a heme binding protein, which we have named hemophilin. An x-ray crystal

structure shows that hemophilin has a previously undescribed heme-binding fold. Recombinant hemophilin, and a hemophilin gene knockout in *H. haemolyticus*, demonstrate that hemophilin is sufficient and necessary for NTHi inhibition *in vitro*. Together, the results presented below suggest that hemophilin has an important role in heme uptake by *H. haemolyticus*, and that growth-inhibition of competing NTHi occurs by heme starvation.

Results

Identification of the hemophilin gene from two *Haemophilus haemolyticus* clinical isolates

We previously showed that an NTHi-inhibitory protein was produced by *H. haemolyticus* isolate BW1 (12). Further screening identified a second *H. haemolyticus* isolate, RHH122, with similar activity (Supplementary Figs. S1 and S2). To identify the NTHi-inhibitory protein, which we now call hemophilin, conditioned growth medium from stationary phase cultures of BW1 and RHH122 was fractionated by ammonium sulfate precipitation, size exclusion chromatography (SEC) and reversed-phase HPLC (RP-HPLC). Identical (mock) separation steps were performed for two control *H. haemolyticus* isolates (BW39 and BWOCT3) that lacked NTHi-inhibitory activity. For BW1 and RHH122 samples, RP-HPLC peaks containing NTHi-inhibitory activity eluted at the same retention time, and these peaks were absent from BW29 and BWOCT3 samples (Fig. 1A). SDS-PAGE and silver staining revealed one predominant polypeptide of ~30 kDa in the active fractions from BW1 and RHH122 (Fig. 1B and Fig. S2), which was identified as a hypothetical protein (GenBank accession EGT80255) from *H. haemolyticus* strain M19107 by mass spectrometry and peptide mass fingerprint (PMF) analysis (Fig. 1C). PMF analysis of whole RP-HPLC fractions identified EGT80255 peptides as the most abundant ions in samples with NTHi-inhibitory activity obtained from BW1 and RHH122, whereas MS analysis of matched RP-HPLC fractions from control strains, BW29 and BWOCT3, did not identify EGT80255 peptides above background (normalised intensity < 1%; Supplementary Table S1). Notably, no peptides corresponding to the first 22 amino acids of EGT80255 were identified, suggesting that a predicted signal peptide (Fig. 1C, blue) had been cleaved to release the mature hemophilin protein into the *H. haemolyticus* growth medium.

PCR-based sequencing of the hemophilin gene loci from BW1 and RHH122 genomic DNA confirmed the presence of ORFs that were identical to EGT80255. We cloned the hemophilin ORF

from BW1 and expressed the mature form (residues 23–272), without the N-terminal signal peptide, in *E. coli*. Purified recombinant hemophilin displayed NTHi-inhibitory activity at micromolar concentrations (Fig. 1D). A truncated recombinant protein (residues 55–272) was expressed and purified from *E. coli* under identical conditions (Fig. S3) but displayed no NTHi-inhibitory activity (Fig. 1E). To show that hemophilin is the NTHi-inhibitory activity in *H. haemolyticus*, we generated a hemophilin gene knockout mutant of *H. haemolyticus* BW1 by insertional inactivation with a kanamycin resistance cassette. Media recovered from cultures of the knockout strain displayed no inhibitory activity against NTHi (Fig. S4). Growth medium recovered from BW1 or the BW1 hemophilin gene knockout contained similar protein content overall, as judged by chromatographic separations and silver staining, with the notable exception that the 27-kDa hemophilin protein band was absent from the knockout samples (Fig. S4). In summary, these results show hemophilin is identical to ORF sequence EGT80255 and is the NTHi-inhibitory protein from *H. haemolyticus* isolates BW1 and RHH122.

Hemophilin is a heme binding protein

The cell pellets from *E. coli* cultures expressing hemophilin had a distinct pink-red appearance that co-purified with the hemophilin polypeptide. The UV-visible absorption spectrum of the hemophilin had strong absorption bands at 415 nm, 540 nm and 575 nm, which are characteristic of the Soret and α/β bands of a hemoprotein (Fig. 2A). The presence of non-covalently bound heme *b* was confirmed by RP-HPLC and mass spectrometry (Fig. 2B). Pools of heme-bound hemophilin holo protein (Fig. 2A, red trace) and largely heme-free apo protein (Fig. 2A, black trace) could be obtained from *E. coli* lysates by ion exchange chromatography. Apo protein could also be obtained by RP-HPLC or acid acetone extraction of the holo protein fraction. The apo protein was able to re-bind heme that was added back to the solution, either as reduced ferrous heme or oxidised ferric heme, and the reconstituted holo proteins were then competent to bind to a variety of small ligands that are specific for ferrous (O_2, CO; Fig. S5) or ferric (CN^-, HS^-; Fig. S6) hemes. The UV-visible spectrum of apo hemophilin reconstituted with ferrous heme and O_2 (Fig. 2A, green trace) is essentially identical to that of hemophilin purified from *E. coli* as the holo protein, suggesting that the *E. coli* derived protein is the O_2-bound form.

The holo and apo hemophilin fractions gave very similar far-UV CD spectra, with maxima at ~200 nm and minima between 220–230 nm, indicating a mixture of β-sheet and α-helical secondary structure elements, and no major change in secondary structure upon heme binding (Fig. 2C). This

is unlike hemoglobins or heme-binding enzymes, which typically undergo at least partial denaturation in the absence of the heme cofactor, and is more similar to the apo state characteristics of transient heme binding proteins, such as heme transport proteins, which essentially have folded apo protein structures (13). In addition to retaining structure, the hemophilin apo protein retained NTHi-inhibitory activity. Indeed, the specific activity of the apo protein was typically 3–4 fold higher than the holo protein (Fig. 2D). Together these data suggest that reversible binding to heme could contribute to the biological function of hemophilin.

Hemophilin has a novel heme binding structure

Holo hemophilin was crystallized and the structure determined by x-ray diffraction to a resolution of 1.6 Å (Figure 3; Table S2). Initial phases were obtained by single wavelength anomalous diffraction (SAD) from the heme iron, above the absorption K-edge. All residues of hemophilin (23–272) were visible in the electron density. After refinement against a model comprising only the hemophilin polypeptide, the shape and position of a bound porphyrin and a small non-protein ligand to the heme iron were clearly visible in electron difference density maps (Fig. S7). Electron density and an anomalous diffraction peak at the ligand position indicated a single heavy atom, which was modelled as a chloride ion (Fig. 4A, B). Although the hemophilin protein used for crystallization was predominantly ferrous O_2 complex there was contaminating ferric protein (Fig. S8), and the crystallization conditions (pH 4.5) favoured further heme oxidation (Fig. S9) and subsequent chloride binding (Fig. S10 shows a K_d of ~3 mM for chloride binding; chloride concentration in the crystallization drop was estimated at 35–70 mM), compared to neutral buffer conditions. The chloride refined at full occupancy and *B*-factors were consistent with those of the iron and other surrounding atoms. The Fe–Cl bond length in the hemophilin structure is 2.47(8) Å, which is in the range observed for Fe–Cl distances in heme protein structures with chloride ligands (14-16) (Fig. S11; Table S3).

Hemophilin comprises an N-terminal region with mixed α/β secondary structure (Fig. 3A, B, blue-green ribbon) that binds a single heme molecule (Fig. 3A, sticks and sphere), and a C-terminal 8-stranded β-barrel (Fig. 3A, B, yellow-red ribbon). The N-terminal α/β region is a sandwich of two antiparallel β-sheets, with insertions between the β-3 and β-4 strands, and between the β-5 and β-6 strands, that cover both faces of the porphyrin ring and provide the majority of protein-heme contacts. The DALI server shows that several bacterial proteins have structural similarity to hemophilin (Fig. S12). Notably, all are cell-surface proteins with ligand binding functions (Fig.

S13), but none have heme binding function or structural similarity within the hemophilin heme-binding site. No heme-binding domain with similarity to hemophilin could be detected by structure- or sequence-based searches.

The heme iron in hemophilin is coordinated in one axial position by the Nε2 atom of His119. The imidazole side chain of His119 makes a hydrogen bond through Nδ1 to the backbone carbonyl of Pro117 and is surrounded by a hydrophobic cage comprising Ala69, Met106, Met109, Met116, Pro117, Leu122 side chains (Fig. 4A) that shields one face of the porphyrin and the Fe–Nε bond from water, providing a similar physicochemical environment to that of the proximal His in metazoan Hbs. His119 and all its caging residues, bar Ala69, lie on the protein loop between β strands 5 and 6 that extends across one face of the heme group (Fig. 4C). The heme Fe atom sits in the plane of the porphyrin—0.08(8) Å from the least squares plane defined by the pyrrole N atoms—consistent with a 6-coordinate octahedral geometry. On the distal side of the heme, the side chains of Arg82 and Gln74 are within hydrogen bonding distance of the heme distal ligand and the porphyrin 17-propionate (Fig. 4A and Fig. S11). The Arg82 side chain extends across the face of the porphyrin with the plane of the guanidinium group contributing cation-π and π-π interactions (17). The heme is oriented with the iron atom and porphyrin ring almost completely buried (solvent accessible non-polar area of 47 $Å^2$, compared to 688 $Å^2$ for free heme) and the ionisable propionate groups pointing out into solvent (Fig. 4D).

To investigate whether the heme binding function of hemophilin is required for NTHi-inhibitory activity, we made a His to Gln substitution of the heme-coordinating His119 (H119Q). Unlike wild-type hemophilin, the H119Q mutant protein purified from *E. coli* without a heme cofactor, indicating a large reduction in heme binding affinity (Fig. S14). Mixing hemin with the H119Q mutant gave a spectrum with peaks at 404, 485 and 602 nm (Fig. S14), similar to spectra of ferric heme:protein complexes without an Fe-coordinating side chain (18), suggesting that the H119Q mutant binds heme, but probably does not form an axial bond to the iron. The H119Q mutant was not inhibitory at up to 8-fold higher concentration than the wild-type hemophilin (the highest concentration tested; Fig. 4E and Fig. S14), indicating that heme binding through H119 contributes substantially to NTHi inhibition.

Hemophilin sequesters a pool of heme and makes it available to *H. haemolyticus* but not to the related species NTHi

H. haemolyticus is unable to synthesise porphyrin and must acquire heme from the environment (19). Therefore, we hypothesised that hemophilin might function as part of a heme uptake pathway in *H. haemolyticus*. To investigate this idea, we compared the growth of BW1 and the BW1 knockout strain under different hemin supplementation regimes. After a 14-hour conditioning period in hemin-deficient media (TSB without hemin supplement), BW1 grew when cultured into fresh medium containing 15 μg mL^{-1} hemin, but not with 0.9 or 0 μg mL^{-1} hemin, indicating a dependence on hemin uptake from the media for growth (Fig. 5A and Figs. S15, S16). After similar conditioning in hemin-deficient medium, the BW1 knockout strain was unable to grow at any of the hemin concentrations tested (0, 0.9 or 15 μg mL^{-1}; Fig. 5B), suggesting that the BW1 knockout cells had a defect in heme uptake. When BW1 was conditioned in heme-replete media (TSB supplemented with 15 μg mL^{-1} hemin), cultures grew without a requirement for additional hemin supplement (Fig. 5C), presumably due to accumulation of heme or porphyrin stores during the conditioning phase (20, 21). After similar conditioning in the hemin-replete medium, the BW1 knockout strain grew poorly at all hemin concentrations (Fig. 5E), when compared to wild type BW1, suggesting that the BW1 knockout strain has reduced capacity to accumulate heme or porphyrin in environments where free hemin is the sole heme source.

To test whether heme-loaded hemophilin can be used directly as a heme source by *H. haemolyticus*, we grew liquid cultures supplemented with hemin or holo hemophilin at matching final molar concentrations. *H. haemolyticus* BW1 or NTHi strain NCTC 4560 were first grown in TSB broth containing hemin at a concentration that was limiting to growth (0.6 μg mL^{-1}). These cultures were then supplemented with additional free hemin or holo hemophilin (each at 7.7 μM). As expected, the addition of free hemin caused an increase in the growth of both organisms (Fig. S17). Addition of holo hemophilin or free hemin to BW1 resulted in equivalent growth (approximately 3-fold increase in culture density at 20 hours compared to baseline; Fig. S17), suggesting that hemophilin-bound heme was readily available to BW1. In contrast, the addition of holo hemophilin to NTHi caused a decrease in growth (approximately 2-fold decrease in cell density at 20 hours; Fig. S17), suggesting that heme bound to hemophilin is not accessible to NTHi. Finally, we found that the inhibition of NTHi in the agar well diffusion assay was overcome by addition of excess hemin (Fig. S18). Together, these results suggest that hemophilin binds heme in a form that can be utilised by *H. haemolyticus* BW1, but is not accessible to NTHi, and that inhibition of NTHi occurs by heme starvation.

Distribution of hemophilin across *H. haemolyticus* strains

Given the prominent role in heme acquisition that is suggested by the BW1 hemophilin gene knockout, we searched the public sequence databases to explore the distribution of hemophilin genes in *H. haemolyticus* and other species. From 46 complete or draft *H. haemolyticus* genomes available in Genbank, 30 genomes contained an ORF with 67–100% nucleotide sequence identity to BW1 hemophilin (Fig. S19; Table S4). In addition, 18 *H. influenzae* genomes (from a total of 700 genome assemblies), 3 *H. quentini* genomes (from a total of 3 assemblies), 1 *H. parainfluenzae* genome (from a total of 40 assemblies) encoded an ORF with similarity to BW1 hemophilin. In each case, the hemophilin gene was immediately downstream of a predicted TonB dependent transporter (Table S4). Across the 25 unique hemophilin-like sequences encoded by all genomes, amino acid differences overwhelmingly mapped to surface exposed sites and surface loops on the hemophilin structure (Fig. S20). Residues surrounding the heme ligand, including the heme coordinating His119 and the distal pocket residues Gln74, Met79 and Arg82 (Fig. 4 and Fig. S20) were 100% conserved in all hemophilin variants, indicating that heme binding is preserved in all variants. In summary, hemophilin homologues are present in a large proportion (63%) of *H. haemolyticus* genomes, compared to a small minority (<3%) of *H. influenzae* genomes, making hemophilin a distinctive feature of *H. haemolyticus* (and also *H. quentini*). Biochemical and genetic studies are now needed to establish if/how hemophilin alleles from the different phylogentic groups contribute to heme uptake in these strains.

The observation that over half of published *H. haemolyticus* genomes carry a hemophilin-like gene raises the question of why only two NTHi-inhibitory strains of *H. haemolyticus* (BW1 and RHH122) were identified in our strain collection. To investigate this question, we screened genomic DNA for the hemophilin ORF by real-time PCR; from our collection of 100 *H. haemolyticus* clinical isolates, 15 were positive for hemophilin genes (this primer set would theoretically have produced an amplicon for 7 of the 46 public *H. haemolyticus* genomes, the same 15% hit rate). After 50-fold concentration of culture supernatants, various levels of NTHi-inhibitory activity could be detected for 9 of these 15 isolates, with the greatest activity in BW1 and RHH122 (Table S5). Notably, the hemophilin ORF sequences from 6 isolates, including BW1 and RHH122, were 100% identical and yet the NTHi-inhibitory activity recovered from these isolates varied considerably. This suggests that high-level production of this hemophilin allele, at least under laboratory conditions, is an unsual feature of BW1/RHH122 and emphasises the importance of understanding how hemophilin production is regulated, in addition to the effects of sequence variation within the hemophilin protein.

Discussion

Hemophilin is a previously undescribed hemophore in H. haemolyticus

In a screen for potential probiotic strains of *H. haemolyticus* that produce inhibitors of NTHi, we identified hemophilin, a small soluble heme binding protein that is secreted at high levels by two *H. haemolyticus* isolates, BW1 and RHH122. Our data suggest that hemophilin plays a positive role in heme acquisition in *H. haemolyticus*, and that the inhibition of NTHi, a pathogenic microorganism that shares a similar ecological niche in the upper respiratory tract of humans, is likely to involve competition for nutrient heme.

H. influenzae and *H. haemolyticus* lack the enzymes for *de novo* porphyrin synthesis and depend for their survival on scavenging heme, or protoporphyrin IX plus iron, from the host (19). *H. influenzae* have multiple pathways to scavenge heme (Table S4) (22); in each pathway the initial heme binding step is performed by a membrane anchored protein or TonB-dependent outer membrane heme transporter. By comparison, hemophilin represents a previously unidentified mechanism in *Haemophilus* spp—that is, secretion of a diffusible heme binding protein (a hemophore) into the surrounding environment. Hemophores are found in a subset of bacterial species. In Gram-positive species, hemophores can be recognised by the prescence of one or more heme-binding NEAT (near iron transporter) domains; these proteins are covalently attached to the cell surface peptidoglycan, or S-layer, or secreted into the environment (23). In Gram-negative organisms, three hemophores have been described: HasA from *Serratia marcesens*, *Pseudomonas* spp., and *Yersinia* spp. (24, 25); HmuY from *Porphyromonas gingivalis* (26); and HusA, also from *P. gingivalis* (18). Hemophilin, HasA, HmuY, HusA and NEAT domains have different folds, suggesting that hemophore functions have been independently acquired on multiple occasions in evolution.

Structural features are consistent with hemophilin acting as a heme scavenging and transport protein: apo and holo hemophilin have very similar secondary structure content and the apo protein was stable in solution and biologically active. In contrast, the heme group typically makes important contributions to the struture and stability of proteins that use heme as a cofactor, such as hemoglobins and various enzymes (13). Nevertheless the heme is deeply buried in hemophilin, requiring at least transient structural changes to allow heme entry and exit; a possible mechanism is a conformational change in β5–6 loop, which contains the iron-coordinating His119 and non-polar side chains that contact all four pyrrole rings on one face of the heme (Fig. 4C). Future solution state studies will be required to address the heme binding/release mechanism.

Heme coordination by a single His side chain, as seen in hemophilin, is unusual for extracellular heme transport proteins, although it is common for heme proteins in general. Known hemophores make 5-coordinate heme complexes through Tyr (15, 27, 28) or 6-coordinate complexes via His/His (26), Tyr/His (29), Tyr/Met (30), or Met/Met (31) ligands. Ligand combinations that include Tyr strongly favour binding to ferric over ferrous heme (32), which is appropriate in the typically oxidising environment of the extracellular milieu. Heme ligation through Met or His ligands is compatible with ferrous or ferric heme binding and would potentially allow hemophilin to capture heme in anaerobic as well as aerobic conditions.

The heme-site structure suggests hemophilin could bind small molecule ligands in vivo

Proteins that coordinate heme through a single axial His side chain commonly function as enzymes or small-molecule sensing/transport proteins. In heme-containing oxidoreductase enzymes the heme-coordinating His side chain is hydrogen bonded to the carboxylate side chain of Asp or Glu, which confers some imidazolate character and negative charge donation to substrates bound in the sixth iron coordination site (33); this contrasts with the neutral imidazole side chain that is inferred from the chemical environment of His119 in hemophilin, which is more similar to the heme coordination sites in hemoglobins. Like hemoglobins, hemophilin forms stable complexes with ferrous and ferric heme ligands (Figs. S5, S6, S8) and purifies from *E. coli* with an O_2 ligand bound, suggesting a potentially biologically meaningful O_2 affinity and resistance to autooxidation; exclusion of water from the proximal face of the heme and the small size and enclosed nature of the distal pocket are consistent with this (34). Other hemophores are typically obtained in the ferric form, and without bound ligands, although CO-adducts of the HasA hemophore can be prepared by reduction under pure CO atmosphere (35, 36).

The chloride ion present as a heme ligand in the hemophilin crystal structure is not likely to be of physiological importance but indicates a strong propensity to bind to anionic ligands in the ferric state. A chloride ligand has been observed in crystal structures of only four heme proteins, including hemophilin (14-16); these proteins have unrelated folds, but all have an Arg side chain with guanadinium group approximately parallel to the porphyrin plane that might facilitate binding of small anionic ligands (Fig. S11). Bacteria from a variety of environments use heme proteins to sense ligands such as NO and HS^- (37), which also have powerful signalling roles in the mammalian immune system (38), making small ligand binding by hemophilin potentially important, although this remains to be investigated.

Proteins with structural similarity to hemophilin bind diverse ligands

The small group of proteins that share structural similarity with hemophilin (Fig. S12) includes hemoglobin-haptoglobin utilisation protein (HpuA) (39), complement factor H binding protein (fHbp) (40) and *Neisseria* heparin binding antigen (NHBA) (39), found in members of the *Neisseriaceae*, and transferrin binding protein B (TbpB) (41), which is found in *Neisseriaceae* and some members of the *Pasteurellaceae*, including *H. influenzae* and *H. haemolyticus* (identifiable by BLAST search). HpuA, fHbp, HNBA and TbpB are lipid anchored in the bacterial outer membrane and have functions in iron uptake, immune evasion or surface attachment. Remarkably, the precise combination of β-barrel topology (8 strands in a meander topology with a shear value of 8) together with hydrophobic residues packed in the barrel core seems to occur only in this group of bacterial proteins (Fig. S12, Legend). Topological similarities between hemophilin, HpuA, TbpA, fHbp and NHBA outside the β-barrel domain suggest that a common ancestor of these proteins also included sequences additional to the β-barrel (Fig. S12, Legend).

The hemophores HasA, HmuY and HusA, as well as HpuA and TbpB, deliver cargo to their respective TonB-dependent outer membrane heme transporters (HasR, HmuR, HusR, HpuB and TbpA) (42, 43) that are expressed from the same operon. It was therefore expected that a gene encoding a transporter should be found at a locus close to hemophilin and, indeed, a predicted TonB-dependent transporter with previously uncharacterised function occurs immediately upstream of the hemophilin gene in all 52 published genomes containing a hemophilin gene sequence (Table S4). PSI-BLAST searches identify HasR (heme transporter for the HasA hemophore) as the closest sequence match to the putative hemophilin receptor, and molecular modelling using the PHYRE and I-TASSER web servers confidently predicts a 22-strand β-barrel and plug structure with closest similarity to TbpA and HasR (Fig. S21).

Hemophilin* genes are prevalent amongst strains of *H. haemolyticus*, but not *H. influenzae

A question arises as to why hemophilin-like genes are common in *H. haemolyticus*, yet extremely rare in *H. influenzae*, given the similar requirement for heme and the close phylogenetic relationship between these species. Part of the answer might lie in the different combinations of heme uptake genes in *H. haemolyticus* and *H. influenzae* genomes. *H. influenzae*, and particularly NTHi, have enormous genetic diversity due to their intrinsic transformability and the high rate of recombination with exogenous DNA from their environment (44), such that less than 50% of ORFs

are present in all strains (the core genome) while the remainder are present variably and represent an accessory genome (45). In a group of 88 *H. influenzae* genome assemblies (NTHi and Hib) (46), 100% of strains have genes for the outer membrane haem transporters, *hup* and *hemR*, as well as the *HxuCBA* genes for heme uptake from hemopexin (Table S4). In addition, 98% of strains have at least 1 gene encoding a transporter for heme uptake from hemoglobin/haptoglobin (*hgpA, hgpB, hgpC*; Table S4). Thus, the majority of these *H. influenzae* strains have highly redundant pathways for accessing heme from a variety of host sources. The number of heme acquisition genes found in *H. haemolyticus* strains is much more variable. From the 46 available *H. haemolyticus* assemblies, 70% have *hup*, only 7% have *hemR*, and 87% have at least 1 gene for *hgpA*/*hgpB*/*hgpC*. Notably, no *H. haemolyticus* strains carry the *huxCBA* system, which is considered a virulence factor in *H. influenzae* (47). Similarly, diagnostic PCR screens performed on large collections of isolates have also shown much lower prevalence of *hup*, *hemR*, and *hxuCBA* genes in *H. haemolyticus* compared to NTHi (22). On balance then, *H. haemolyticus* have fewer heme uptake pathways than *H. influenzae*, which may have created stronger selective pressure for *H. haemolyticus* to acquire a hemophore. In this context, finding hemophilin-like genes in 65% of *H. haemolyticus* genome assemblies suggests hemophilin may be of considerable importance in the overall heme economy of this species.

By scavenging heme, hemophilin might contribute to interference competition between Haemophilus species

We undertook our initial search for strains of *H. haemolyticus* that could inhibit NTHi with the goal of developing respiratory tract probiotics. The paradigm on which we based this search was that some bacterial strains produce bacteriocins to specifically inhibit related species that share the same ecological niche. The fact that we identified hemophilin, a putative hemophore, suggests that sequestering heme might be an alternative mechanism for bacteria to inhibit their competitors, particularly in cases where competition occurs between heme auxotrophs such as *Haemophilus* spp. Competition for iron and heme between bacterial pathogens and their hosts is well accepted, and the ability of the host to impose low concentrations of free iron is one of the most important forms of nutritional immunity (48). Iron availability is also an important determinant of competition dynamics between microbial species (49). Competition for iron between microbes in the context of infection has, however, received less attention than the concomitant pathogen-host interaction. Now, the emerging picture is that non-pathogenic probiotic bacteria, as well as pathogenic species, have enhanced iron and heme uptake capabilities that facilitate inhibition of microbial neighbours as well as colonisation of the host (50).

Our work identifies hemophilin as a previously unrecognised heme uptake mechanism in *H. haemolyticus* with the potential to block uptake of the essential heme micronutrient by pathogenic NTHi. In general, probiotics require a combination of attributes, including enhanced iron transport, immune modulation, competition for mucosal binding sites, and bacteriocin production. Since *H. haemolyticus* co-colonises the upper respiratory tract with NTHi and competes for binding sites on epithelial cells (51), *H. haemolyticus* strains with high-level expression of particular hemophilin alleles, as seen in the BW1 and RHH122, might starve NTHi of heme and locally and specifically inhibit NTHi colonisation. Co-culture studies with hemophilin producing *H. haemolyticus* and NTHi in cell culture and animal models are warranted to further explore this potential.

Materials and Methods

The origin, identification and culture of the bacterial strains, and agar well diffusion assay have been described (12). For PMF analysis, tryptic peptide digests were analysed by nanoLC-MS/MS and MS/MS spectra were searched against *H. haemolyticus* sequences downloaded from NCBI on Sept 19th, 2016. Data are available from the ProteomeXchange PRIDE repository (accession PXD013687). Hemophilin was expressed from pET28a in *E. coli* and purified by Ni-affinity. The His-tag was removed with thrombin and untagged hemophilin was obtained by cation exchange. A hemophilin gene knockout was constructed by insertion of a kanamycin resistance cassette into a unique *Bsp*TI site in the hemophilin gene of *H. haemolyticus* strain BW1. Hemophilin crystals were grown in hanging drops (mother liquor was 0.1 M sodium acetate trihydrate pH 4.5, 2 M ammonium sulfate) and cryoprotected in 25% glycerol. Diffraction data were collected at the MX2 beamline at the Australian Synchrotron; phases were obtained by single-wavelength anomalous dispersion; PDB accession is 6om5. Full methods are provided as supplementary information.

Acknowledgements

This work was funded in part by a grant from the Clifford Craig Foundation, Launceston, Tasmania (Grant # 170). This research was undertaken in part using the MX2 beamline at the Australian Synchrotron, part of ANSTO, and made use of the Australian Cancer Research Foundation (ACRF) detector.

References

1. Lemon KP *et al.* (2010) Comparative analyses of the bacterial microbiota of the human nostril and oropharynx. *MBio* **1**.
2. Ngo CC, Massa HM, Thornton RB, Cripps AW (2016) Predominant Bacteria Detected from the Middle Ear Fluid of Children Experiencing Otitis Media: A Systematic Review. *PLoS One* **11**: e0150949.
3. Rosell A *et al.* (2005) Microbiologic determinants of exacerbation in chronic obstructive pulmonary disease. *Arch Intern Med* **165**: 891-897.
4. Sethi S, Murphy TF (2008) Infection in the pathogenesis and course of chronic obstructive pulmonary disease. *N Engl J Med* **359**: 2355-2365.
5. Van Eldere J, Slack MP, Ladhani S, Cripps AW (2014) Non-typeable *Haemophilus influenzae*, an under-recognised pathogen. *Lancet Infect Dis* **14**: 1281-1292.
6. Cerquetti M, Giufre M (2016) Why we need a vaccine for non-typeable *Haemophilus influenzae*. *Hum Vaccin Immunother* **12**: 2357-2361.
7. Tristram S, Jacobs MR, Appelbaum PC (2007) Antimicrobial resistance in *Haemophilus influenzae*. *Clin Microbiol Rev* **20**: 368-389.
8. Di Pierro F *et al.* (2016) Effect of administration of *Streptococcus salivarius* K12 on the occurrence of streptococcal pharyngo-tonsillitis, scarlet fever and acute otitis media in 3 years old children. *Eur Rev Med Pharmacol Sci* **20**: 4601-4606.
9. Zipperer A *et al.* (2016) Human commensals producing a novel antibiotic impair pathogen colonization. *Nature* **535**: 511-516.
10. Murphy TF *et al.* (2007) *Haemophilus haemolyticus*: a human respiratory tract commensal to be distinguished from *Haemophilus influenzae*. *J Infect Dis* **195**: 81-89.
11. Zhang B, Kunde D, Tristram S (2014) *Haemophilus haemolyticus* is infrequently misidentified as *Haemophilus influenzae* in diagnostic specimens in Australia. *Diagn Microbiol Infect Dis* **80**: 272-273.
12. Latham RD *et al.* (2017) An isolate of *Haemophilus haemolyticus* produces a bacteriocin-like substance that inhibits the growth of nontypeable *Haemophilus influenzae*. *Int J Antimicrob Agents* **49**: 503-506.
13. Smith LJ, Kahraman A, Thornton JM (2010) Heme proteins—diversity in structural characteristics, function, and folding. *Proteins* **78**: 2349-2368.
14. Kuwada T *et al.* (2011) Involvement of the distal Arg residue in Cl^- binding of midge larval haemoglobin. *Acta Crystallogr D Biol Crystallogr* **67**: 488-495.
15. Kumar R *et al.* (2013) The hemophore HasA from *Yersinia pestis* (HasAyp) coordinates hemin with a single residue, Tyr75, and with minimal conformational change. *Biochemistry* **52**: 2705-2707.
16. Singh R, Grigg JC, Armstrong Z, Murphy ME, Eltis LD (2012) Distal heme pocket residues of B-type dye-decolorizing peroxidase: arginine but not aspartate is essential for peroxidase activity. *J Biol Chem* **287**: 10623-10630.
17. Kumar K *et al.* (2018) Cation-pi interactions in protein-ligand binding: theory and data-mining reveal different roles for lysine and arginine. *Chem Sci* **9**: 2655-2665.
18. Gao JL *et al.* (2018) Structural properties of a haemophore facilitate targeted elimination of the pathogen *Porphyromonas gingivalis*. *Nat Commun* **9**: 4097.
19. Norskov-Lauritsen N (2014) Classification, identification, and clinical significance of *Haemophilus* and *Aggregatibacter* species with host specificity for humans. *Clin Microbiol Rev* **27**: 214-240.
20. Vogel AR *et al.* (2012) SapF-mediated heme-iron utilization enhances persistence and coordinates biofilm architecture of *Haemophilus*. *Front Cell Infect Microbiol* **2**: 42.

21. Mason KM, Raffel FK, Ray WC, Bakaletz LO (2011) Heme utilization by nontypeable *Haemophilus influenzae* is essential and dependent on Sap transporter function. *J Bacteriol* **193**: 2527-2535.
22. Hariadi NI *et al.* (2015) Comparative Profile of Heme Acquisition Genes in Disease-Causing and Colonizing Nontypeable *Haemophilus influenzae* and *Haemophilus haemolyticus*. *J Clin Microbiol* **53**: 2132-2137.
23. Choby JE, Skaar EP (2016) Heme Synthesis and Acquisition in Bacterial Pathogens. *J Mol Biol* **428**: 3408-3428.
24. Ghigo JM, Letoffe S, Wandersman C (1997) A new type of hemophore-dependent heme acquisition system of *Serratia marcescens* reconstituted in Escherichia coli. *J Bacteriol* **179**: 3572-3579.
25. Izadi N *et al.* (1997) Purification and characterization of an extracellular heme-binding protein, HasA, involved in heme iron acquisition. *Biochemistry* **36**: 7050-7057.
26. Wojtowicz H *et al.* (2009) Unique structure and stability of HmuY, a novel heme-binding protein of *Porphyromonas gingivalis*. *PLoS Pathog* **5**: e1000419.
27. Grigg JC, Vermeiren CL, Heinrichs DE, Murphy ME (2007) Haem recognition by a *Staphylococcus aureus* NEAT domain. *Mol Microbiol* **63**: 139-149.
28. Kanadani M, Sato T, Hino T, Nagano S, Ozaki S (2015) The crystal structure of heme acquisition system A from *Yersinia pseudotuberculosis* (HasAypt): Roles of the axial ligand Tyr75 and two distal arginines in heme binding. *J Inorg Biochem* **151**: 26-33.
29. Arnoux P *et al.* (1999) The crystal structure of HasA, a hemophore secreted by *Serratia marcescens*. *Nat Struct Biol* **6**: 516-520.
30. Gaudin CFM, Grigg JC, Arrieta AL, Murphy MEP (2011) Unique Heme-Iron Coordination by the Hemoglobin Receptor IsdB of *Staphylococcus aureus*. *Biochemistry* **50**: 5443-5452.
31. Ran Y *et al.* (2007) Bis-methionine ligation to heme iron in the streptococcal cell surface protein Shp facilitates rapid hemin transfer to HtsA of the HtsABC transporter. *J Biol Chem* **282**: 31380-31388.
32. Reedy CJ, Elvekrog MM, Gibney BR (2008) Development of a heme protein structure-electrochemical function database. *Nucleic Acids Res* **36**: D307-313.
33. Shepherd M *et al.* (2010) The single-domain globin from the pathogenic bacterium *Campylobacter jejuni*: novel D-helix conformation, proximal hydrogen bonding that influences ligand binding, and peroxidase-like redox properties. *J Biol Chem* **285**: 12747-12754.
34. Carver TE *et al.* (1992) A novel site-directed mutant of myoglobin with an unusually high O_2 affinity and low autooxidation rate. *J Biol Chem* **267**: 14443-14450.
35. Lukat-Rodgers GS, Rodgers KR, Caillet-Saguy C, Izadi-Pruneyre N, Lecroisey A (2008) Novel heme ligand displacement by CO in the soluble hemophore HasA and its proximal ligand mutants: implications for heme uptake and release. *Biochemistry* **47**: 2087-2098.
36. Ozaki SI *et al.* (2014) Spectroscopic studies on HasA from *Yersinia pseudotuberculosis*. *J Inorg Biochem* **138**: 31-38.
37. Martinkova M, Kitanishi K, Shimizu T (2013) Heme-based globin-coupled oxygen sensors: linking oxygen binding to functional regulation of diguanylate cyclase, histidine kinase, and methyl-accepting chemotaxis. *J Biol Chem* **288**: 27702-27711.
38. Toliver-Kinsky T *et al.* (2019) H2S, a Bacterial Defense Mechanism against the Host Immune Response. *Infect Immun* **87**: doi: 10.1128/IAI.00272-00218.
39. Wong CT *et al.* (2015) Structural analysis of haemoglobin binding by HpuA from the *Neisseriaceae* family. *Nat Commun* **6**: 10172.
40. Schneider MC *et al.* (2009) *Neisseria meningitidis* recruits factor H using protein mimicry of host carbohydrates. *Nature* **458**: 890-893.

41. Moraes TF, Yu RH, Strynadka NC, Schryvers AB (2009) Insights into the bacterial transferrin receptor: the structure of transferrin-binding protein B from *Actinobacillus pleuropneumoniae*. *Mol Cell* **35**: 523-533.
42. Noinaj N *et al.* (2012) Structural basis for iron piracy by pathogenic *Neisseria*. *Nature* **483**: 53-58.
43. Krieg S *et al.* (2009) Heme uptake across the outer membrane as revealed by crystal structures of the receptor-hemophore complex. *Proc Natl Acad Sci U S A* **106**: 1045-1050.
44. Mell JC, Shumilina S, Hall IM, Redfield RJ (2011) Transformation of natural genetic variation into *Haemophilus influenzae* genomes. *PLoS Pathog* **7**: e1002151.
45. De Chiara M *et al.* (2014) Genome sequencing of disease and carriage isolates of nontypeable *Haemophilus influenzae* identifies discrete population structure. *Proc Natl Acad Sci U S A* **111**: 5439-5444.
46. Pinto M *et al.* (2018) Insights into the population structure and pan-genome of *Haemophilus influenzae*. *Infect Genet Evol* **67**: 126-135.
47. Morton DJ *et al.* (2007) The haem-haemopexin utilization gene cluster (*hxuCBA*) as a virulence factor of *Haemophilus influenzae*. *Microbiology* **153**: 215-224.
48. Parrow NL, Fleming RE, Minnick MF (2013) Sequestration and scavenging of iron in infection. *Infect Immun* **81**: 3503-3514.
49. Butaite E, Baumgartner M, Wyder S, Kummerli R (2017) Siderophore cheating and cheating resistance shape competition for iron in soil and freshwater *Pseudomonas* communities. *Nat Commun* **8**: 414.
50. Deriu E *et al.* (2013) Probiotic bacteria reduce *salmonella typhimurium* intestinal colonization by competing for iron. *Cell Host Microbe* **14**: 26-37.
51. Pickering JL *et al.* (2016) *Haemophilus haemolyticus* Interaction with Host Cells Is Different to Nontypeable *Haemophilus influenzae* and Prevents NTHi Association with Epithelial Cells. *Front Cell Infect Microbiol* **6**: 50.

Figures and Legends

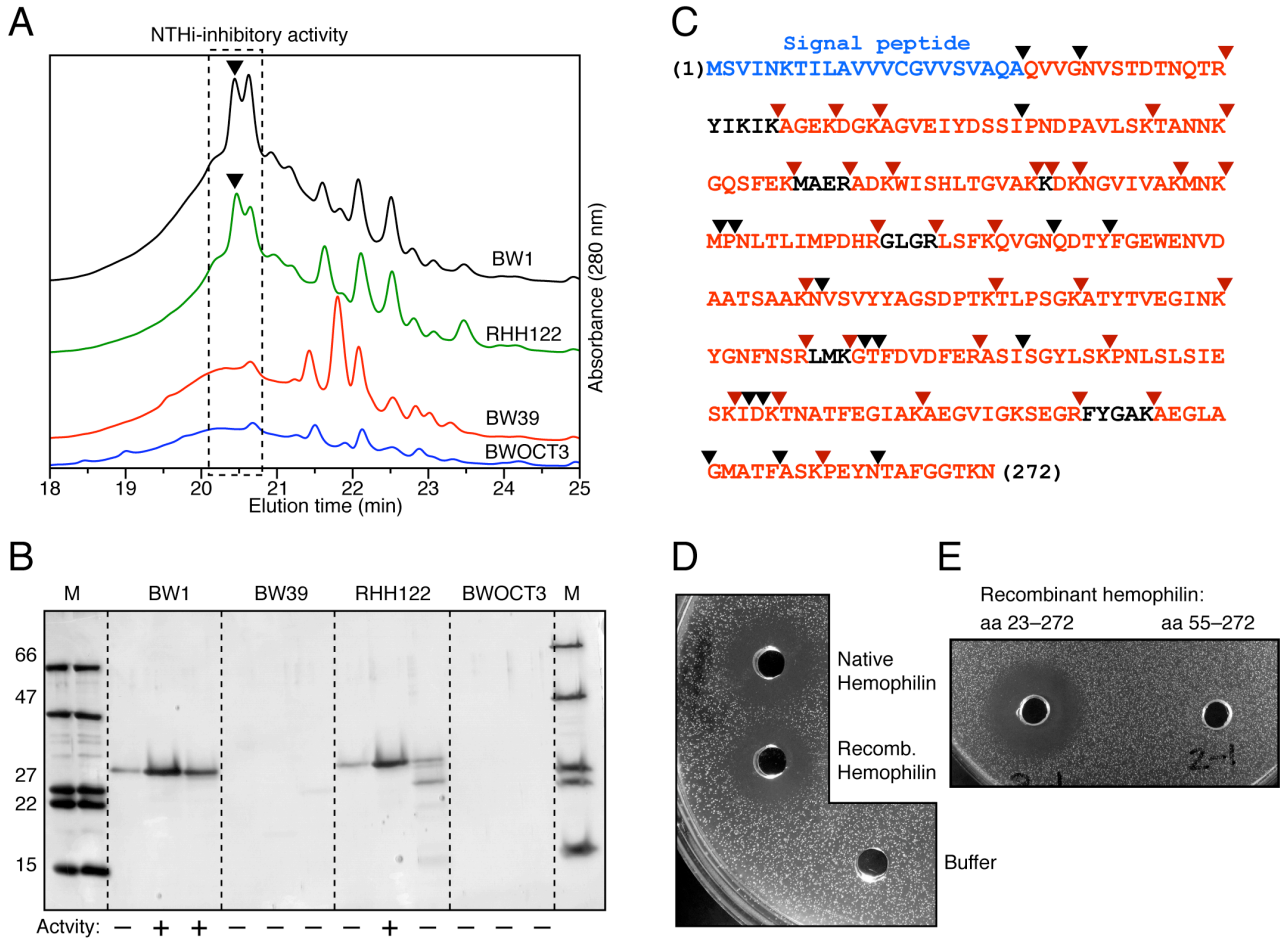

Fig. 1. Hemophilin is a 27-kDa protein isolated from *H. haemolyticus* that inhibits the growth of NTHi. (*A*) RP-HPLC profiles show NTHi-inhibitory activity (filled triangles) coincided with elution peaks present in BW1 and RHH122 samples, and absent from control strains BW39 and BWOCT3. (*B*) Tris-tricine SDS-PAGE and silver staining of RP-HPLC fractions indicated by a dashed box in A. Markers (M; kDa). (*C*) Tryptic peptides covering 92% of the sequence of hypothetical protein EGT80255 (red) were identified from purified hemophilin samples (red); canonical trypsin cleavage sites (red triangles) and other protein fragmentation sites (black triangles) are indicated. (*D*) NTHi-inhibitory activity obtained from 1 mL of *H. haemolyticus* culture (OD_{600} = 1.2) is approximately equivalent to 200 pmole of recombinant hemophilin (10 μM). (*E*) No NTHi growth inhibition was observed for truncated hemophilin.

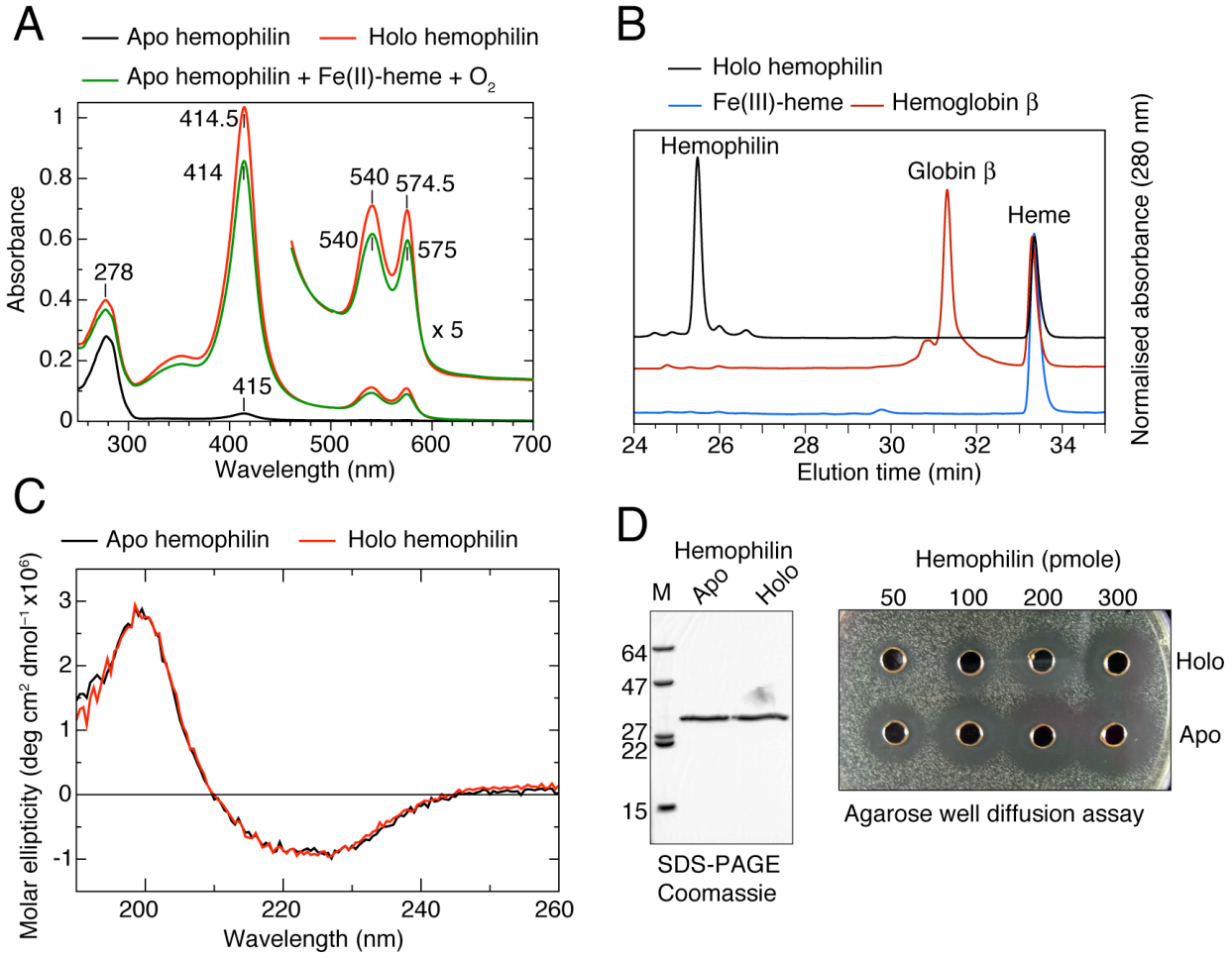

Fig. 2. Hemophilin is a heme-binding protein. (*A*) UV-visible absorption spectra of recombinant holo (red) and apo hemophilin (black). Addition of ferrous heme and O_2 to apo hemophilin yields a spectrum (green) essentially identical to that of holo hemophilin. (*B*) RP-HPLC shows heme is non-covalently bound to hemophilin; the retention time for the dissociated prosthetic group is identical to that of heme dissociated from human hemglobin β chain, or purified hemin. (*C*) CD spectra of holo and apo hemophilin (25 mM sodium phosphate, 125 mM NaF, pH 7.1) are highly similar, suggesting no large change in secondary structure upon heme binding. (*D*) Apo hemophilin has greater NTHi-inhibitory activity than the holo protein.

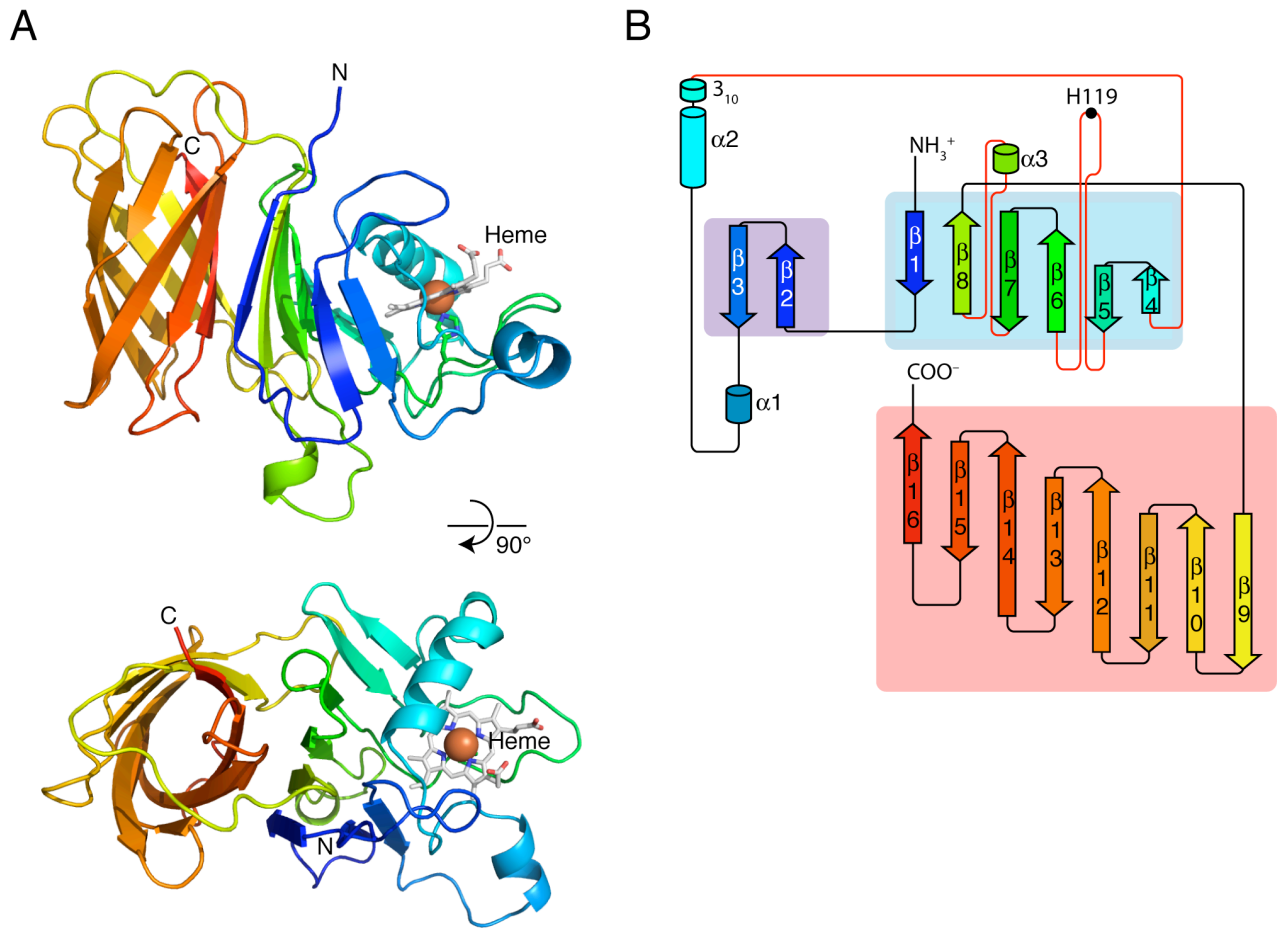

Fig. 3. Hemophilin is a new heme-binding fold. (*A*) Richardson diagrams of the hemophilin crystal structure in two orthogonal views. (*B*) Topology diagram of hemophilin prepared based on the output from PROORIGAMI; comparison with proteins of similar structure is shown in Fig. S12.

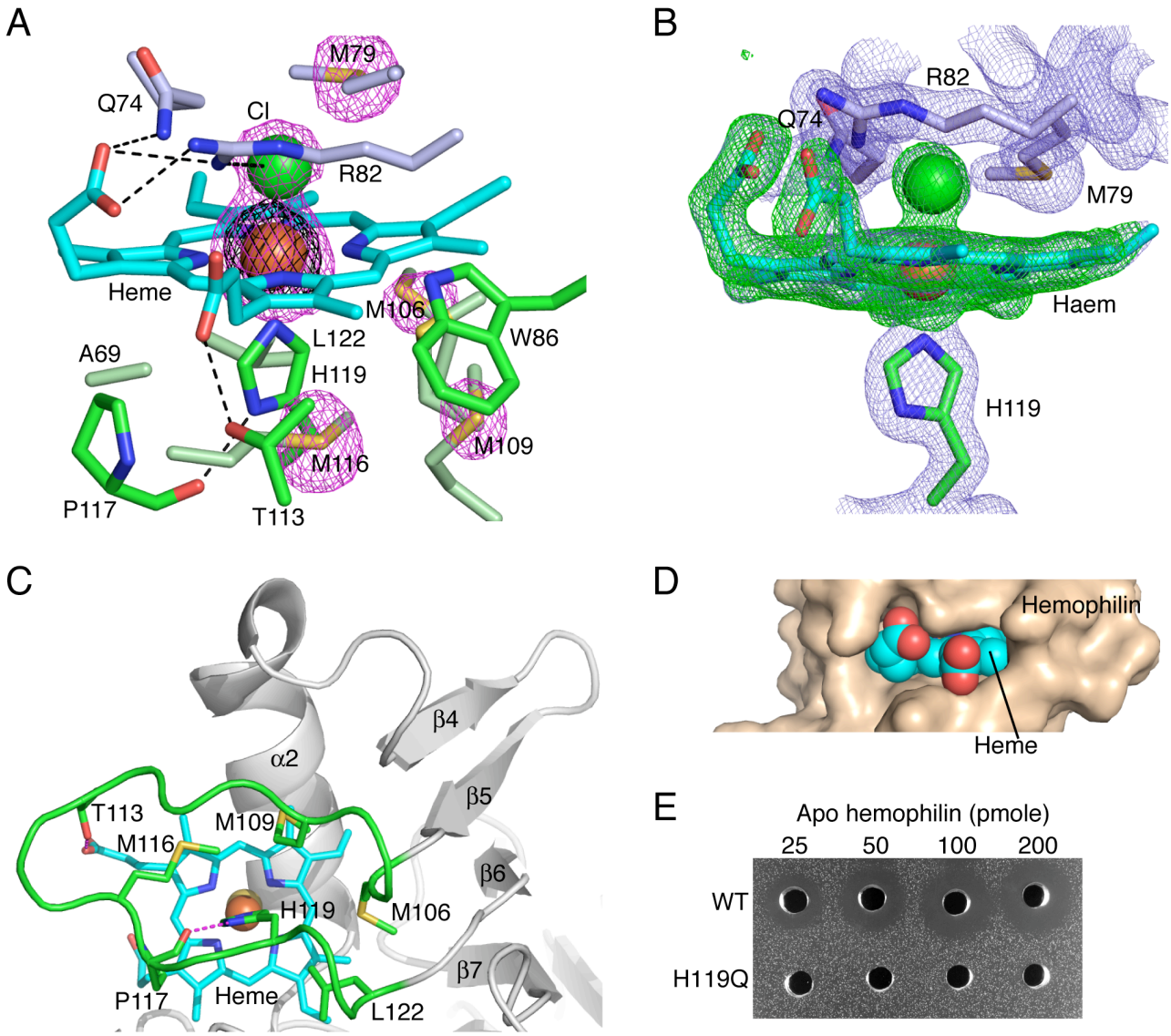

Fig. 4. Features of the hemophilin heme pocket. (*A*) Protein side chains make extensive contacts with both faces of the porphyrin (cyan sticks). Anomalous diffraction map contoured at 25 σ (black mesh) and 5 σ (pink mesh) identifies the position of the heme iron atom (orange sphere), S atoms of methionines, and a chloride ion (green sphere). (*B*) Figure shows $2F_O–F_C$ electron density map (contoured at 2 σ) and $F_O–F_C$ OMIT map (contoured at 4σ; refined without the heme and heme ligand). (*C*) All the molecular contacts between hemophilin and the H119-proximal side of the heme are provided by a single protein loop (green). (*D*) The heme group (spheres) is buried with the central iron obscured. (*E*) Mutation of the heme-coordinating histidine (H119Q) led to loss of NTHi-inhibtory activity.

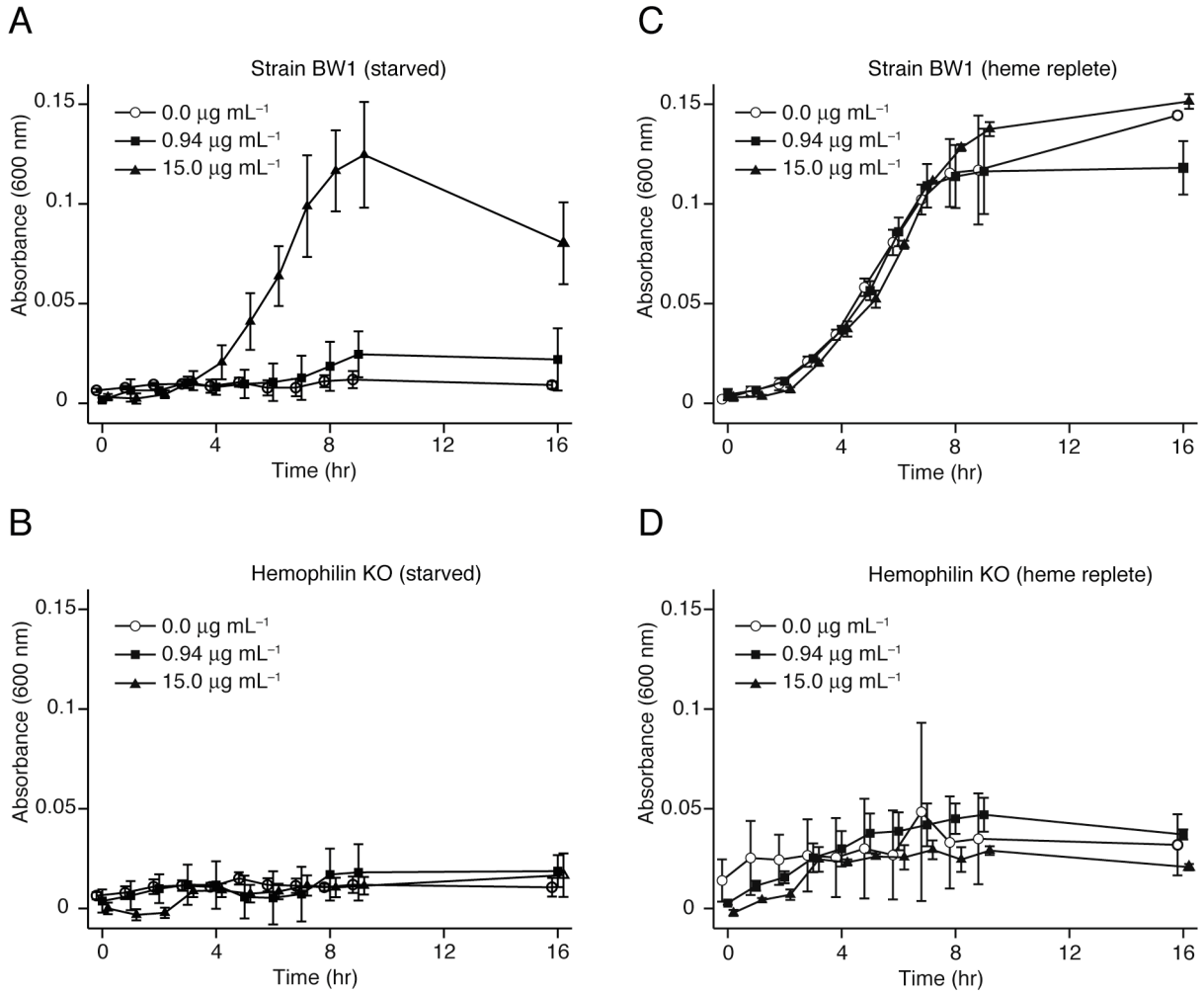

Fig. 5. A hemophilin gene knockout strain of BW1 has defects in heme utilisation. Heme-replete and heme-starved bacterial populations were inoculated into medium containing varying concentrations of hemin as follows: (*A*) hemin-starved BW1; (*B*) hemin-starved BW1 knockout; (*C*) hemin-replete BW1; (*D*) hemin-replete BW1 knockout. Error bars represent ±1 SD (n=3).