

Bacterial hemophilin homologs and their specific type eleven secretor proteins have conserved roles in heme capture and are diversifying as a family

Alex S. Grossman,¹ David A. Gell,² Derek G. Wu,^{1,3} Dana L. Carper,⁴ Robert L. Hettich,⁴ Heidi Goodrich-Blair¹

AUTHOR AFFILIATIONS See affiliation list on p. 26.

ABSTRACT Cellular life relies on enzymes that require metals, which must be acquired from extracellular sources. Bacteria utilize surface and secreted proteins to acquire such valuable nutrients from their environment. These include the cargo proteins of the type eleven secretion system (T11SS), which have been connected to host specificity, metal homeostasis, and nutritional immunity evasion. This Sec-dependent, Gram-negative secretion system is encoded by organisms throughout the phylum Proteobacteria, including human pathogens *Neisseria meningitidis*, *Proteus mirabilis*, *Acinetobacter baumannii*, and *Haemophilus influenzae*. Experimentally verified T11SS-dependent cargo include transferrin-binding protein B (TbpB), the hemophilin homologs heme receptor protein C (HrpC), hemophilin A (HphA), the immune evasion protein factor-H binding protein (fHbp), and the host symbiosis factor nematode intestinal localization protein C (NilC). Here, we examined the specificity of T11SS systems for their cognate cargo proteins using taxonomically distributed homolog pairs of T11SS and hemophilin cargo and explored the ligand binding ability of those hemophilin cargo homologs. *In vivo* expression in *Escherichia coli* of hemophilin homologs revealed that each is secreted in a specific manner by its cognate T11SS protein. Sequence analysis and structural modeling suggest that all hemophilin homologs share an N-terminal ligand-binding domain with the same topology as the ligand-binding domains of the *Haemophilus haemolyticus* heme binding protein (Hpl) and HphA. We term this signature feature of this group of proteins the hemophilin ligand-binding domain. Network analysis of hemophilin homologs revealed five subclusters and representatives from four of these showed variable heme-binding activities, which, combined with sequence-structure variation, suggests that hemophilins are diversifying in function.

IMPORTANCE The secreted protein hemophilin and its homologs contribute to the survival of several bacterial symbionts within their respective host environments. Here, we compared taxonomically diverse hemophilin homologs and their paired Type 11 secretion systems (T11SS) to determine if heme binding and T11SS secretion are conserved characteristics of this family. We establish the existence of divergent hemophilin sub-families and describe structural features that contribute to distinct ligand-binding behaviors. Furthermore, we demonstrate that T11SS are specific for their cognate hemophilin family cargo proteins. Our work establishes that hemophilin homolog-T11SS pairs are diverging from each other, potentially evolving into novel ligand acquisition systems that provide competitive benefits in host niches.

KEYWORDS heme, porphyrin, protein structure, ligand binding, nutritional immunity, secretion, metalloprotein

Editor Patricia A. Champion, University of Notre Dame, Notre Dame, Indiana, USA

Address correspondence to Heidi Goodrich-Blair, hgblair@utk.edu, or David A. Gell, david.gell@utas.edu.au.

The authors declare no conflict of interest.

See the funding table on p. 26.

Received 5 January 2024

Accepted 18 February 2024

Published 20 March 2024

Copyright © 2024 American Society for Microbiology. All Rights Reserved.

Many enzymes have evolved expanded catalytic potential through the incorporation of metallic cofactors and prosthetic groups. Iron cofactors are essential to most living organisms due to their functional contributions to enzymes required for DNA synthesis, photosynthesis, respiration, and nitrogen metabolism (1). Without the biochemical flexibility provided by metallic cofactors, life as we know it would be impossible. Because of this, and the limited bioavailability of essential metals, competition among organisms for these ions can be fierce. In some cases, including many marine environments, competition for bioavailable iron is the major limiting factor of microbial growth (2). Animals exploit the fact that iron is typically a growth-limiting nutrient for bacteria through a process known as nutritional immunity, in which valuable metals, such as iron, are sequestered to slow or deter the pathogenic growth of microbes. Within healthy animal hosts, iron is sequestered by proteins such as hemopexin, transferrin, lactoferrin, and ferritin (3). Medical conditions, such as hemochromatosis, that increase the serum iron concentration or prevent effective storage of iron increase a patient's risk of opportunistic infection by bacteria, including *Escherichia coli*, *Listeria monocytogenes*, and *Yersinia enterocolitica* (4). To overcome iron limitation within an animal host, bacteria have evolved means of countering nutritional immunity, including adaptations to use alternative catalytic metals (5), production of high-affinity siderophores (6), and/or membrane-bound uptake receptors (3) that facilitate the acquisition of iron from host metalloproteins.

The type eleven secretion system (T11SS) is a family of outer membrane proteins (OMPs) present throughout proteobacteria which possess a conserved 14-stranded β -barrel (7, 8). These proteins are both necessary and sufficient to secrete their cognate cargo proteins across the outer membrane (8–10). Recent studies have linked some T11SS and their cargo proteins to iron uptake strategies in Gram-negative bacteria. In *Neisseria*, the T11SS proteins Slam1 and Slam2 surface expose cargo proteins that are responsible for binding host-metal carriers: transferrin-binding protein B (TbpB) and lactoferrin-binding protein B are surface exposed by Slam1, and hemoglobin/haptoglobin-binding protein A (HpuA) is surface exposed by Slam2 (7, 11–13). These surface-exposed outer membrane lipoproteins facilitate bacterial colonization by capturing their respective host factors (transferrin, lactoferrin, or hemoglobin/haptoglobin) and complexing with a TonB-dependent uptake channel capable of importing the iron cofactor. Since surface exposure is essential for the function of these lipoproteins, genetic inactivation of the T11SS OMP, Slam1 prevents effective colonization and pathogenesis by *Neisseria* (10). While *Neisseria* Slam1 and Slam2 have specificity for their respective cargo (7), no underlying mechanism for specificity has been proposed yet and it is unknown if all T11SS have specificity for their cognate cargo. Bioinformatic analyses revealed a large number of potential T11SS-dependent cargo, lipid anchored and unanchored, which frequently exist in cognate pairs/groupings according to genomic co-occurrence analyses (8, 10). To date, all verified or predicted T11SS-dependent cargo have two distinct domains: an N-terminal domain that varies in predicted structure and ligand-binding function, and a C-terminal, 8-stranded β -barrel domain from either the TbpBBD or the so-called lipoprotein C families (7, 8, 10).

Although originally ascribed as facilitators of lipoprotein surface exposure, T11SS are capable of secreting unlipidated cargo proteins, such as the soluble hemophores heme receptor protein C from *Xenorhabdus nematophila* (HrpC) and hemophilin A from *Acinetobacter baumannii* (HphA) (8, 14). HphA likely captures heme from hemoglobin and other host hemoproteins and contributes to the virulence of *A. baumannii* in a murine infection model through its role as a co-receptor to the TonB-dependent heme receptor HphR (14). Thus, hemophilin proteins represent a high-affinity heme acquisition system comparable to HasA from *Serratia marcescens* (15) or IsdB from *Staphylococcus aureus* (16). Known members of the hemophilin protein family function to import heme from a host environment as depicted in Fig. 1. Compiling the results of published data from multiple organisms into a single model suggests that hemophilin crosses the inner membrane through the Sec translocon to reach the periplasm (8), may interact

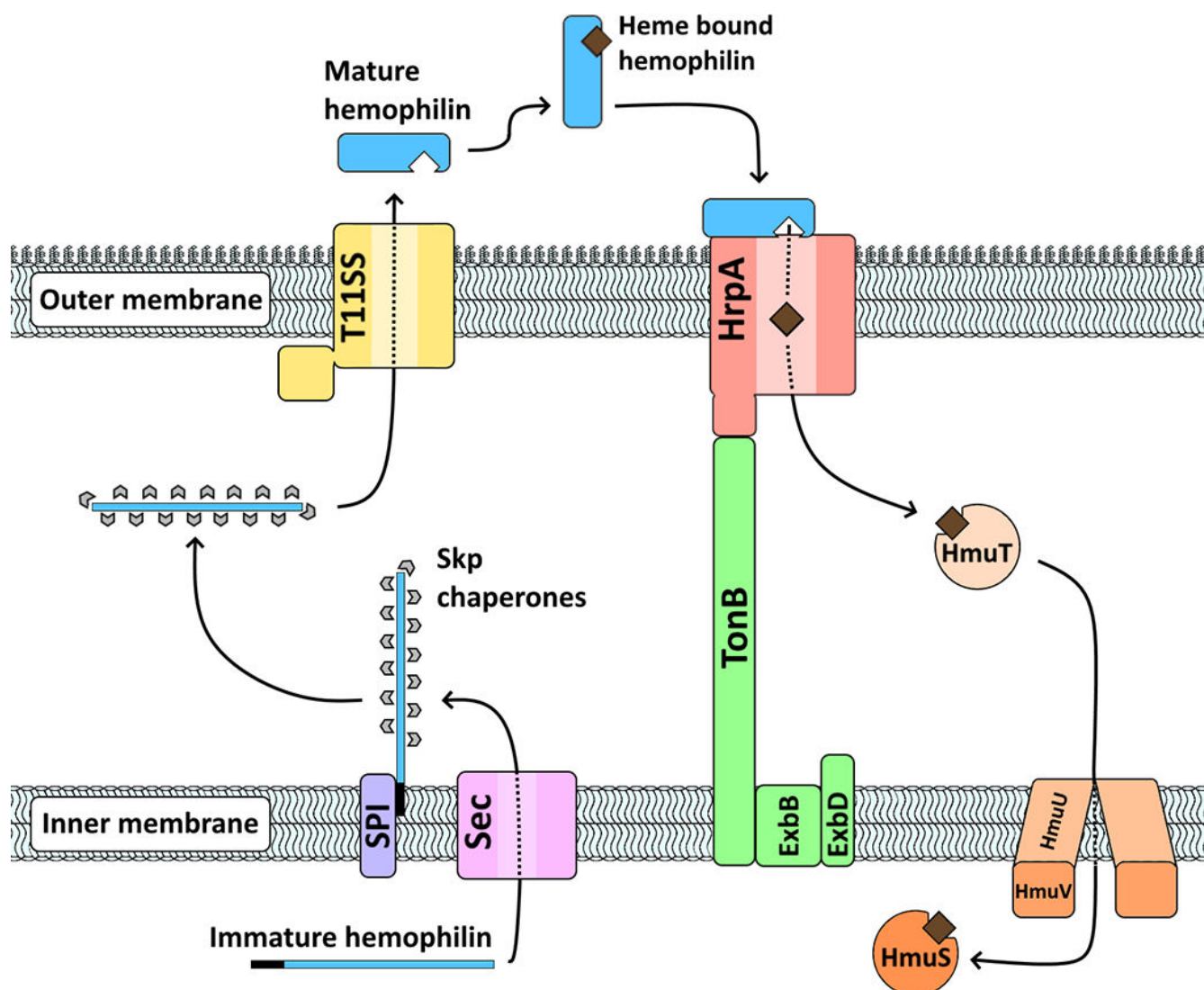


FIG 1 Conceptual model of hemophilin secretion and heme acquisition. The hemophilin homologs examined in this study (HrpC, Hpl, HsmA, and CrpC) are translated with a signal peptide directing them to the Sec translocon and signal peptidase I. This allows them to cross the inner membrane in an unfolded state. These hemophilin proteins may maintain an unfolded state in the periplasm through interaction with Skp chaperones, based on analogy to the T11SS-dependent lipoprotein TbpB (17). Each hemophilin protein we examined has a cognate T11SS which translocates it through the outer membrane when present (HrpB_{X,nem}, HrpB_{H,haem}, HphA, and CrpB, respectively). Mature apo-hemophilin proteins then bind extracellular heme, becoming holo-hemophilin. Holo-hemophilin interacts with a cognate TonB-dependent co-receptor for heme uptake across the outer membrane. While these TonB-dependent co-receptors are nearly ubiquitous in hemophilin encoding genomic loci (92.3%), they are absent in our expression strain of *E. coli*, and these uptake steps, therefore, are not expected to occur in our experiments. After uptake, periplasmic heme would likely be transported across the inner membrane by the heme utilization system (Hmu). Once brought into the cytoplasm, heme can be degraded by heme oxygenase (HmuS), an enzyme which is regularly encoded within hemophilin loci (28.5%).

with chaperones such as Skp to traverse the periplasm (17), and then crosses the outer membrane in a T11SS-dependent manner to reach the extracellular milieu (7, 8, 14, 18). From here, hemophilin captures heme that is released from hemoglobin or other host hemoproteins using a high-affinity-binding domain (14, 19). Holo-hemophilin is predicted to interact with a TonB-dependent outer membrane co-receptor which imports the heme molecule into the periplasm and releases apo-hemophilin. Finally, the heme utilization system delivers periplasmic heme into the cytoplasm for incorporation into cellular processes or digestion by heme oxygenase to free the iron cation (20).

Within this overall framework, hemophilin can have diverse functional roles within different organisms and environments. For example, in *Haemophilus haemolyticus*, hemophilin can act as a probiotic factor by making bioavailable iron inaccessible to nontypeable *Haemophilus influenzae* (21). Hemophilin-producing strains of *H. haemolyticus* inhibit the growth of *H. influenzae* significantly more than do non-hemophilin encoding strains in co-cultured media (19) and cell cultures (21). Additionally, oropharyngeal sampling of human subjects indicates that individuals who carry hemophilin encoding *H. haemolyticus* are approximately twofold less likely to carry nontypeable *H. influenzae* (22). Conversely, within the opportunistic pathogen *A. baumannii*, hemophilin can act as a virulence factor by facilitating systemic infection in a murine model (14). Furthermore, predicted hemophilin homologs found in sequence databases display sequence variation within the heme-binding handle domain, suggesting variability in ligand binding. To better understand the fundamental biochemical and biological functions of hemophilin family proteins, we explored a series of hemophilin homologs selected to cover a broad taxonomic range and sequence level divergence.

RESULTS

Sequence similarity networks reveal hemophilin families that have genomic associations with metal-related metabolic functions

To explore the relatedness of hemophilin family proteins and to identify subcluster divisions that may reflect divergent function, a sequence similarity network generated through Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST) was overlaid with a taxonomic framework (23) (Fig. S1; Supplemental File 1). The network was populated with the previously published data set of T11SS-associated cargo that were not predicted to be lipidated or membrane anchored (8). This analysis revealed a single major cluster containing all previously described hemophilin proteins (88/107 nodes), one smaller cluster containing uncharacterized proteins from predominantly *Pseudomonas* and *Neisseria* species (9/107 nodes), and a few unassociated doublets and singletons (10/107 nodes). To focus this study specifically on hemophilin and its direct evolutionary relatives, all nodes not within the central cluster were removed. The remaining nodes were labeled according to taxonomic family and separated using the ForceAtlas2 force-directed separation algorithm, resulting in five subclusters predominantly populated by seven families from Alphaproteobacteria, Betaproteobacteria, and Gammaproteobacteria (Fig. 2A through C). The three subclusters that contained Hpl, HrpC, and HphA, respectively, were named Hpl-like, HrpC-like, and HphA-like after their respective characterized member. One novel subcluster was termed “Cobalt/molybdenum associated” due to the genomic co-occurrence of its members with genes predicted to encode cobalt- or molybdenum-dependent enzymes. Another novel subcluster was termed “Plant/Environmental” due to the dominant presence of homologs encoded by microbes found in soil, water, and plant-associated environments. The subclusters did not fall exclusively along taxonomic lines. For instance, the Hpl-like subcluster includes *Neisseria* and *Pasteurella* hemophilin homologs, indicating that these genes may have been horizontally exchanged.

To identify potential distinguishing features among the five subclusters, we examined the identities of genes that commonly occur within genomic neighborhoods (± 6 genes) surrounding the hemophilin homologs in our data set. All sequences from the network were submitted to the Rapid ORF Detection & Evaluation Online web tool that uses profile hidden Markov models to identify co-occurring protein domains (Table S1) (24). All hemophilin homologs had a T11SS OMP encoded within their genomic neighborhood since the original data set was generated by searching for predicted T11SS-cargo pairings (8). Consistent with previous observations, 92.3% (240/260 sequences) of hemophilin homologs, regardless of subcluster, were encoded in association with genes predicted to encode TonB and TonB-dependent receptors, suggesting a strong and consistent link between T11SS cargo and TonB-dependent uptake across the outer membrane (8, 10, 19). However, some subclusters had further

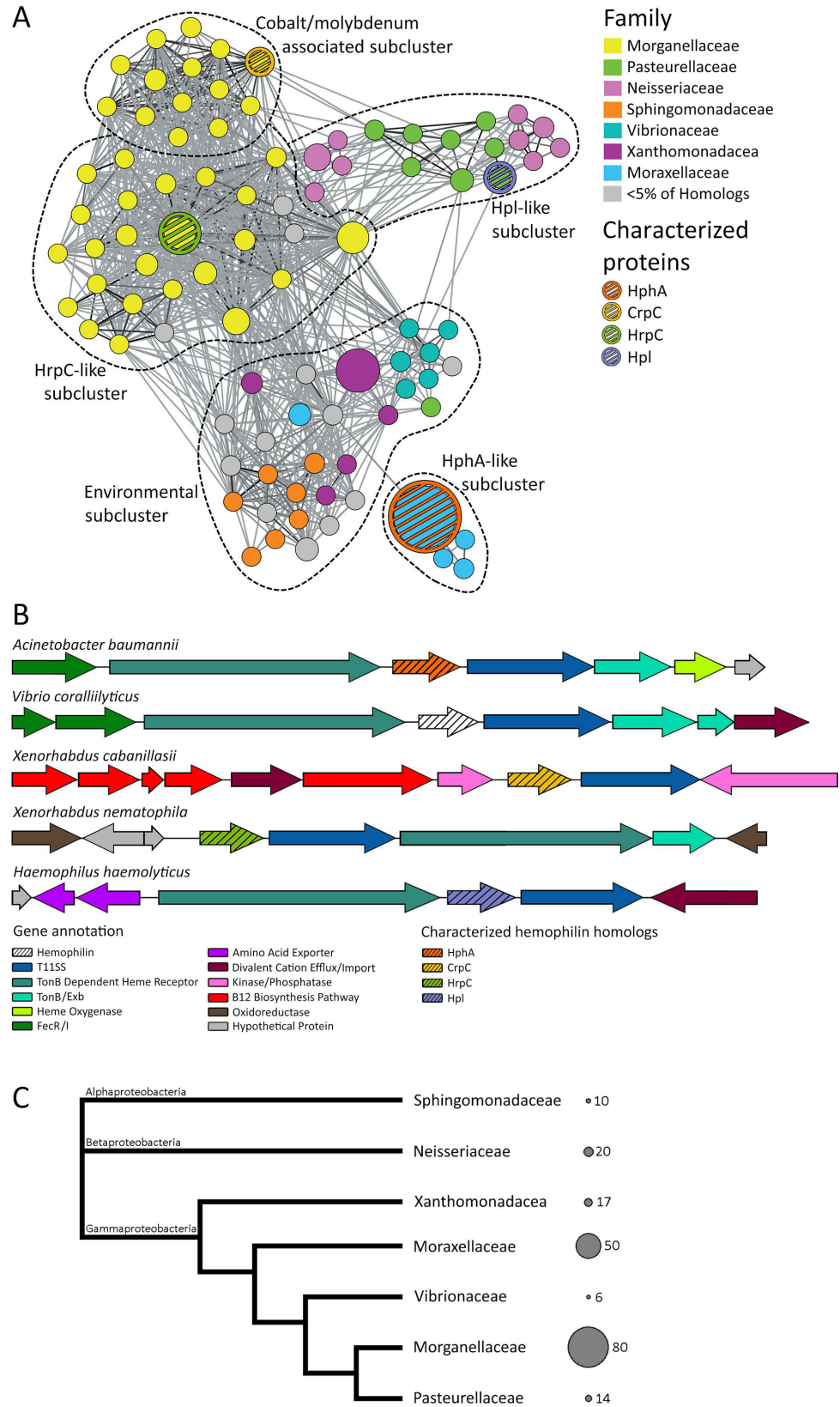


FIG 2 Distribution and relatedness of hemophilin family proteins. (A) A sequence similarity network of hemophilin homologs generated with EFI-EST. Each node represents one or more protein sequences with 80% or greater identity, the larger the node the more sequences it contains. Edges indicate an alignment score of 35 or greater. Edge darkness indicates shared (Continued on next page)

FIG 2 (Continued)

sequence identity, with the darkest edges being the most identical. Dotted lines indicate proposed subclusters as defined using the ForceAtlas2 force-directed algorithm and the distribution of characterized proteins. (B) Representative genomic neighborhoods from the subclusters identified within the sequence similarity network demonstrating common co-occurring genes. (C) A cladogram of the seven genera which encode the most hemophilin family proteins, split between Alpha-, Beta-, and Gammaproteobacteria. The circles and numbers to the right of the cladogram indicate the relative abundance of known hemophilin homologs encoded by each genus.

informative co-occurrences. For example, hemophilin homologs in the HphA-like subcluster showed nearly universal co-occurrence with genes predicted to encode heme oxygenase (51/53) and the iron-sensing regulator FecR (48/53). Similarly, homologs in the Plant/Environmental subcluster typically were encoded near genes predicted to encode metal responsive regulatory proteins (FecR or Fur) (64/72) and occasionally near genes encoding heme oxygenase (19/72). Additionally, 69/72 Plant/Environmental-subcluster loci encoded additional regulatory genes such as RpoE, IscR-family regulators, and LysR-family regulators. RpoE is a sigma factor that responds to extra-cytoplasmic stress and is essential for metal resistance in *E. coli* (25). IscR regulates iron-sulfur cluster biosynthesis according to cellular demand (26), and LysR-family regulators drive diverse pathways by binding DNA directly in response to co-inducing/co-repressing ligands (27). Homologs in the Cobalt/molybdenum-associated subcluster occurred alongside other predicted T11SS-dependent cargo (14/18) and were either located adjacent to a B12 biosynthetic locus (3/18) or near a formate dehydrogenase locus (13/18). Vitamin B12, also known as cobalamin, is a bacterially derived nutrient which contains a central corrin ring containing a single atom of cobalt. This corrin ring has structural similarity to the porphyrin ring of heme which contains a single atom of iron. Formate dehydrogenase activity relies on molybdenum cofactors, which typically complex with molybdenum or tungsten atoms. The co-occurrence of both B12 and formate dehydrogenase loci with T11SS OMPs may hint at a role for T11SS-dependent cargo in the acquisition of metal ions other than iron. The HrpC-like subcluster includes genes predicted to encode redox enzymes, such as formate dehydrogenase (48/81) and NADPH:quinone oxidoreductase (12/81), tRNA synthases and modification systems (30/81), and regulatory proteins including TetR family regulators, FaeA family regulators, FecR, and FecI (32/81). Homologs in the Hpl-like subcluster had few unifying co-occurrences; however, many co-occurred with tRNA synthases/modification systems (14/36), specifically selenocysteine tRNA synthases (6/36) (Fig. 2B). Overall, network analysis of hemophilin homologs indicates that the family is diversifying into distinct subfamilies independent of their taxonomic lineage and that those subfamilies are genomically associated with metal-related cellular activities such as ion uptake, metal responsive regulatory proteins, and metal-dependent enzymes.

Hemophilin homologs are transported by their cognate type eleven secretor proteins

Hemophilin homologs in each of the network subfamilies are predicted to be secreted by a T11SS, and many were encoded adjacent to a T11SS OMP. We considered the possibility that co-diversification of hemophilins and their T11SS OMPs has resulted in specificity between cognate pairs. To investigate specificity between hemophilin cargo proteins and their paired T11SS OMP secretors, we chose representative hemophilin homologs from each of the subclusters identified in the sequence similarity network (except for the Plant/Environmental subcluster) and expressed them within *E. coli*; either in isolation, co-expressed alongside their cognate T11SS, or co-expressed alongside a non-cognate T11SS from *X. nematophila*. T11SS/hemophilin cargo pairs from *X. nematophila* (HrpB_{X.nem}/HrpC), *H. haemolyticus* (HrpB_{H.haem}/Hpl), *A. baumannii* (HsmA/HphA), and *Xenorhabdus cabanillasii* (CrpB/CrpC) were cloned, each with a C-terminal FLAG-tag insertion, into pETDuet-1-based expression vectors to perform co-expression and secretion experiments in *E. coli* BL21 C43. Additionally, plasmids were

constructed to co-express HrpB_{X,nem} alongside the non-cognate hemophilin homologs Hpl, HphA, and CrpC. Western blotting of supernatant and cellular lysates was performed using anti-FLAG antibodies to detect both the cargo and T11SS proteins. As previously observed with HrpB expression (8), occasionally, the supernatant fraction contained T11SS-FLAG proteins despite them being outer membrane anchored. This reflects the presence of outer membrane vesicles or membrane fragments which, when removed, did not impact T11SS-secreted cargo levels (8).

In each expression assay, the observed migration of hemophilin cargo proteins varied from predictions. Extracellular HrpC and HphA each ran as a single-protein band at ~30 kD (~24 kD theoretical) and ~29 kD (~25 kD theoretical), respectively (Supplemental file 2: sheets 2 and 4) (27). The ~30 kD band was verified by mass spectrometry to be HrpC (accounting for the most abundant ions in the sample, with 46 unique peptides and 80% peptide coverage). Extracellular Hpl appeared as two bands, a predominant band at ~32 kD and a minor one at ~26 kD, both of which were verified through mass spectrometry (56% and 52% peptide coverage, respectively) to include the full-length mature protein (excluding the signal peptide). Extracellular CrpC appeared as ~19 and ~27 kD bands (Supplemental file 2: sheets 3 and 5) that were both identified as CrpC by mass spectrometry (39% coverage and 40% coverage, respectively). Mass spectrometry allowed us to confirm that the detected protein bands contained the expected hemophilin homologs but did not reveal protein modifications (e.g., peptide cleavage) that could explain their migration as multiple bands. To accurately reflect the levels of Hpl and CrpC cargo proteins in our calculations, we opted to sum both protein bands within each condition.

We found that co-expression of hemophilin homologs with a cognate T11SS protein always significantly increased the levels of cargo protein found in the supernatant though the relative impact varied greatly among T11SS proteins (HrpB_{X,nem}/HrpC: 34.2-fold, HrpB_{H,haem}/Hpl: 59.3-fold, HsmA/HphA: 4.6-fold, CrpB/CrpC: 56.9-fold) (Fig. 3; Supplemental File 2). HrpB_{X,nem} significantly increased (4.8-fold) the average extracellular levels of the non-cognate cargo Hpl though it was significantly less effective at

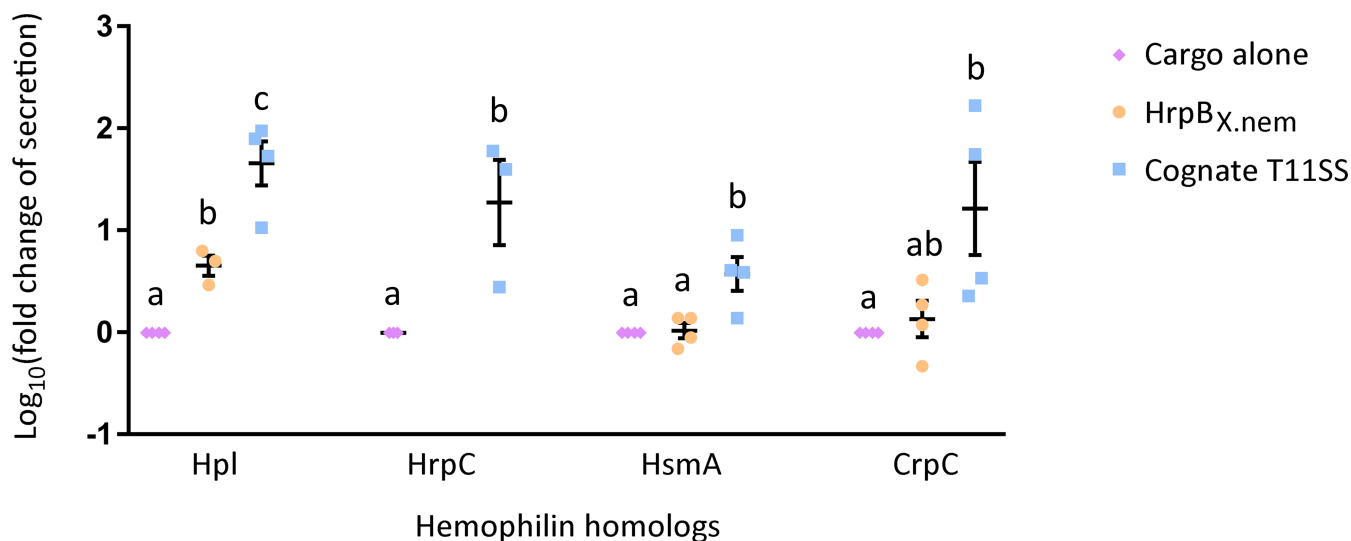


FIG 3 Secretion of hemophilin proteins by cognate T11SS proteins or a non-cognate T11SS protein. Each cargo protein was expressed in *E. coli* in isolation (cargo alone; pink diamonds) or co-expressed with their cognate T11SS (blue squares) or with non-cognate HrpB_{X,nem} (orange circles). Supernatant proteins were precipitated and quantified via immuno-blotting with anti-FLAG antibody to detect extracellular cargo proteins. Fold change of secretion was determined by dividing the amount of extracellular cargo detected in co-expression treatments by the amount seen in the respective cargo-alone treatment. Data were transformed with a log₁₀ function prior to performing a Tukey's HSD test for each hemophilin homolog. Letters indicate significance groups when comparing all treatments containing the indicated hemophilin homolog. The error bars indicate the standard error of the mean. All four T11SS proteins assayed significantly increased the amount of cognate cargo that was present in the extracellular milieu. The non-cognate T11SS HrpB_{X,nem} significantly increased extracellular levels of Hpl relative to no T11SS, but less effectively than did the cognate T11SS of Hpl. HrpB_{X,nem} did not significantly affect extracellular levels of HsmA or CrpC.

doing so than the cognate T11SS HrpB_{H.haem}. In contrast, HrpB_{X.nem} did not significantly impact average extracellular levels of non-cognate cargo: HphA (1.1-fold) and CrpC (1.7-fold). These data show that T11SS from four representative species facilitate transport of a cognate hemophilin cargo to the extracellular milieu and that T11SS from *X. nematophila* selectively transports HrpC over cargo from other species. These data are consistent with the hypothesis that T11SS secretor proteins display specificity for cognate cargo proteins.

The hemophilin C-terminal β -barrel domain may contribute to secretion specificity by T11SS

Published literature has implicated the C-terminal β -barrel domain of cargo proteins in directing secretion by T11SS OMPs (10). To assess the role of the hemophilin C-terminal domain on T11SS specificity, two chimeric hemophilin cargo were engineered. The first chimeric cargo had the N-terminal handle domain from HrpC and the C-terminal β -barrel domain from Hpl (henceforth HrpC-Hpl), while the second had the N-terminal handle from Hpl and the C-terminal β -barrel domain from HrpC (henceforth Hpl-HrpC). pETDuet-1 constructs, with C-terminal FLAG-tagged T11SS and cargo, were assembled to independently co-express HrpB_{H.haem} and HrpB_{X.nem} alongside both chimeric cargo proteins (Fig. 4A). Western blotting of supernatant and cellular lysates was again performed to monitor cargo secretion (Supplemental file 2: sheet 6). The HrpC-Hpl chimera ran as a single band at ~27 kD (~25 kD theoretical). The Hpl-HrpC chimera ran as two bands with apparent sizes of ~33 and ~24 kD (~27 kD theoretical). Since Hpl-HrpC ran as two bands, similarly to Hpl, we summed both bands for the purpose of quantification. Co-expression with HrpB_{H.haem} increased, on average, the extracellular levels of its cognate cargo Hpl by 65.0-fold, the HrpC-Hpl chimera by 51.0-fold, and the Hpl-HrpC chimera by 7.8-fold, indicating that HrpB_{H.haem} was more effective at transporting a chimera with the cognate C-terminal domain (here the C-terminal domain of Hpl). The observed differences in secretion between chimeric cargo proteins were not as profound as those seen in the cognate vs non-cognate secretion experiment,

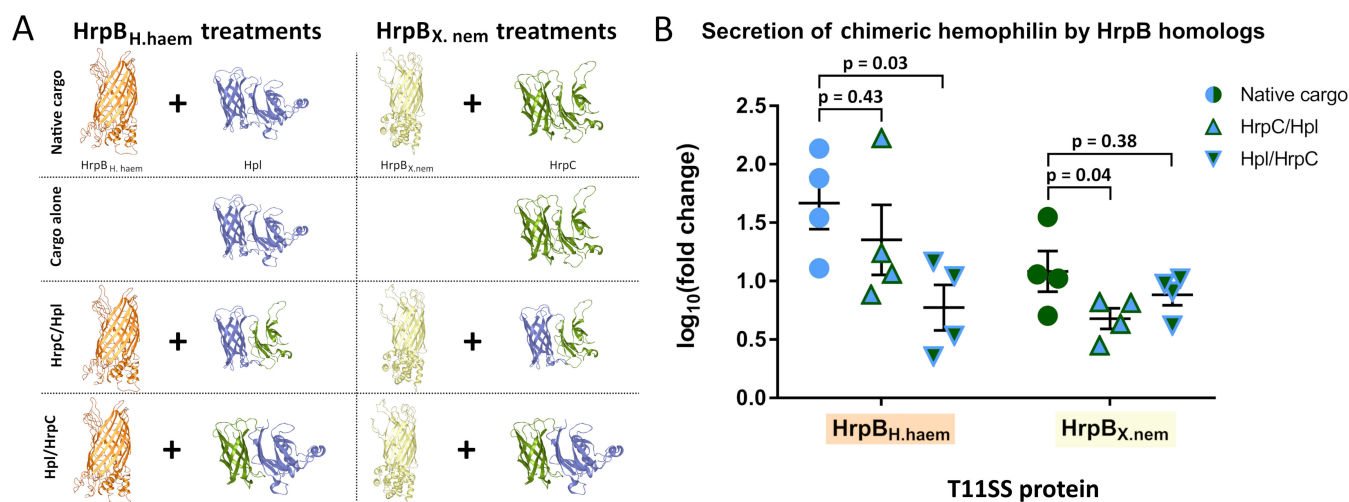


FIG 4 Secretion of domain-swapped chimeric hemophilin proteins by HrpB_{X.nem} and HrpB_{H.haem}. (A) A pictorial depiction of the experimental treatments used to assess secretion of chimeric hemophilin cargo proteins by HrpB T11SS secretors from *H. haemolyticus* (HrpB_{H.haem}; orange) and *X. nematophila* (HrpB_{X.nem}; yellow). Each T11SS protein was co-expressed alongside its cognate cargo protein [Hpl (blue) and HrpC (green), respectively], and two chimeric cargo proteins (HrpC-Hpl and Hpl-HrpC) generated by swapping the two domains (N-terminal effector-C-terminal barrel) of Hpl and HrpC. Fold change of secretion was determined by immunoblotting with anti-FLAG antibodies and dividing the amount of extracellular cargo in detected co-expression treatments by the amount seen in the respective cargo-only treatment. (B) Data were transformed with a log₁₀ function prior to performing a Kruskal-Wallis ANOVA with an uncorrected Dunn's comparison to the secretion of the cognate cargo protein. The error bars indicate the standard error of the mean. Chimeric hemophilin proteins were preferentially secreted by the T11SS that was cognate to the C-terminal barrel domain they contained. However, neither chimeric protein (triangles) was as effectively secreted as the native (non-chimeric; circles) hemophilin protein of each T11SS.

so in order to compare treatments we opted to perform a Kruskal-Wallis ANOVA with an uncorrected Dunn's test. This analysis focused strictly on comparing the secretion of chimeric proteins to the secretion of cognate cargo instead of directly comparing chimeras. Co-expression with HrpB_{X.nem} increased on average the extracellular levels of the cognate HrpC by 15.6-fold, the HrpC-Hpl chimera by 5.1-fold, and the Hpl-HrpC chimera by 8.1-fold on average (Fig. 4B; Supplemental File 2). Again, a higher level of secretion was seen when the C-terminal domain of the chimeric protein was in the presence of its cognate HrpB transporter. Thus, in both cases, we observed that the C-terminal β -barrel domain of the cargo had a dominant role in determining the level of secretion although this level was always lower than for the native hemophilin homolog. It is possible that T11SS/cargo interactions also occur outside the cargo C-terminal domain or that interdomain interactions within the chimeras result in loss of structural integrity that reduces transport efficiency. Nevertheless, our results are consistent with a conclusion that the C-terminal β -barrel domain of hemophilin homologs helps direct cargo for T11SS-mediated secretion (28).

Hemophilin homologs Hpl, HphA, and HrpC, but not CrpC, bind heme

We next explored the molecular function of the N-terminal domains of hemophilin homologs. To begin to investigate if the heme-binding properties of Hpl and HphA are conserved in other hemophilin homologs, we expressed Hpl from *H. haemolyticus*, HrpC from *X. nematophila*, HphA from *A. baumannii*, and CrpC from *X. cabanillasii* (without signal peptides) in the cytoplasm of *E. coli* and purified these. As expected, Hpl from *H. haemolyticus* was recovered from *E. coli* cytoplasm as an approximately 50:50 mix of heme-bound and heme-free protein, with the level of heme saturation likely reflecting competition for heme binding and limitations of heme biosynthesis *in vivo* (19). Preparations of HrpC (*X. nematophila*) and HphA (*A. baumannii*) had a brownish appearance and an absorbance peak at ~413 nm, indicating the presence of sub-saturating levels of a porphyrin ligand, whereas CrpC (*X. cabanillasii*) was colorless with no peaks in the visible absorption spectrum, indicating the lack of a porphyrin. Heme-free (apo-protein) preparations of Hpl, HphA, and HrpC were produced by acid-acetone extraction and reversed-phase HPLC, and heme-binding activities were investigated by titration (Fig. 5; Table 1). Large changes in the UV-visible spectrum of hemin occurred upon titration with *H. haemolyticus* Hpl, *A. baumannii* HphA, or *X. nematophila* HrpC (Fig. 5A through C). Similarities in the Soret (412–414 nm) and Q-band regions (500–600 nm) between HrpC and HphA suggest that the heme coordination structure of HrpC is similar to that in HphA (14). The binding curves for Hpl, HphA, and HrpC (Fig. 6A) yielded K_d values of 9, 7, and 20 nM respectively; however, the curves were close to linear, indicating that binding might be too strong to reliably extract binding constants. Using simulated data with added Gaussian noise, we determined that simulated data generated from K_d values <15 nM produced fits that did not differ significantly (at $P \leq 0.05$ by F test) from fits where the K_d parameter was fixed at an extreme low value. Thus, we suggest that Hpl and HphA bind heme with K_d values ≤ 15 nM and HrpC binds with $K_d = 20$ (10–50) nM. In contrast, spectral changes upon the addition of CrpC to heme were more gradual (Fig. 5D) and were fit with a one-to-one binding model with apparent $K_d \approx 5$ μ M (Fig. 6A). These values are similar to the binding affinity we determined for BSA with apparent $K_d \approx 2$ μ M (Fig. S2).

To further distinguish between the porphyrin-binding affinities of Hpl, HphA, and HrpC, we assayed binding to Zn(II)PPIX, a fluorescent heme analog, in the presence of competition from excess BSA (Fig. 7). We reasoned that fluorescence detection and competition binding would shift the useful detection range to a high-affinity regime, compared to the previous absorption design. These experiments (Fig. 7) yielded IC₅₀ values for hemophilin-binding Zn(II)-PPIX in competition with BSA as shown in Table 1, together with calculated affinities based on the affinity of BSA for Zn(II)PPIX. The IC₅₀ values for Hpl (10 nM) and HphA (20 nM) did not differ within error of the measurements, suggesting that Hpl and HphA bind Zn(II)-PPIX with similar affinity. In comparison, the

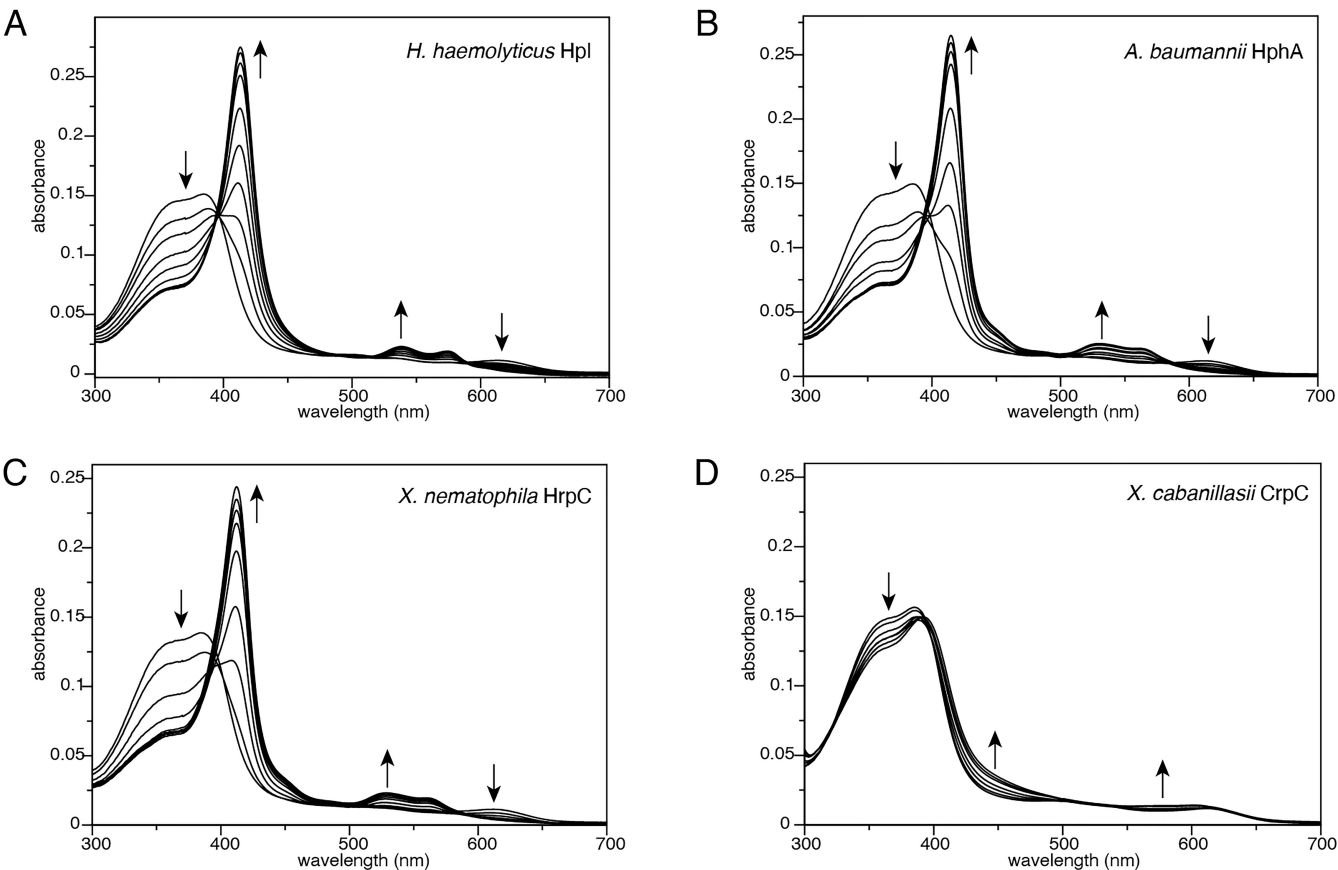


FIG 5 Titration of hemophilin homologs with hemin. Titrations of hemophilin proteins into hemin (Fe(III)-PPIX) solution (2.5 μ M hemin in 20 mM Tris-HCl, pH 8.0 at 21°C) were monitored by UV-visible absorption spectroscopy. (A) Spectra recorded after the addition of *H. haemolyticus* Hpl to a final concentration in the range of 0.3–4.5 μ M. (B) Spectra recorded after the addition of *A. baumannii* HphA to a final concentration in the range of 0.4–3.8 μ M. (C) Spectra with the addition of *X. nematophila* HrpC in the concentration range of 0.2–5.4 μ M. (D) Spectra with the addition of *X. cabanillasii* CrpC in the concentration range of 0.3–9.1 μ M. Arrows indicate the direction of spectral changes.

IC₅₀ value for *X. nematophila* HrpC binding to Zn(II)-PPIX (110 nM) was approximately 5- to 10-fold higher, indicating weaker binding to Zn(II)-PPIX. Using the same fluorescence competition assay, the interaction of CrpC with Zn(II)-PPIX was undetectable, consistent with the finding that CrpC and BSA have similar propensities to bind heme. In summary, the above results suggest that heme-binding affinities of the hemophilin homologs proceed from higher to lower affinity in the order Hpl $\approx$ HphA > HrpC >> CrpC.

To investigate the importance of the porphyrin metal to hemophilin binding, we performed spectroscopic titrations with unmetallated PPIX. Additionally, because the

TABLE 1 Relative porphyrin binding affinity of hemophilin homologs^d

Protein	Hemin <i>K_d</i> (nM) ^a	Zn(II)-PPIX <i>IC</i> ₅₀ (nM) ^b	Zn(II)-PPIX <i>K_d</i> (nM) ^a	Zn(II)-PPIX <i>K_d</i> (nM) ^c	PPIX <i>K_d</i> (nM) ^a	Co(III)-PPIX <i>K_d</i> (nM) ^a
Hh Hpl	≤ 15	10 (1–20)	N/A	0.2	70 (30–140)	≤ 15
Ab HphA	≤ 15	20 (2–50)	N/A	0.4	≤ 15	30 (10–50)
Xn HrpC	20 (10–50)	110 (70–180)	N/A	2.2	20 (10–30)	24 (20–29)
Xc CrpC	$5 (1-9) \times 10^3$	N/A	$3 (2-4) \times 10^2$	N/A	$7 (3-11) \times 10^3$	24 (21–26)
BSA	$2 (1-5) \times 10^3$	N/A	$3 (2-4) \times 10^2$	N/A	$11 (6-14) \times 10^3$	70 (50–80)

^aCalculated from absorbance data.

^bCalculated from fluorescence measurements in competition with BSA.

^cCalculated from IC₅₀ values according to (29), using measured *K_d* = 0.3 μ M for BSA binding Zn(II)PPIX and [BSA] = 15 μ M.

^dValues in parentheses are 95% confidence intervals.

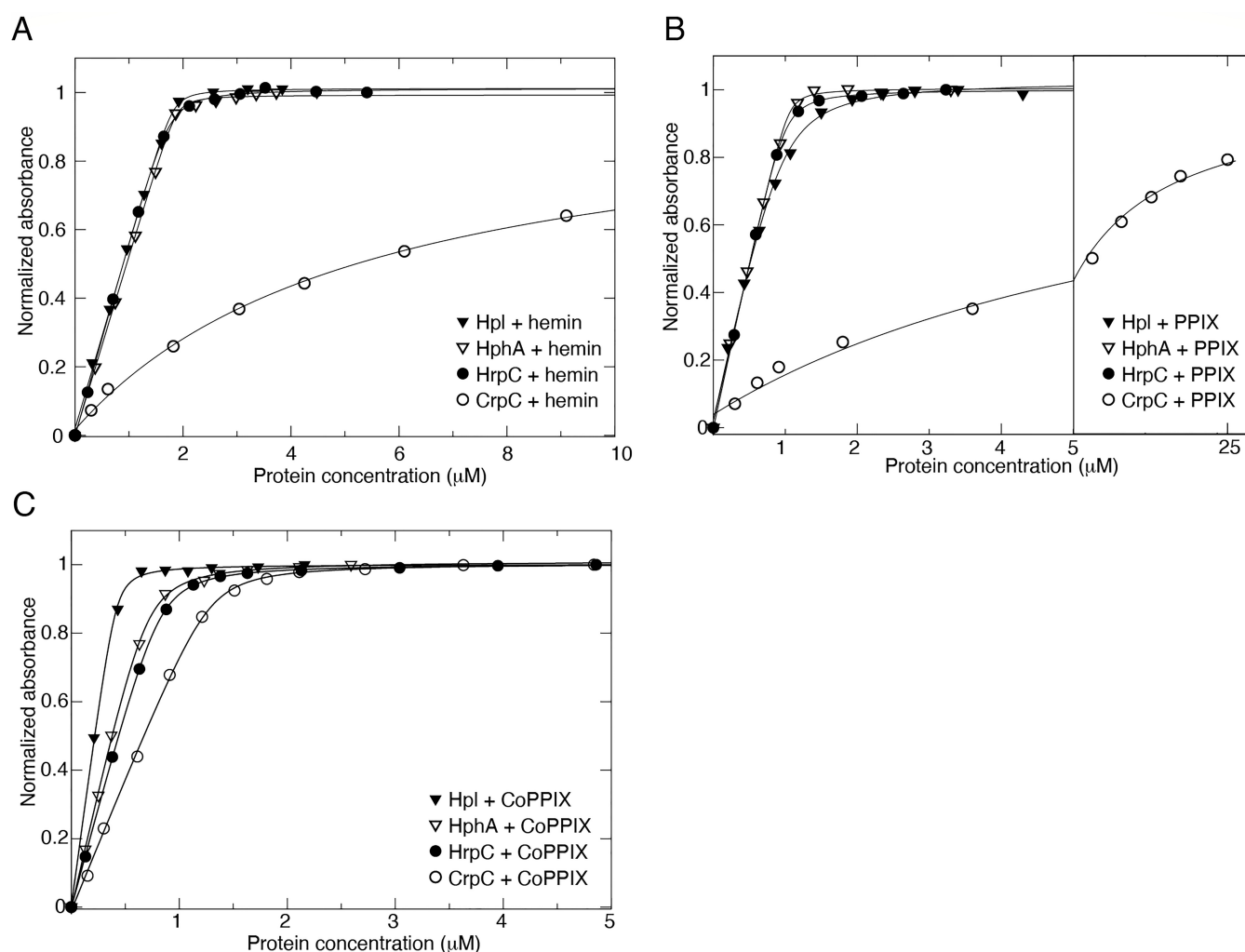


FIG 6 Binding isotherms for hemophilin homologs binding to porphyrins. Binding isotherms were generated from absorption data (symbols) by fitting to peak wavelength changes using a 1:1 binding model accounting for ligand depletion (lines). (A) Binding isotherms for Hpl, HphA, HrpC, or CrpC with hemin, generated from data in Fig. 5. (B) Binding isotherms for Hpl, HphA, HrpC, or CrpC with PPIX, generated from data in Fig. S2. (C) Binding isotherms for Hpl, HphA, HrpC, or CrpC with Co(III)PPIX, generated from data in Fig. S4.

CrpC gene is encoded adjacent to a cobalamin biosynthesis locus, we also screened for binding to Co(III)PPIX. Titration with PPIX, monitored by UV-visible spectroscopy, indicated binding by Hpl ($K_d \approx 70$ nM), HphA ($K_d < 15$ nM), and HrpC ($K_d \approx 20$ nM), whereas CrpC binding was much weaker ($K_d \approx 7$ μM), in a pattern (affinity HphA > Hpl $\approx$ HrpC $\gg$ CrpC) similar to that seen for a hemin ligand (Fig. 6B; Fig. S3). A different pattern was seen with Co(III)-PPIX. Titration experiments with Co(III)-PPIX indicated binding by Hpl ($K_d < 15$ nM), HphA ($K_d \approx 30$ nM), HrpC ($K_d \approx 24$ nM), and CrpC ($K_d \approx 24$ nM) ranked in the order Hpl > HphA $\approx$ HrpC $\approx$ CrpC (Fig. 6C; Fig. S4). To understand the potential significance of this, we looked at the interaction of BSA with Co(III)-PPIX. We found that Co(III)-PPIX and Zn(II)-PPIX bound more strongly to BSA than did PPIX or hemin, suggesting that this pattern is not specific to CrpC. We tested several other porphyrin-related molecules, including coproporphyrin III, biliverdin, and cobalamin and found no evidence for these binding to CrpC (Fig. S5). Taken together, these results suggest that Hpl, HphA, and HrpC might effectively scavenge metalated or unmetalated porphyrins from the environment. *X. nematophila* HrpC appeared to bind metalated and unmetalated porphyrin with similar affinities (10–20 nM). In contrast, CrpC binds very weakly to PPIX and hemin. Given the genomic context of CrpC, it is interesting to consider whether

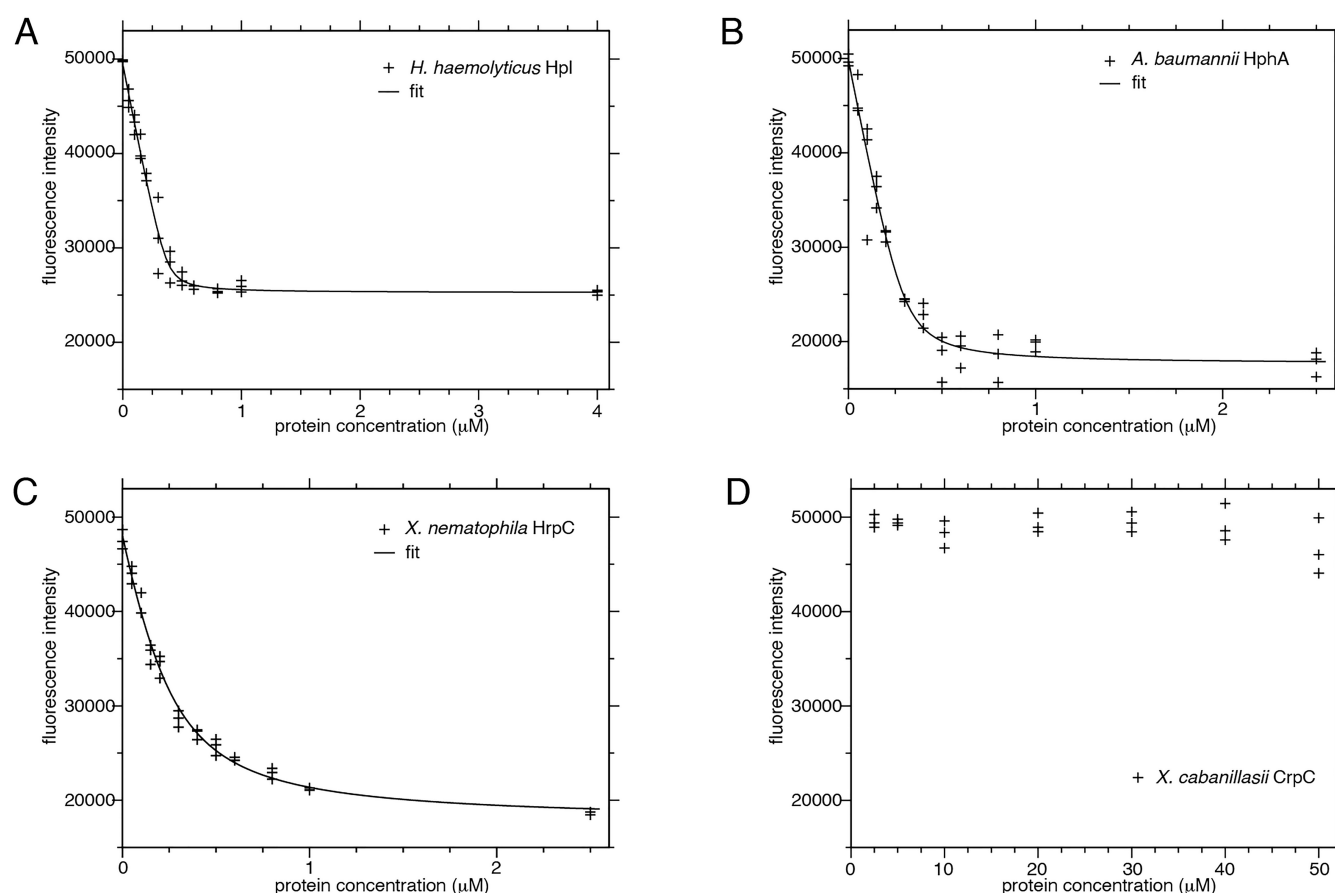


FIG 7 Titration of hemophilin homologs with Zn(II)PPIX. Binding of hemophilin homologs was monitored by changes in fluorescence intensity of the Zn(II)PPIX ligand (prepared in 20 mM Tris-HCl, pH 8.0 at 21°C) in the presence of excess BSA (15 μM). Data points (+) were fit (solid lines) with a 1:1 binding model accounting for ligand depletion. (A–C) Titrations of Zn(II)PPIX (0.5 μM) with Hpl, HphA, or HrpC. (D) Titration of CrpC into Zn(II)PPIX (1.0 μM).

this weak binding to PPIX and hemin could allow CrpC to preferentially bind cobalt porphyrins, within a biological setting containing hemin.

Hemophilin family proteins adopt a multi-domain structure characteristic of T11SS cargo

Our observations that hemophilin homologs form a distinct cluster of sequences prompted us to ask if these proteins also comprise a distinct structural class within the T11SS cargo proteins. The Hpl and HphA crystal structures (14, 19) were compared against the protein structure database [DALI server (30)] and manually compared against the protein fold classification databases CATH (31) and SCOP (32). The C-terminal β-barrel domain structure (Fig. 8A and C) that appears to be characteristic of all T11SS cargo proteins (not just hemophilins) is recognized as a distinct protein fold in CATH (superfamily 2.40.160.90) and SCOP (superfamily 3002098) (19). More limited similarity was detected between the N-terminal ligand-binding domains of Hpl or HphA (Fig. 8B and D) and the N-terminal handle domain of TbpB proteins (CATH superfamilies 2.40.128.240/250; SCOP fold 2001281). A subset of secondary structural elements in the N-terminal domains of T11SS cargo proteins, including Hpl and HphA, adopt a conserved β-sheet topology (strands 1, 8, 7, 6, and 5 in Hpl and HphA) that packs against the C-terminal β-barrel domain. The remainder of the N-terminal regions of the Hpl and HphA is variable in structure compared to each other and the other known T11SS cargo proteins and potentially give rise to the different ligand-binding properties. While the N-terminal domains of Hpl and HphA share features with the handle domain of TbpB, the insertion of one or more α-helical elements between β-strands 3 and 4, and a change

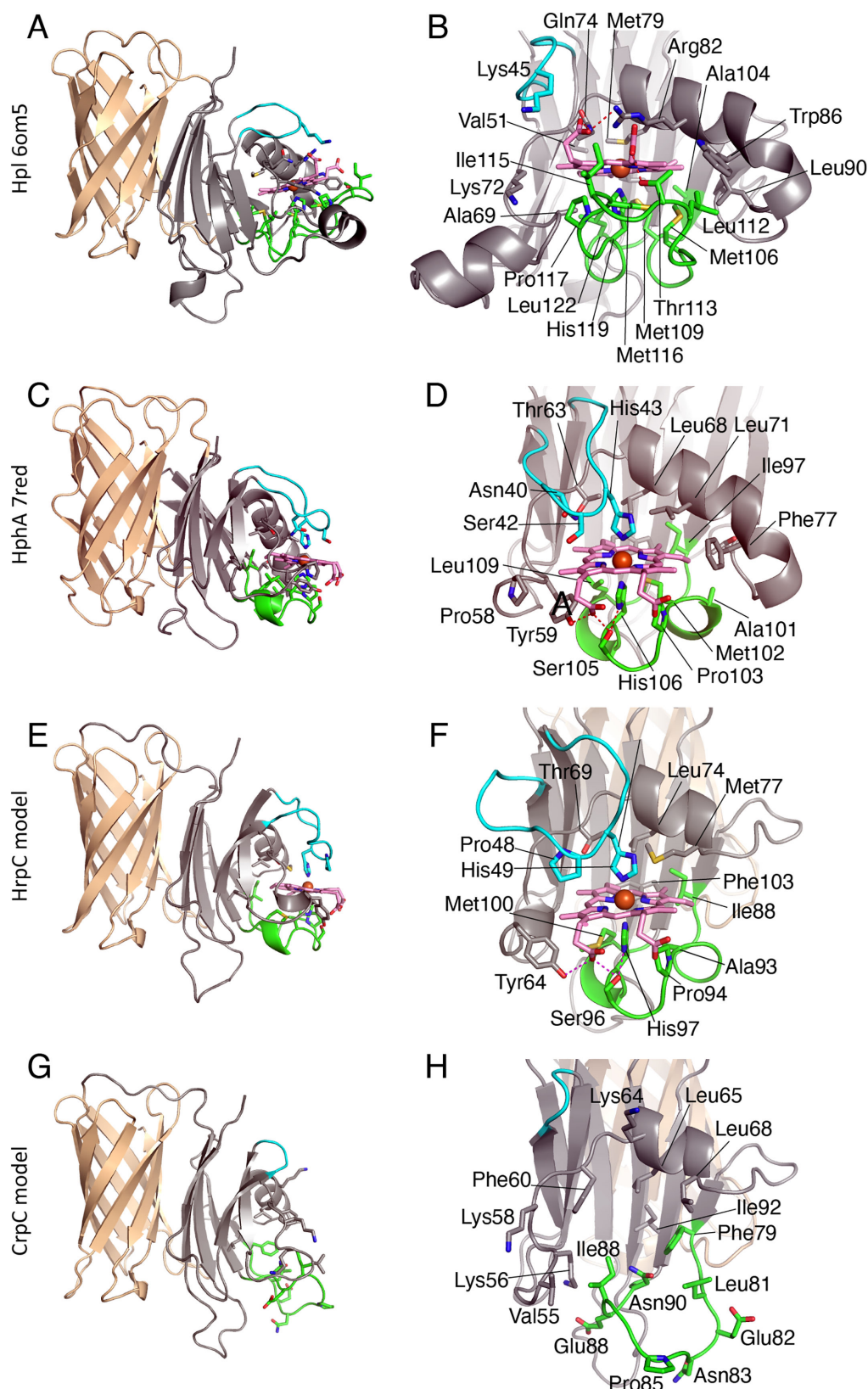


FIG 8 Structural modeling of HrpC and CrpC. (A) Domain structure of Hpl (pdb 6om5) showing the N-terminal hemophilin ligand-binding domain (gray, cyan, green) and C-terminal 8-stranded β -barrel domain (BBD; tan color). (B) Heme-binding site of Hpl showing structural elements that contribute to heme binding including the β 2- β 3 loop (cyan), β 3- β 4 α -helix that carries (Continued on next page)

FIG 8 (Continued)

Arg82 and Trp86, and the β 5- β 6 loop (green) that carries the heme-coordinating His119. The heme porphyrin and central iron atom are shown in pink and orange, respectively. (C) Domain structure HphA (pdb 7red) shown with the same coloring as A. (D) Heme-binding site of HphA colored as B. The heme iron is coordinated by two His side chains: His43 in the β 2- β 3 loop and His106 in the β 4- β 5 loop. (E and F) Structural model of *X. nematophila* HrpC based on the alignment in Fig. 10, modeled with bound heme. The heme-binding site has features similar to HphA, including bis-histidyl heme coordination. (G and H) Modeled structure of *X. cabanillasii* CrpC without a heme ligand. The precise conformation of the β 5- β 6 loop cannot be modeled by threading due to the lack of a suitable template.

in hydrogen-bond connectivity within β -strands 1–4 suggest that the hemophilin ligand binding domain belongs to a separate domain family, as a subgroup of the TbpB handle domain topology. Overall, sequence similarity between hemophilins is higher in the β -barrel domain (40% identity between Hpl and HrpC; 36% identity between Hpl and CrpC) and it is lower in the N-terminal domain (24% identity between Hpl and HrpC; 20% identity between Hpl and CrpC), consistent with variation in ligand-binding properties within the hemophilin subgroup.

We also considered the predicted structures of HrpC and CrpC as representatives of the two other subclusters of hemophilin family proteins. We made structural models using the program MODELLER (33) with the crystal structures of Hpl and HphA (14, 19) as templates (see Materials and Methods) and found that HrpC (Fig. 8E and F) and CrpC (Fig. 8G and H) are predicted to adopt a similar overall structure as Hpl and HphA.

Hemophilin homologs share a signature N-terminal ligand-binding domain

Using knowledge of the conserved hemophilin ligand-binding domain in Hpl and HphA, we wanted to ask how conserved this structure was predicted to be across the different hemophilin subclusters. Alignments were used to make structural models of representative members of each subcluster, including *X. nematophila* HrpC (Fig. 8E and F) and *X. cabanillasii* CrpC (Fig. 8G and H) for which we had measured porphyrin ligand binding, using the program MODELLER (33) with the crystal structures of Hpl and HphA (14, 19) as templates. The ligand-binding domains of Hpl and HphA have a conserved hydrophobic core that was also preserved in alignments with HrpC and CrpC and other members of each subcluster, including the Plant/Environmental subcluster (e.g., *Lysobacter enzymogenes* ALN55974 and *Stenotrophomonas maltophilia* KUJ02124) (Fig. 9). We also found conserved pair-wise contacts between side chains that are distant in the primary sequence and that provided additional confidence in modeling local structures. For example, we identified a DXNG[V/I] motif corresponding to a β -hairpin in the ligand-binding domain of *H. haemolyticus* Hpl that makes a bifurcated hydrogen bond with a Tyr side chain in the β -barrel domain (Fig. 10) and is conserved in Hpl-like, HrpC-like, and CrpC-like sequences but is absent in the HphA-like cluster (Fig. 9). With respect to this motif, the Plant/Environmental subcluster comprises two subsets, one containing the DXNG motif (represented by ALN55974 from *L. enzymogenes*) and one lacking the DXNG motif and with a divergent sequence/structure in β 4- β 5 hairpin (represented by KUJ02124 from *S. maltophilia*; Fig. 9). Several groups of sequences in the Hpl-like subcluster could not be aligned or modeled against the hemophilin domain due to greater sequence (and, therefore, presumably structural) diversity within this subcluster. With the exception of this group, structural modeling suggests that hemophilin homologs (members of the HrpC, CrpC, HphA, Plant/Environmental subclusters, and a subset of the Hpl subcluster) have an N-terminal domain with the same topology as the ligand-binding domains of *H. haemolyticus* Hpl and *A. baumannii* HphA and that this hemophilin ligand-binding domain is the signature feature of an extensive family of hemophilin homologs.

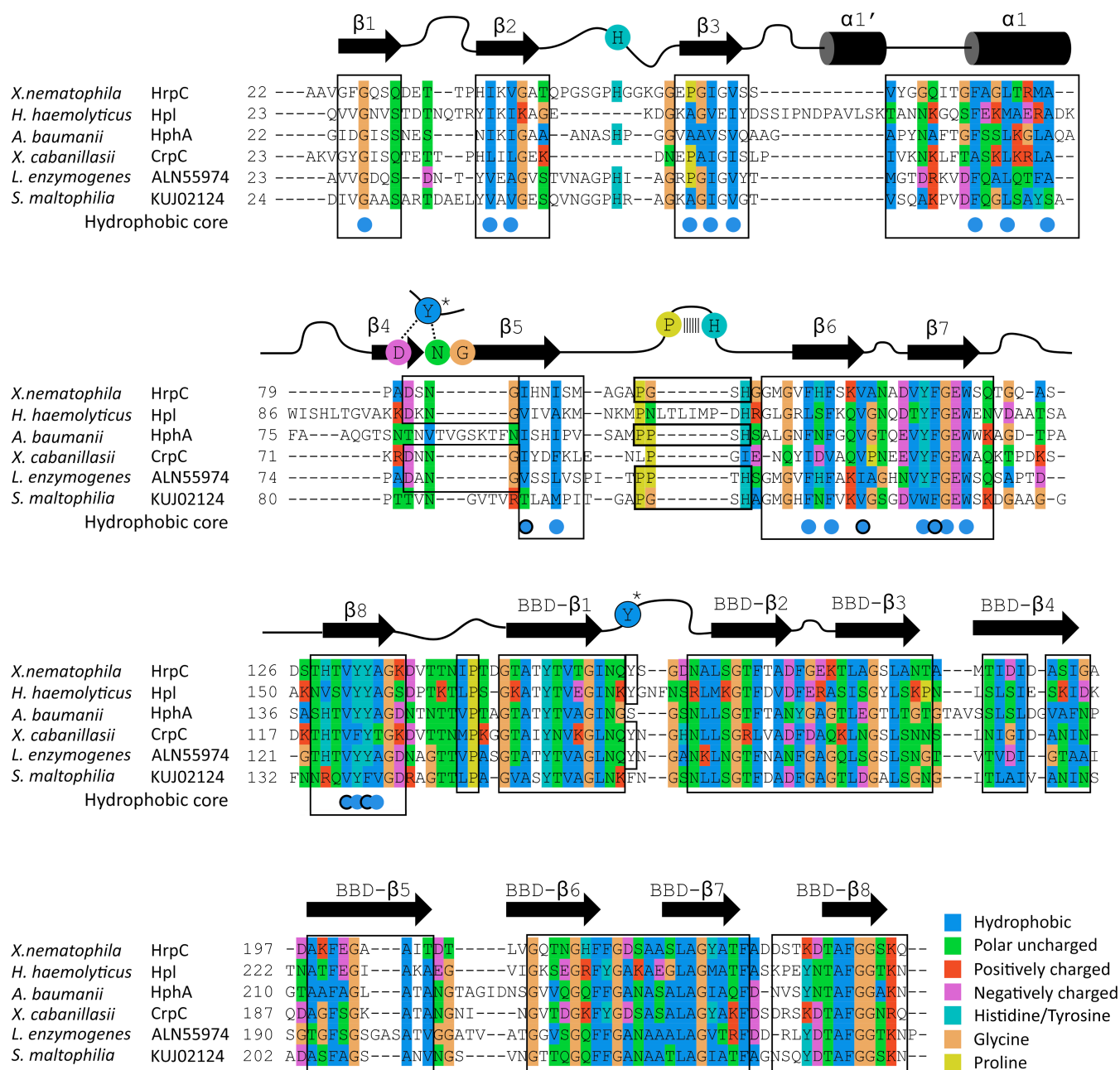


FIG 9 Aligned sequences of representative members from hemophilin subclusters defined by sequence similarity network analysis. The alignment is a composite of information including structure-based alignment of *H. haemolyticus* Hpl and *A. baumannii* HphA sequences and sequence-based alignments using the full membership of each hemophilin subcluster output from network analysis. Only representative members of the HrpC-like, CrpC-like, and Plant/Environmental clusters used in modeling (see Fig. 8; Fig. S6) are shown in the figure. The approximate position of secondary structure elements is displayed in cartoon, with residues of the DXNG and PX[S/T]H motifs (see Fig. 10) shown in colored circles. Boxes indicate regions where query sequences could be modeled on one or both structure templates. Residues of the conserved hydrophobic core of the N-terminal hemophilin domain, or interdomain contact, are marked with blue circles without or with a black outline, respectively.

At least two distinct heme-binding modes occur among hemophilin homologs

Having identified the conserved fold common to hemophilins, we used models to investigate possible structural reasons for common and distinct ligand-binding properties of hemophilin homologs. The crystal structures of *H. haemolyticus* Hpl and

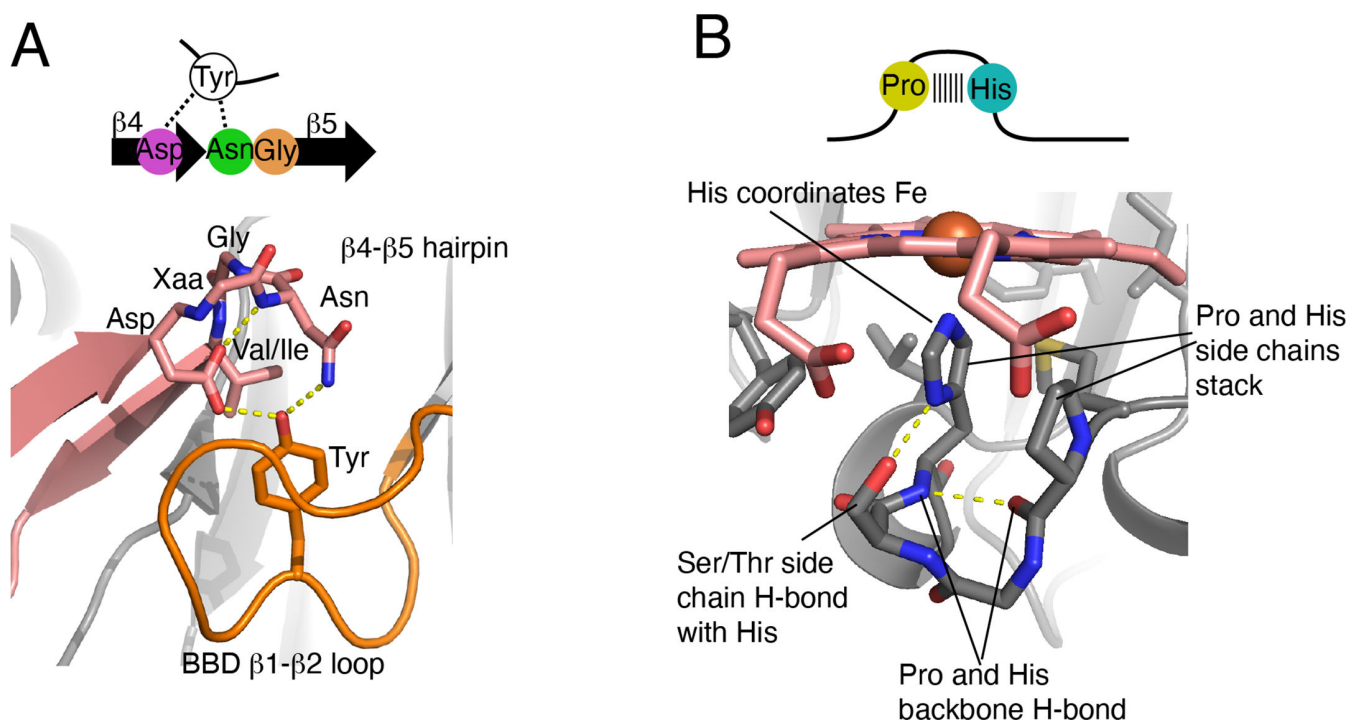


FIG 10 Conserved sequence-structure motifs. (A) DXNG[V/I] motif structure in the $\beta 4$ - $\beta 5$ hairpin. (B) PX[S/T]H motif incorporating the heme-ligating His in the $\beta 5$ - $\beta 6$ loop.

A. baumannii HphA, both of which bind heme, show that residues in loops $\beta 2$ - $\beta 3$, $\beta 3$ - $\beta 4$ (including the $\beta 3$ - $\beta 4$ α -helix), and $\beta 5$ - $\beta 6$ contribute the majority of heme-binding contacts and that the mode of heme binding in Hpl and HphA differs in the relative contributions of these loops and in the heme-iron coordination structure (Fig. 8B and D). Of particular interest, heme in Hpl is coordinated through a single His side chain in the $\beta 5$ - $\beta 6$ loop, whereas heme in HphA is coordinated by two His side chains, one from the $\beta 5$ - $\beta 6$ loop and a second donated by the $\beta 2$ - $\beta 3$ loop.

The bis-histidyl heme coordination observed for HphA appears to be ubiquitous in the HphA subcluster, based on the conservation of the amino acid sequence of this motif (data not shown). A subset of sequences within the Hpl subcluster had >65% identity to *H. haemolyticus* Hpl, and these are predicted to all coordinate heme through a single histidine side chain. The heme coordination pattern for the remaining members of the Hpl subcluster is unclear due to the high level of sequence/structure diversity in this group. Molecular models show that HrpC subcluster and Plant/Environmental subcluster proteins contain His side chains in both the $\beta 2$ - $\beta 3$ loop and $\beta 5$ - $\beta 6$ loops, and modeling in the presence of a heme ligand indicates that the position of these His residues is compatible with bis-histidyl heme coordination similar to HphA (Fig. 8F; Fig. S6). Other structural similarities among HphA, HrpC, and Plant/Environmental subclusters, but not for the Hpl subcluster, exist in the $\beta 2$ - $\beta 3$ loop, including a Gly-, Ser-, Pro-rich sequence, and similar loop length (Fig. 9). The mechanism of heme binding and release by *A. baumannii* HphA is proposed to involve unfolding of the $\beta 2$ - $\beta 3$ loop (14), and if so, this mechanism is likely to be conserved in hemophilin homologs with this motif. The second heme-coordinating His of HphA occurs in the $\beta 5$ - $\beta 6$ loop, which contains a conserved PX[S/T]H sequence motif.

Like the $\beta 2$ - $\beta 3$ loop, the PX[S/T]H motif of the $\beta 5$ - $\beta 6$ loop is conserved across the HphA, HrpC, and Plant/Environmental clusters (Fig. 9 and 10B). In *A. baumannii* HphA, the His side chain of this motif is packed against the pyrrolidine ring of Pro and makes a H-bond with the Ser side chain (substitution of Thr for Ser is expected to preserve this interaction). The carbonyl oxygen of Pro also accepts a H-bond from the backbone amide

of His. These interactions are expected to stabilize the orientation of the heme-ligating His, and conservation of this motif in HrpC and Plant/Environmental cluster hemophilin homologs further implies that the mode of heme binding in HphA, HrpC, and Plant/Environmental proteins is similar. The PX[S/T]H motif is absent in Hpl, which has a substantially longer $\beta 5$ - $\beta 6$ loop (20 residues in Hpl vs 14 residues in HphA) that more completely covers the face of the porphyrin.

The predicted N-terminal structure of CrpC is consistent with loss of heme and porphyrin binding

In addition to iron-ligating side chains, other amino acids contribute to ligand binding through non-covalent interactions with apolar and polar sites on the porphyrin. Models of *X. nematophila* HrpC show multiple hydrophobic contacts with the porphyrin ring contributed by side chains in the $\beta 3$ - $\beta 4$ α -helix, $\beta 5$ - $\beta 6$ loop, and underlying β sheet, and a hydrogen bond network around the charged porphyrin 17-propionate comprising Tyr64, Ser96, and His97 side chains—features similar to the HphA heme site (Fig. 8, compare D, F). Differences include a shorter $\beta 3$ - $\beta 4$ α -helix in HrpC than either HphA or Hpl, and a $\beta 4$ - $\beta 5$ hairpin in HrpC that is more similar to Hpl than to HphA. The pyrrole rings B and C appear to be more exposed to solvent in HrpC than in either Hpl or HphA, which may partly explain a relatively lower affinity for the metalated porphyrins, hemein and Zn(II)PPIX, because hydration is expected to increase the rate of scission of the Fe-metal bond (although the lack of effect on Co(III)PPIX binding is not explained).

Our porphyrin-binding assays revealed that *X. cabanillasii* CrpC does not bind heme. Consistent with this observation, while *X. cabanillasii* CrpC is predicted to adopt the hemophilin ligand-binding domain fold, it lacks His or other residues that typically coordinate heme-iron, in either the $\beta 2$ - $\beta 3$ or $\beta 5$ - $\beta 6$ loops (Fig. 8G and H; Fig. S6). The $\beta 2$ -3 loop is reduced to a three-residue hairpin and is too small to contribute to a porphyrin pocket of the kind seen in HphA and HrpC. Neither does CrpC look like the Hpl heme-binding site, which employs a much longer $\beta 2$ - $\beta 3$ α -helix and longer $\beta 5$ - $\beta 6$ loop. Nevertheless, the distribution of apolar and polar side chains in Hpl/HphA that contact the porphyrin skeleton and ionizable propionates, respectively, is largely conserved in CrpC although we cannot model the precise conformation of the $\beta 5$ -6 loop in CrpC due to lack of a suitable template. Thus, CrpC could potentially accommodate a porphyrin-like ligand in a more solvent exposed, and therefore presumably lower affinity, binding site. This *X. cabanillasii* distinct non-histidine ligand-binding domain is shared among the other CrpC-like subcluster homologs, except for a small number of sequences that have a large deletion in the hemophilin domain that removes the porphyrin-binding site altogether.

DISCUSSION

A cornerstone of microbial existence is the extracellular deployment of metal-binding molecules that facilitate the competitive uptake of metals that are essential for cellular physiology. Here, we investigated the structural features, secretion, and ligand binding of a recently described family of such metal-acquisition molecules, the hemophilins. These proteins occur in five distinct sequence similarity subclusters encoded by Proteobacteria from human, animal, plant, built, and free-living environments. We found that the entire hemophilin family shares an N-terminal domain with a common α/β scaffold and a C-terminal β -barrel domain. These domains are present in the two solved hemophilin structures, *H. haemolyticus* Hpl and *A. baumannii* HphA (14, 19), and the C-terminal barrel is a previously described characteristic of T11SS cargo proteins (8, 10). In contrast, the N-terminal ligand-binding domain is specific to hemophilin homologs, making these a distinct structural subgroup of T11SS cargo proteins. We established that, like the hemophilins *X. nematophila* HrpC and *A. baumannii* HphA (8, 34, 35), two other hemophilin homologs, *H. haemolyticus* Hpl and *X. cabanillasii* CrpC, rely on a T11SS secretor to reach the extracellular milieu. This establishes T11SS-dependence across all four hemophilin sequence subclusters tested and indicates T11SS-dependent secretion

is a hallmark of the entire family. Although the hemophilin ligand-binding domain structure is specific to that group, a subset of β -strands that comprise a β -sheet that packs against the C-terminal barrel is also recognizable in TbpB, HpuA, and other lipidated T11SS cargo. Overall, the results presented here support the concept that the β -sheet- β -barrel architecture is an ancestral scaffold for T11SS-dependent secretion, onto which new effector variants are emerging through the evolution of the N-terminal domain.

The results presented here provide a comparative framework in which to consider the overall evolution and relatedness of T11SS cargo protein structures. These can be described as comprising three domains: an N-terminal variable effector domain, a mid-protein β -sheet domain, and a C-terminal β -barrel domain. The C-terminal 8-stranded, hydrophobic core β -barrel domain (19) is a unique and defining feature of T11SS cargo proteins that, as we have established here, contributes to the recognition of a specific cargo partner by the cognate T11SS secretor. The mid-protein β -sheet domain is also universally present among known T11SS cargo. It is formed in large part by amino acids in the central portion of the protein sequence and stacks against the β -barrel forming a physical scaffold on which the variable structural elements of the N-terminal effector domain are built. Variation in the structural elements of the N-terminal domain allows cargo such as the hemophilins to bind large organic ligands such as porphyrins, whereas other elaborations of the same underlying scaffold allow binding to protein ligands—such as in the case of HpuA, which binds specifically to hemoglobin (19).

Despite the overall conservation of the gross structural motifs of T11SS-dependent cargo, there must be inherent differences among these domains since different cargo proteins show specificity for different T11SS secretors (7). Here, we capitalized on the fact that each hemophilin homolog is associated with a cognate T11SS secretor, enabling different combinations of related cargo-secretor pairs to be assessed for secretion activity. Using HrpB from *X. nematophila* as our representative type eleven secretor, we found that, despite sequence, structural, and functional similarity among hemophilin homologs and among their cognate T11SS secretors, HrpB_{X.nem} could not substitute for HrpB_{H.haem}, HsmA, or CrpB in an *E. coli* secretion assay, indicating co-adaptation and co-diversification of cargo-secretor pairs. These results suggest that T11SS specificity is not shared between cargo, even when comparing closely related homologs, although future studies employing T11SS from diverse species would be required to determine the generality of T11SS selectivity. When expressed in a common *E. coli* background under identical conditions, the T11SS secretors we studied ranged by an order of magnitude in their impact on the levels of extracellular cargo and were more effective at secreting the cargo (wild type or chimeric) containing the C-terminal β -barrel domain of their cognate (vs non-cognate) hemophilin. The latter finding supports the idea that the β -barrel domain contributes more to T11SS cargo specificity than does the N-terminal domain (7, 17). However, for both T11SS tested, chimeric proteins were not secreted as well as the cognate wild-type cargo proteins, possibly due to structural differences that slow transposition or increase degradation. Analyzing a series of hemophilin homolog chimeras with different junction points could yield insights into the structural motifs or inter-domain contacts that are important for optimal secretion by, and specificity for, T11SS. Regardless, our work has highlighted that T11SS specificity exists even among closely related cargo proteins due in part to features within the cargo C-terminal β -barrel domain. This raises the possibility that novel T11SS-dependent cargo could be engineered for specific secretion by fusing novel N-terminal functional domains to the C-terminal domain from a characterized cargo protein with the desired level of secretion activity. Such a strategy could be useful for T11SS-driven surface presentation of immunogenic antigens that has been proposed as a potential vaccination strategy (36). An important next step toward achieving this goal will be to identify the individual sequence and structural motifs responsible for the specificity to enable automatic annotation of T11SS cognate pairs and informed engineering of novel pairings.

A distinctive feature of hemophilins revealed by our study is an N-terminal, β 1- β 7 region, hemophilin ligand-binding effector domain that is diversifying within the family. The hemophilin ligand-binding domain shares a parent fold (SCOP fold 2001281) with the TbpB handle domain (SCOP family 4006246; TbpB handle domain-like) and with HpuA (SCOP family 4005058, hemoglobin receptor HphA), highlighting the relatedness of hemophilins with other metal-related T11SS cargo. However, relative to these other cargo proteins, hemophilins display differences in connectivity in the β 1- β 3 strands and inclusion of one or more α -helical elements in the β 3- β 4 loop. The hemophilin ligand-binding domain contains the most highly conserved sequence motif among hemophilins: a YFGEW pentapeptide that maps to β 7. This motif contributes side chains to interfacial contacts between the β -sheet and β -barrel domains, as well as to other residues within the conserved N-terminal hemophilin ligand-binding domain. Also within this region is a characteristic, active site PX[S/T]H motif (β 5- β 6 loop) that features a heme-ligating His side stabilized by π -stacking and H-bonding interactions and that likely is responsible for heme binding. A search of the protein database using RASMOT-3D (37) identified only one other heme-binding protein (nine-heme cytochrome *c*, pdb 19hc) with a similar 3D arrangement of Pro, Ser, and heme-coordinating His side chains, but different primary sequence arrangement. Thus, the PX[S/T]H appears to be specific, defining feature of the hemophilin ligand-binding domain, though the modeled CrpC structure lacks this PX[S/T]H motif, suggesting an evolutionary trajectory for loss of heme binding in CrpC. Overall, our structural analysis results imply that β -strand topology and the hydrophobic core of the hemophilin ligand-binding domain comprise an ancestral scaffold which has been elaborated on in evolution to bind a diverse range of ligands for T11SS cargo proteins: a variety of protein targets in the cases of HpuA, TbpB, and fHbp, heme in the case of Hpl, HphA, and HrpC, and unknown ligands in the case of CrpC and NilC.

To gain further insights into the diversification of hemophilins, we categorized them in two ways: using network analysis, which revealed five subclusters, and by considering active site residues and topology. The latter suggested three general active site classes of hemophilins which cross network subcluster boundaries: the HphA/HrpC type including representatives from the Plant/Environmental network subcluster, with a bis-histidyl heme coordination site; the Hpl type with a single histidine heme coordination site; and the CrpC type which lacked any histidines in the predicted active site. The bis-histidyl coordination HphA/HrpC type might be considered a canonical heme-binding structure with heme coordinated between two histidines: one in a low complexity, Gly, Pro, and polar residue-rich β 2- β 3 loop, and a second in the PX[S/T]H motif of the β 5- β 6 loop noted above. The conserved low complexity of the β 2- β 3 loop suggests that conformational heterogeneity may be a general feature of this class; in heme-bound *A. baumannii* HphA the loop and heme-coordinating His cover one face of the porphyrin, whereas the same loop points out away from the binding site in the apo protein (14). In contrast to the HphA/HrpC class, the Hpl-like network subcluster appears to be more diverse in structure and arrangement of heme ligands. One common feature within the Hpl subcluster is a D[R/K/S]NGV motif in the β 4- β 5 hairpin that is predicted to make conserved interactions at the interface between the β -sheet and the β -barrel domains. A similar motif appears in HrpC, CrpC, and a subset of the Plant/Environmental subcluster. *H. haemolyticus* Hpl itself has the single histidyl coordination-binding site in the conserved β 5- β 6 loop PX[S/T]H motif. Although most sequences within the Hpl subcluster carry this His, one group (sequence OFR67839 is one example) lacks this residue and others potentially have two His ligands.

These distinguishing active site characteristics among the hemophilin homologs contextualized our findings that purified hemophilin homologs displayed variable affinity for hemin (Fig. 5; Table 1). For example, the heme-associated α -helix in both HphA and Hpl is longer than predicted in HrpC and Plant/Environmental cluster proteins, possibly contributing to higher heme affinity in HphA/Hpl by protecting a larger surface area of the porphyrin from solvent. At the other end of the spectrum, CrpC, a member

of the Cobalt/molybdenum-associated subcluster of hemophilin homologs, did not bind heme, consistent with the fact that the predicted active site had neither the heme-coordinating His residues noted for the other hemophilin classes nor any other typical heme ligands such as Cys, Met, or Tyr. Other potential metal ligands include Lys, Asn, Gln, and Glu are known only as a second heme ligands, not as primary ligands (38–42). Thus, the absence of a strong primary heme-iron ligand provides a structural explanation for the failure of CrpC to bind to heme or Zn(II)PPIX. In this context, the binding of Co(III)PPIX to CrpC and the low-affinity heme-binding protein, BSA, with comparable affinity to other hemophilins was unexpected. This result is particularly intriguing given the genomic context of the genes encoding CrpC and its T11SS partner, CrpB, which are found adjacent to an anaerobic B12/cobalamin biosynthesis locus in at least three strains of *Xenorhabdus* (Fig. 2B; Supplemental file 2). This may hint at a role for the CrpC/CrpB cargo/secretion pair in binding certain intermediates of cobalamin biosynthesis. Based on these data, we suggest that CrpC has evolved to lose a heme-binding function while maintaining the ability to bind an alternative porphyrin-related ligand, such as an intermediate in cobalamin biosynthesis or metabolism. Future studies could attempt to isolate and identify a primary ligand *in vivo* to better understand the role of CrpC and other divergent hemophilin homologs, such as those which clustered near CrpC within the sequence similarity network (Fig. 2A).

Variable affinity for heme may reflect differences in the ecological roles of the hemophilin homologs. Hpl and HphA, which have established roles in competing for heme under stringent conditions, had the highest heme affinity. *H. haemolyticus* Hpl is a heme chelating protein for nutritional heme uptake and can prevent the related organism, *H. influenzae*, from accessing environmental heme (19). This role for *H. haemolyticus* Hpl as a tool for competing against other organisms by sequestering a limiting nutrient (a form of nutritional immunity) would naturally favor the evolution of high heme affinity. In a similar vein, *A. baumannii* HphA competes with mammalian host nutritional immunity for heme. In contrast to these two hemophilins, HrpC, encoded by *X. nematophila*, displayed lower heme-binding affinity. This may reflect the distinctive animal host niches occupied by *X. nematophila* relative to *H. haemolyticus* and *A. baumannii*. As a pathogen, *Xenorhabdus* can infect diverse insects, which are notoriously heme poor (43), and as a mutualist, it colonizes the intestinal tissues of its nematode host, which is a heme auxotroph (44, 45). The insect *Drosophila melanogaster* transferrin-1 binds iron with lower affinity than mammalian transferrin and is susceptible to low pH conditions (46). Therefore, to access iron in an insect environment, *X. nematophila* HrpC may not need high affinity to overcome host chelation. In turn, lower affinity may offer a selective advantage by enabling resource sharing with its mutualistic nematode host, *S. carpocapsae*, with the resulting improved fitness benefiting both mutualistic partners. Dissociation of heme from HrpC would be essential for heme to be transferred to nematode iron chelating proteins. Entomopathogenic nematodes can live for months at a time without feeding while in their free-living infective stage. During this time, they have a sealed intestine and no access to exogenous nutrients and they may rely on their *Xenorhabdus* symbionts to provide heme chaperoned by hemophilin (47). While this study did not test the heme-binding affinity of hemophilin homologs from the Plant/Environmental cluster, these homologs encode His residues in the β 2- β 3 loop and the β 5- β 6 loop like HphA and HrpC (Fig. S6). Future studies could investigate how these hemophilin homologs facilitate plant and soil-associated bacterial lifestyles while also examining their respective heme-binding affinities. Our work has revealed that not all hemophilins have heme-binding activity, which opens the possibility that this family is diversifying and potentially gaining new ligand-binding activities.

Overall, our findings suggest that bacteria can encode multiple members of the hemophilin family, both bona-fide hemophilins, such as HrpC, and paralogs like CrpC, that might have evolved new ligand-binding activities by divesting themselves of their affinity for heme. An exciting avenue for the discovery of new metal-, or other

ligand-binding proteins will be to examine the activities of additional members of the hemophilin family of proteins.

MATERIALS AND METHODS

Sequence similarity network analysis

Protein sequence similarity networks were generated using the Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST) (23). As an input, we used the previously reported database of soluble TbpBBD cargo-encoding genes which co-occurred with T11SS encoding genes (8). EFI-EST performs an all-by-all BLAST of query sequences to assess relatedness and then generates a network where each node represents a protein sequence, and the color of the edges indicates relatedness between nodes. A minimum alignment score of 35 was chosen to reduce total network edges enough to visualize protein subclusters. To simplify visualization, proteins sharing $\geq 80\%$ identity were compressed into representative nodes. Networks were visualized and interpreted using Cytoscape v3.7.1 (48) and Gephi v0.9.5 (49). For the complete network of soluble TbpBBD domain proteins, nodes were organized with the Fruchterman-Reingold force-directed algorithm (50). For the network containing only hemophilin family proteins, nodes were organized with the ForceAtlas2 algorithm for continuous force-directed arrangement (51).

Bacterial culture conditions

All strains, plasmids, and primers utilized in this study are described in Supplemental File 3. All cultures were grown in glucose minimal media (34), LB stored in the dark to prevent the formation of oxidative radicals (henceforth dark LB), or glucose minimal media supplemented with 1% dark LB. Plate-based cultures were grown on either LB supplemented with pyruvate to prevent the formation of reactive oxygen radicals (henceforth, LBP or glucose minimal plates) (34). For plasmid-based expression, chemically competent *E. coli* strain BL21-DE3 (C43) were chosen for ease of transformation and their ability to tolerate expression of membrane proteins (52, 53). Strains of *E. coli* were grown at 37°C. Where appropriate, media were supplemented with the following antibiotics and concentrations (unless otherwise stated): ampicillin (150 $\mu\text{g}/\text{mL}$), chloramphenicol (15 $\mu\text{g}/\text{mL}$), or kanamycin (50 $\mu\text{g}/\text{mL}$). Protein expression was induced at the mid-log point of bacterial growth via the addition of isopropyl β -D-thiogalactopyranoside (henceforth, IPTG) at a concentration of 0.5 mM.

Construction of T11SS and T11SS-dependent cargo expression plasmids

Expression plasmids for HrpC alone (HGB2531) and HrpBC co-expression (HGB2530) were previously generated and reported (8). FLAG-*hrpB* was amplified from pETDuet-1/*hrpBC_{Xnem}* using primers 1–2. *hpl*-FLAG and its adjacently encoded T11SS neighbor, FLAG-*hrpB_{H.haemr}*, were amplified from the purified genome of *H. haemolyticus* BW1 using primers 3–6. *crpC*-FLAG and its adjacently encoded T11SS neighbor, FLAG-*crpB*, were amplified from the purified genome of *X. cabanillasii* (HGB2490) using primers 7–10. *hphA* (ACJ40780.1) and *hsmA* (ACJ40781.1) were generated via gene synthesis by Genscript. To make cargo-only expression plasmids, pETDuet-1, *hpl*-FLAG, *crpC*-FLAG, and *hphA*-FLAG were digested with NcoI and NotI. Each cargo protein was independently ligated into MCS1 of pETDuet-1 via T4 DNA ligase, resulting in pETDuet-1/*hpl* (HGB2526), pETDuet-1/*crpC* (HGB2525), and pETDuet-1/*hphA* (HGB2532). Integration of each T11SS-dependent cargo was confirmed via digestion with NcoI and NotI as well as Sanger sequencing using primers 11–12 at the University of Tennessee (UT) Genomics Core.

To make T11SS/cargo co-expression plasmids, each of the above cargo-only expression plasmids was digested with KpnI and NdeI, alongside the PCR products for FLAG-*hrpB_{H.haemr}*, FLAG-*crpB*, and FLAG-*hsmA*. Each T11SS protein was then

independently ligated into MCS2 of the plasmid containing its cognate cargo via T4 DNA ligase, resulting in pETDuet-1/*hpl/hrpB_{H.haem}* (HGB2523), pETDuet-1/*crpC/crpB* (HGB2524), and pETDuet-1/*hphA/hsmA* (HGB2533). Additionally, the PCR product for FLAG-*hrpB* from *X. nematophila* was digested with KpnI and NdeI and ligated into MCS2 of all the cargo-only expression plasmids, resulting in pETDuet-1/*hpl/hrpB_{X.nem}* (HGB2529), pETDuet-1/*crpC/hrpB_{X.nem}* (HGB2528), and pETDuet-1/*hphA/hrpB_{X.nem}* (HGB2527). Integration of each T11SS protein was confirmed via digestion with KpnI and NdeI as well as Sanger sequencing using primer 13 at the University of Tennessee (UT) Genomics Core.

To construct chimeric hemophilin homologs, *hrpC* and *hpl* were split into two domains based on multiple sequence alignment and the NCBI conserved domain database. *hrpC* was split between nucleotide position 402 and 403, while *hpl* was split between nucleotide position 474 and 475. Primers 14–15 were used to amplify the hemophilin handle domain from pETDuet-1/*hrpC/hrpB_{X.nem}* (HGB2530). Primers 16–17 were used to amplify *hrpB* and the hemophilin β -barrel domain from pETDuet-1/*hrpB_{X.nem}/hpl* (HGB2529). These two products were assembled into pETDuet-1/Chimeric hemophilin(*hrpC-hpl*)/*hrpB_{X.nem}* (HGB2595). Primers 18–19 were used to amplify the hemophilin handle domain from pETDuet-1/*hrpB_{X.nem}/hpl* (HGB2529). Primers 20–21 were used to amplify *hrpB_{X.nem}* and the hemophilin β -barrel domain from pETDuet-1/*hrpC/hrpB_{X.nem}* (HGB2530). These two products were assembled into pETDuet-1/Chimeric hemophilin(*hpl-hrpC*)/*hrpB_{X.nem}* (HGB2596). Primers 14–15 were used to amplify the hemophilin handle domain from pETDuet-1/*hrpC/hrpB_{X.nem}* (HGB2530). Primers 16–17 were used to amplify *hrpB_{H.haem}* and the hemophilin β -barrel domain from pETDuet-1/*hpl/hrpB_{H.haem}* (HGB2523). These two products were assembled into pETDuet-1/Chimeric hemophilin (*hrpC-hpl*)/*hrpB_{H.haem}* (HGB2597). Finally, pETDuet-1/Chimeric hemophilin (*hpl-hrpC*)/*hrpB_{X.nem}* (HGB2596) and pETDuet-1/*hpl/hrpB_{H.haem}* (HGB2523) were digested with NotI and NcoI, liberating the hemophilin homolog from each vector. The chimeric hemophilin from HGB2596 was isolated via gel electrophoresis and then ligated into MCS1 of the vector isolated from HGB2523, resulting in pETDuet-1/Chimeric hemophilin (*hpl-hrpC*)/*hrpB_{H.haem}* (HGB2598). Integration of each T11SS protein and chimeric cargo was confirmed via Sanger sequencing using primers 12–13 at the University of Tennessee (UT) Genomics Core.

Protein expression and immunoblotting

E. coli strains used for expression experiments were taken fresh from storage at -80°C for each experiment. Strains were cultured on glucose minimal media plates + ampicillin overnight. For each biological replicate, 10 colonies were pooled and inoculated into 5 mL of fresh minimal media glucose + ampicillin broth and incubated rotating overnight. Each replicate of each strain was rinsed 2 $\times$ in PBS and normalized to an OD₆₀₀ of 0.05 in 60 mL of glucose minimal media + 1% LB + ampicillin. These were grown shaking at 225 rpm until they reached mid log growth (OD₆₀₀ $\approx$ 1), typically between 5 and 8.5 h. Upon reaching mid-log growth, 25 mL of each culture was removed and used as an uninduced T0 control. The remaining 35 mL was supplemented with IPTG to a concentration of 0.5 mM. One-milliliter samples of supernatant were taken at 1 and 2.5 h post induction. Supernatant samples were clarified via centrifugation and filter sterilized. At 2.5 h post induction, the remaining cultures were concentrated via centrifugation, rinsed 2 $\times$ in PBS, and lysed via sonication (30 s at $\sim$ 500-rms volts). Supernatant samples and cellular lysate samples were supplemented with PMSF (1.7 $\mu\text{g/mL}$), Leupeptin (4.75 $\mu\text{g/mL}$), and Pepstatin A (0.69 $\mu\text{g/mL}$) to inhibit proteinase activity. The no plasmid control was performed identically except without the presence of ampicillin in the media.

The protein concentration of cellular lysates was normalized via the Pierce 660 nm Protein Assay (REF22660). For supernatant samples, 600 μL of each filtered sample was precipitated via 10% Trichloroacetic acid precipitation as previously described (8, 54). Samples were boiled for 10–25 min prior to performing SDS-PAGE to ensure complete

unfolding in the protein sample buffer. SDS-PAGE was performed in duplicate using 10% polyacrylamide gels. The first gel was used to perform Coomassie staining for total protein content, while the second gel was transferred to a PVDF membrane for Western immunoblotting. Immunoblots were incubated in 50% Ly-cor blocking buffer:50% Tris-buffered saline (TBS) for 1 h to block. Immunoblots were then incubated in 50% Ly-cor blocking buffer:50% TBS supplemented 0.1% Tween20 and 1:5,000 rat α -FLAG antibody for 1 h. Subsequently, the blots were incubated in 50% Ly-cor blocking buffer:50% TBS supplemented 0.1% Tween20 and 1:5,000 goat α -rat antibody bound to a 680CW fluorophore for 1 h. Finally, immunoblots were visualized using a Li-cor odyssey imaging the 700 nm wavelength. The intensity of supernatant samples was normalized to a clearly visible, non-target protein band in the Coomassie stain to control for protein concentration. Efficacy of secretion was measured as the fold change of cargo protein present in the supernatant when co-expressed with a T11SS protein relative to cargo protein present in the supernatant when expressed alone. Fold changes were not normally distributed initially, so they were $\log_{10}$ transformed prior to analysis. Cognate vs non-cognate protein secretion data were analyzed via a one-way ANOVA and a Tukey's HSD test (55). Chimeric protein secretion data were analyzed via a Kruskal-Wallis ANOVA with an uncorrected Dunn's comparison.

Mass spectrometry

Gel bands for proteins: Hpl (32 kD), Hpl (26 kD), CrpC (27 kD), CrpC (19 kD), and HrpC (30 kD) were processed for HPLC-mass spectrometry based on reference (56). Gel bands were rinsed twice with HPLC MS-grade water. To each gel piece, 100 μ L of 50:50 100 mM ammonium bicarbonate (ABC)/acetonitrile (ACN) was added and then gel pieces were pulverized using a pestle. Pieces were incubated for 30 min with occasional vortexing to remove the staining. Acetonitrile (500 μ L) was added and incubated until gel pieces shrank and became white and then all liquid was removed. Trypsin buffer (10 mM ABC, 10% acetonitrile, 13 ng/ μ L trypsin) was added to cover gel pieces, and they were left to incubate for 30 min at 4°C when more trypsin buffer was added to cover the gel pieces. After another 90 min, 20 μ L of ABC was used to cover the gel pieces during digestion. Gel pieces in buffer were incubated overnight at 37°C with constant shaking. Extraction buffer [ACN with 5% formic acid (FA)] was added at a ratio of (2:1 vol/vol, buffer to sample) and incubated for 15 min at 37°C with constant shaking. Digested peptides were filtered through a 10 kD MWCO filter and freeze dried before analysis. Directly before running on the instrument peptides were resolubilized in 30 μ L solvent A (5% ACN and 0.1% FA).

Peptide mixtures were analyzed using one-dimensional liquid chromatography-tandem mass spectrometry (LC-MS/MS) on a Vanquish uHPLC (Thermo Scientific) coupled to a Q Exactive Plus mass spectrometer (Thermo-Fisher Scientific). For each sample, a 10 μ L inject (representing 50% of the total in-gel digest) was loaded to an in-house-built nanospray emitter (75 μ m inner diameter) packed with 1.7 μ m Kinetex C18 reversed phase resin (Phenomenex) to 15 cm. Sample loading proceeded at 2 μ L/min at 100% solvent A (2% acetonitrile [ACN], 98% H₂O and 0.1% formic acid [FA]) for 30 min. The loaded peptides were then separated, eluted, and analyzed by data-dependent acquisition MS/MS over a 90 min organic gradient (300 nL/min flow rate; 0%–30% solvent B [70% ACN, 30% H₂O and 0.1% FA]). All MS data were acquired with Xcalibur (Thermo Scientific; version 4.2.47) software using the top N method (N up to 20). Target values for the full-scan MS spectra were 1×10^6 charges in the 400–1,500 m/z range at 70,000 resolution with a maximum injection time of 25 ms. Analyte precursors with +2 or +3 charges were isolated within a 1.8 m/z window (with 0.3 m/z offset) and fragmented in the HCD cell by applying a normalized collision energy of 27 eV. Resulting MS/MS scans were measured at a resolution of 17,500 with an ion target value of 1×10^5 and a 50 ms maximum injection time. Dynamic exclusion was set to 30 s to avoid repeated sequencing of peptides. All MS raw data files were analyzed using the Proteome Discoverer software (Thermo-Fisher Scientific, version 2.5) (57). Each MS

raw data file was processed by the SEQUEST HT database search algorithm (58) and confidence in peptide-to-spectrum (PSM) matching was evaluated by Percolator (59). Peptides were analyzed using the semi-tryptic search feature in SEQUEST HT.

Purification of hemophilin homologs

Hemophilin from *H. haemolyticus* was expressed and purified as previously described to yield low and high heme-content fractions after anion exchange chromatography (19). Heme was removed by cold acid acetone treatment to yield an apo hemophilin fraction, as previously described (60). Residual Fe(III) heme was estimated at 1.8% of sites, based on extinction coefficients of met-hemophilin being $96,100 \text{ M}^{-1} \text{ cm}^{-1}$ and $38,600 \text{ M}^{-1} \text{ cm}^{-1}$ at 414 nm and 280 nm, respectively, and extinction coefficient of the apo-protein being $25,900 \text{ M}^{-1} \text{ cm}^{-1}$ at 280 nm.

Expression constructs encoding the hemophilin homologs from *X. nematophila* (amino acid residues 23–247), *X. cabanillasii* (amino acid residues 23–238), and *A. baumannii* (amino acid residues 21–264) were constructed in pET28a. In each case, the native N-terminal signal peptide was omitted and replaced with a hexa-histidine tag and engineered tobacco etch virus (TEV) protease cleavage site. Clones were transformed into *E. coli* strain Rossetta-2 (Novagen), grown in LB containing 34 $\mu\text{g/mL}$ chloramphenicol and 25 $\mu\text{g/mL}$ kanamycin; expression was induced with 1 mM IPTG for 3 h shaking at 37°C. Cells were suspended in lysis buffer (0.5 M NaCl, 0.05 M sodium phosphate, 0.02 M imidazole, 100 μM phenylmethylsulfonyl fluoride, pH 7.2) and lysed by sonication (Branson); then hemophilin homologs were captured by Ni-affinity chromatography. TEV protease was expressed and purified as described (61). The His-tag was cleaved from hemophilin homologs by TEV protease treatment overnight at room temperature, to liberate hemophilin proteins with an additional N-terminal Gly-His-Met tripeptide residual from the TEV cleavage site. TEV protease and His-tag peptides were removed over a second Ni-affinity column. Hemophilin preparations from *X. nematophila* and *A. baumannii* had a brownish appearance and an absorbance peak at ~413 nm characteristic of a porphyrin ligand, as well as less intense absorption peaks at 533 and 659 nm. Ligand was estimated to occupy ~25% of sites based on comparison with spectra of hemophilin from *H. haemolyticus*. In contrast, CrpC from *X. cabanillasii* was colorless. Acid acetone or methyl ethyl ketone extraction was not effective to remove colored contaminants from HrpC of *X. nematophila* or *A. baumannii*. Apo-protein fractions of these proteins were prepared by reversed-phase HPLC over a C4 stationary phase (Waters Symmetry) developed with a CH_3CN :water mobile phase gradient containing 0.1% trifluoroacetic acid. Solvent was removed by lyophilization. Apo-CrpC from *X. cabanillasii* was applied to a strong anion exchange resin (Q sepharose, Pharmacia) in 25 mM Tris-HCl buffer (pH 8.25 at 21°C) and collected in the flow-through. All apo-proteins were dialyzed into 20 mM Tris-HCl buffer (pH 7.9 at 21°C) prior to storage at –80°C. Apo-protein concentrations were determined by absorption extinction coefficient at 280 nm calculated from amino acid composition.

UV-visible absorption and fluorescence spectroscopy

UV-visible spectra were recorded on a Jasco V-630 spectrophotometer fitted with a temperature-controlled sample holder (Jasco) and spectroil quartz cuvettes with a path length of 1.0 cm (Starna, Baulkham Hills, Australia). Porphyrin concentrations were determined according to the following molar extinction coefficients and solvent conditions: hemin chloride, $\epsilon_{385} = 58,400 \text{ M}^{-1} \text{ cm}^{-1}$ in 0.1 M NaOH (62); PPIX $\epsilon_{554} = 13,500 \text{ M}^{-1} \text{ cm}^{-1}$ in 2.7 M HCl (62); coproporphyrin III, $\epsilon_{548} = 16,800 \text{ M}^{-1} \text{ cm}^{-1}$ in 0.1 M HCl (63); Co(III)PPIX $\epsilon_{424} = 180,000 \text{ M}^{-1} \text{ cm}^{-1}$ in NaOH (0.1 M):pyridine:H₂O 3:10:17 (64). Zn(II)PPIX was determined as free PPIX after decomposition in 2.7 M HCl. To prepare Zn-PPIX, 0.5 g PPIX (Frontier Scientific) was dissolved in boiling chloroform (100 mL) to which a saturated solution of Zn acetate in MeOH (1 mL) was added. The mixture was refluxed for 20 min and then a small amount of MeOH was added, and after cooling, the dark red solid was filtered off (50:50 Zn-PPIX:PPIX by HPLC). Zn-PPIX was purified by

RP-HPLC over a C18 solid phase (Phenomenex) with isocratic acetone:MeOH:water:formic acid (280:120:100:1) mobile phase, which achieved baseline separation of the Zn-PPIX fraction.

Hemin-binding measurements in absorbance mode were made by successive additions of apo-protein (~0.4 mM stock) into porphyrin solution (1.5 μ M) in 20 mM Tris-HCl, pH 8 at 21°C. All binding experiments were performed at 21°C. Data were fitted to a 1:1 binding model accounting for ligand depletion,

$$F_{\text{obs}} = F_0 - F_{\text{sat}}((L + K_d + M) - \sqrt{(L + K_d + M)^2 - 4ML})$$

where F_{obs} is fluorescence signal, F_0 is the starting fluorescence, F_{sat} is a scaling factor for fluorescence at saturation, L and M are the ligand and macromolecule concentrations, respectively, and K_d is the equilibrium dissociation constant. Data were fitted using GNUPLLOT version 4.6. A 95% confidence interval for the K_d parameter was obtained by determining a threshold sum-of-squares for which a fit with all fixed parameters would not be significantly different from the best-fit model at a significance level of $P = 0.05$ (65).

$$SS_{\text{fixed}} = SS_{\text{best}}((F \times \text{Dfn})/(\text{Dfd} + 1))$$

Here, SS_{best} is the sum-of-squares for the fit with all parameters floated; Dfn and Dfd are the degrees of freedom in the numerator and denominator, respectively, for the calculation of F . The K_d value was then fixed at values above or below the best-fit K_d until fits exceeded the threshold sum-of-squares. To determine a lower limit of K_d that we could expect to fit from absorbance measurements we simulated data for different K_d values with Gaussian noise added to give fits with SS_{best} that matched our experimental data. We determined by F test that K_d values ≤ 15 nM were not significantly different (at $P = 0.05$) from arbitrarily high K_d (approximating a straight-line fit); thus, we decided on 15 nM as a lower cutoff for reporting K_d .

Fluorescence measurements for Zn(II)PPIX binding were made in 96-well format in 20 mM Tris-HCl, pH 8.0 at 21°C using individually mixed samples covering an appropriate range of hemophilin protein concentrations based on preliminary experiments. IC_{50} for hemophilins binding to Zn(II)PPIX (0.5 or 1.0 μ M) in competition with BSA (15 μ M) were determined by fitted to the same 1:1 binding model as described above. The Cheng and Prusoff equation (29) was used to convert to K_d ,

$$K_d = IC_{50} / (1 + [BSA]/K_{d,BSA})$$

where [BSA] is the molar concentration of competing BSA (15 μ M) and $K_{d,BSA}$ was the dissociation equilibrium constant for BSA binding to Zn(II)PPIX, determined to be 0.3 (0.2–0.4) μ M from absorbance titration data.

Generation of threading models for hemophilin active sites

Threading models were generated with the program MODELLER version 10.2 (33), using the X-ray crystal structures of *H. Haemolyticus* Hpl (PDB 6om5) and *A. baumannii* HphA (PDB 7red) as templates (14, 19). Residues of the N-terminal signal peptide, as identified by SIGNALP 5.0 (66), were removed prior to analysis or modeling of all sequences. The threading approach used by MODELLER relies on accurate alignment of template and query sequences. To identify residues likely to be important for structural integrity of the hemophilin ligand binding domain, we collected the small number of sequences that shared similarity with Hpl, HphA, and *X. nematophila* HrpC in the N-terminal region, or with pair-wise combinations of these, based on BLASTP searches. We aligned these sequences with the structure-based alignment of Hpl and HphA produced by MODELLER to generate a profile against which subclusters from the sequence similarity network analysis were then aligned using CLUSTALO. Alignments

were inspected against several criteria to see that they were structurally plausible. We specifically examined the preservation of the conserved hydrophobic core ligand-binding domains of Hpl and HphA, corresponding to Hpl residues G26, I38, I40, A49, V51, I53, F76, M79, A83, A104, L125, F127, Y136, G138, W140, Y156, A158 (Fig. 9). An initial structure-based alignment of *H. haemolyticus* Hpl and *A. baumannii* HphA was produced using MODELLER, and this pairing was fixed in all subsequent alignments. Amino acid residues with potential importance for the hemophilin fold were identified by sequence homology (with the signal peptides and BBD removed) using BLASTP (67). Sequences were identified with similarity to the hemophilin domains of Hpl, HphA, and the N-terminal region of *X. nematophila* HrpC (WP_019473020, WP_057440571, WP_244182492), or Hpl and HphA only (WP_005758278), or HphA and HrpC only (MBP6115507, WP_228864429, WP_232888613), or Hpl and HrpC only (WP_038256617, WP_092512525, WP_244182492), or HphA only (WP_121975315), or Hpl only (RKV63521). These sequences were aligned with the structure-based alignment of Hpl and HphA using CLUSTAL OMEGA (68). A profile was generated from this sequence alignment, and this profile was aligned with the profiles of the hemophilin network clusters, Plant/Environmental, HphA-like, HrpC-like, and CrpC-like, using CLUSTAL OMEGA. Alignments were colored using MVIEW (69).

ACKNOWLEDGMENTS

We thank Dr. Brianna Atto for useful discussions, Sarah Kauffman for technical assistance, and Dr. Richard Giannone for assistance with the mass spectrometry measurements.

This work was funded, in part, by funds awarded to H.G.-B. from the University of Tennessee-Knoxville (David and Sandra White Endowment) and the National Science Foundation (IOS-2128266). This work was funded, in part, by a grant from the Clifford Craig Foundation awarded to Dr. Stephen Tristram (CCF192). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

The authors have no competing interests that might be perceived to influence the results and/or discussion reported in this paper.

AUTHOR AFFILIATIONS

¹Department of Microbiology, University of Tennessee Knoxville, Knoxville, Tennessee, USA

²School of Medicine, University of Tasmania, Hobart, Tasmania, Australia

³Department of Plant and Soil Sciences, University of Delaware, Newark, Delaware, USA

⁴Bioscience Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA

PRESENT ADDRESS

Alex S. Grossman, Department of Microbiology, ADA Forsyth Institute, Cambridge, Massachusetts, USA

AUTHOR ORCID*s*

Alex S. Grossman  <http://orcid.org/0000-0003-1661-0383>

David A. Gell  <http://orcid.org/0000-0003-0382-1181>

Heidi Goodrich-Blair  <http://orcid.org/0000-0003-1934-2636>

FUNDING

Funder	Grant(s)	Author(s)
University of Tennessee, Knoxville (UT)		Heidi Goodrich-Blair
National Science Foundation (NSF)	IOS-2128266	Heidi Goodrich-Blair
Clifford Craig Foundation	CCF192	David A. Gell

AUTHOR CONTRIBUTIONS

Alex S. Grossman, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review and editing | David A. Gell, Formal analysis, Investigation, Methodology, Resources, Writing – original draft, Writing – review and editing | Derek G. Wu, Investigation | Dana L. Carper, Formal analysis, Investigation, Writing – review and editing | Robert L. Hettich, Formal analysis, Investigation, Writing – review and editing | Heidi Goodrich-Blair, Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review and editing

ADDITIONAL FILES

The following material is available [online](#).

Supplemental Material

Supplemental figures (JB00444-23-s0001.pdf). Figures S1 to S6.

Supplemental file 1 (JB00444-23-s0002.xlsx). Rodeo and network subclusters.

Supplemental file 2 (JB00444-23-s0003.xlsx). Western images.

Supplemental file 3 (JB00444-23-s0004.xlsx). Strains and primers.

REFERENCES

- Bertini I, Gray HB, Stiefel EI, Selverstone Valentine J. 2007. Biological inorganic chemistry: structure and reactivity. University Science Books, Sausalito.
- Zhang H-R, Wang Y, Xiu P, Qi Y, Chai F. 2021. Roles of iron limitation in phytoplankton dynamics in the Western and Eastern subarctic Pacific. *Front Mar Sci* 8. <https://doi.org/10.3389/fmars.2021.735826>
- Hood MI, Skaar EP. 2012. Nutritional immunity: transition metals at the pathogen–host interface. *Nat Rev Microbiol* 10:525–537. <https://doi.org/10.1038/nrmicro2836>
- Khan FA, Fisher MA, Khakoo RA. 2007. Association of hemochromatosis with infectious diseases: expanding spectrum. *Int J Infect Dis* 11:482–487. <https://doi.org/10.1016/j.ijid.2007.04.007>
- Posey JE, Gherardini FC. 2000. Lack of a role for iron in the Lyme disease pathogen. *Science* 288:1651–1653. <https://doi.org/10.1126/science.288.5471.1651>
- Holden VI, Bachman MA. 2015. Diverging roles of bacterial siderophores during infection. *Metallomics* 7:986–995. <https://doi.org/10.1039/c4mt00333k>
- Hooda Y, Lai CC-L, Judd A, Buckwalter CM, Shin HE, Gray-Owen SD, Moraes TF. 2016. Slam is an outer membrane protein that is required for the surface display of lipidated virulence factors in *Neisseria*. *Nat Microbiol* 1:16009. <https://doi.org/10.1038/nmicrobiol.2016.9>
- Grossman AS, Mauer TJ, Forest KT, Goodrich-Blair H, Comstock L. 2021. A widespread bacterial secretion system with diverse substrates. *mBio* 12:e0195621. <https://doi.org/10.1128/mBio.01956-21>
- Huynh MS, Hooda Y, Li R, Jagielnicki M, Lai C-L, Moraes TF. 2021. Reconstitution of surface lipoprotein translocation reveals Slam as an outer membrane translocon in Gram-negative bacteria. *bioRxiv*. <https://doi.org/10.1101/2021.08.22.457263>
- Hooda Y, Lai CCL, Moraes TF. 2017. Identification of a large family of Slam-dependent surface lipoproteins in Gram-negative bacteria. *Front Cell Infect Microbiol* 7:207. <https://doi.org/10.3389/fcimb.2017.00207>
- Chen CJ, Sparling PF, Lewis LA, Dyer DW, Elkins C. 1996. Identification and purification of a hemoglobin-binding outer membrane protein from *Neisseria gonorrhoeae*. *Infect Immun* 64:5008–5014. <https://doi.org/10.1128/iai.64.12.5008-5014.1996>
- Calmettes C, Alcantara J, Yu R-H, Schryvers AB, Moraes TF. 2012. The structural basis of transferrin sequestration by transferrin-binding protein B. *Nat Struct Mol Biol* 19:358–360. <https://doi.org/10.1038/nsmb.2251>
- Ostan NKH, Yu R-H, Ng D, Lai CC-L, Pogoutse AK, Sarpe V, Hepburn M, Sheff J, Raval S, Schriener DC, Moraes TF, Schryvers AB. 2017. Lactoferrin binding protein B - a bi-functional bacterial receptor protein. *PLoS Pathog* 13:e1006244. <https://doi.org/10.1371/journal.ppat.1006244>
- Bateman TJ, Shah M, Ho TP, Shin HE, Pan C, Harris G, Fegan JE, Islam EA, Ahn SK, Hooda Y, Gray-Owen SD, Chen W, Moraes TF. 2021. A Slam-dependent hemophore contributes to heme acquisition in the bacterial pathogen *Acinetobacter baumannii*. *Nat Commun* 12:6270. <https://doi.org/10.1038/s41467-021-26545-9>
- Arnoux P, Haser R, Izadi-Pruneyre N, Lecroisey A, Czjzek M. 2000. Functional aspects of the heme bound hemophore HasA by structural analysis of various crystal forms. *Proteins* 41:202–210. [https://doi.org/10.1002/1097-0134\(20001101\)41:2<202::aid-prot50>3.0.co;2-8](https://doi.org/10.1002/1097-0134(20001101)41:2<202::aid-prot50>3.0.co;2-8)
- Gianquinto E, Moscetti I, De Bei O, Campanini B, Marchetti M, Luque FJ, Cannistraro S, Ronda L, Bizzarri AR, Spyarakis F, Bettati S. 2019. Interaction of human hemoglobin and semi-hemoglobins with the *Staphylococcus aureus* hemophore IsdB: a kinetic and mechanistic insight. *Sci Rep* 9:18629. <https://doi.org/10.1038/s41598-019-54970-w>
- Huynh MS, Hooda Y, Li YR, Jagielnicki M, Lai CC-L, Moraes TF. 2022. Reconstitution of surface lipoprotein translocation through the Slam translocon. *Elife* 11:e72822. <https://doi.org/10.7554/eLife.72822>
- Fantappiè L, Irene C, De Santis M, Armini A, Gagliardi A, Tomasi M, Parri M, Cafardi V, Bonomi S, Ganfini L, Zerbini F, Zanella I, Carnemolla C, Bini L, Grandi A, Grandi G. 2017. Some Gram-negative lipoproteins keep their surface topology when transplanted from one species to another and deliver foreign polypeptides to the bacterial surface. *Mol Cell Proteomics* 16:1348–1364. <https://doi.org/10.1074/mcp.M116.065094>
- Latham RD, Torrado M, Atto B, Walshe JL, Wilson RK, Guss JM, Mackay JP, Tristram S, Gell DA. 2020. A heme-binding protein produced by *Haemophilus haemolyticus* inhibits non-typeable *Haemophilus influenzae*. *Mol Microbiol* 113:381–398. <https://doi.org/10.1111/mmi.14426>
- Schwiesow L, Mettert E, Wei Y, Miller HK, Herrera NG, Balderas D, Kiley PJ, Auerbuch V. 2018. Control of *hmu* heme uptake genes in *Yersinia pseudotuberculosis* in response to iron sources. *Front Cell Infect Microbiol* 8:47. <https://doi.org/10.3389/fcimb.2018.00047>
- Atto B, Latham RD, Kunde D, Gell DA, Tristram S. 2020. *In vitro* probiotic potential of hemophilin-producing strains of *Haemophilus haemolyticus*. *bioRxiv*. <https://doi.org/10.1101/2020.01.02.893487>
- Atto B, Kunde D, Gell DA, Tristram S. 2021. Oropharyngeal carriage of *hpl*-containing *Haemophilus haemolyticus* predicts lower prevalence and density of NTHi colonisation in healthy adults. *Pathogens* 10:577. <https://doi.org/10.3390/pathogens10050577>
- Zallot R, Oberg NO, Gerlt JA. 2019. The EFI web resource for genomic enzymology tools: leveraging protein, genome, and metagenome databases to discover novel enzymes and metabolic pathways.

- Biochemistry 58:4169–4182. <https://doi.org/10.1021/acs.biochem.9b00735>
24. Tietz JI, Schwalen CJ, Patel PS, Maxson T, Blair PM, Tai H-C, Zakai UI, Mitchell DA. 2017. A new genome-mining tool redefines the lasso peptide biosynthetic landscape. *Nat Chem Biol* 13:470–478. <https://doi.org/10.1038/nchembio.2319>
 25. Egler M, Grosse C, Grass G, Nies DH. 2005. Role of the extracytoplasmic function protein family sigma factor RpoE in metal resistance of *Escherichia coli*. *J Bacteriol* 187:2297–2307. <https://doi.org/10.1128/JB.187.7.2297-2307.2005>
 26. Schwartz CJ, Giel JL, Patschkowski T, Luther C, Ruzicka FJ, Beinert H, Kiley PJ. 2001. IscR, an Fe-S cluster-containing transcription factor, represses expression of *Escherichia coli* genes encoding Fe-S cluster assembly proteins. *Proc Natl Acad Sci U S A* 98:14895–14900. <https://doi.org/10.1073/pnas.251550898>
 27. Maddocks SE, Oyston PCF. 2008. Structure and function of the LysR-type transcriptional regulator (LTTR) family proteins. *Microbiology (Reading)* 154:3609–3623. <https://doi.org/10.1099/mic.0.2008/022772-0>
 28. Grossman AS, Escobar CA, Mans EJ, Mucci NC, Mauer TJ, Jones KA, Moore CC, Abraham PE, Hettich RL, Schneider L, Campagna SR, Forest KT, Goodrich-Blair H. 2022. A surface exposed, two-domain lipoprotein cargo of a type XI secretion system promotes colonization of host intestinal epithelia expressing glycans. *Front Microbiol* 13:800366. <https://doi.org/10.3389/fmicb.2022.800366>
 29. Cheng Y, Prusoff WH. 1973. Relationship between the inhibition constant (K_i) and the concentration of inhibitor which causes 50 per cent inhibition (I_{50}) of an enzymatic reaction. *Biochem Pharmacol* 22:3099–3108. [https://doi.org/10.1016/0006-2952\(73\)90196-2](https://doi.org/10.1016/0006-2952(73)90196-2)
 30. Holm L, Rosenström P. 2010. Dali server: conservation mapping in 3D. *Nucleic Acids Res* 38:W545–W549. <https://doi.org/10.1093/nar/gkq366>
 31. Sillitoe I, Bordin N, Dawson N, Waman VP, Ashford P, Scholes HM, Pang CSM, Woodridge L, Rauer C, Sen N, Abbasian M, Le Cornu S, Lam SD, Berka K, Varekova IH, Svobodova R, Lees J, Orengo CA. 2021. CATH: increased structural coverage of functional space. *Nucleic Acids Res* 49:D266–D273. <https://doi.org/10.1093/nar/gkaa1079>
 32. Andreeva A, Kulesha E, Gough J, Murzin AG. 2020. The SCOP database in 2020: expanded classification of representative family and superfamily domains of known protein structures. *Nucleic Acids Res* 48:D376–D382. <https://doi.org/10.1093/nar/gkz1064>
 33. Webb B, Sali A. 2016. Comparative protein structure modeling using MODELLER. *Curr Protoc Bioinformatics* 54:5. <https://doi.org/10.1002/cpbi.3>
 34. Bhasin A, Chaston JM, Goodrich-Blair H. 2012. Mutational analyses reveal overall topology and functional regions of NilB, a bacterial outer membrane protein required for host association in a model of animal-microbe mutualism. *J Bacteriol* 194:1763–1776. <https://doi.org/10.1128/JB.06711-11>
 35. Cowles CE, Goodrich-Blair H. 2008. The *Xenorhabdus nematophila* nilABC genes confer the ability of *Xenorhabdus* spp. to colonize *Steinernema carpocapsae* nematodes. *J Bacteriol* 190:4121–4128. <https://doi.org/10.1128/JB.00123-08>
 36. Moraes TF, Lai C-L, Hooda Y, Judd A, Schryvers AB, Gray-Owen SD. 2021. Slam polynucleotides and polypeptides and uses thereof. US20210379176A1.
 37. Debret G, Martel A, Cuniasse P. 2009. RASMOT-3D PRO: a 3D motif search webserver. *Nucleic Acids Res* 37:W459–W464. <https://doi.org/10.1093/nar/gkp304>
 38. Leys D, Backers K, Meyer TE, Hagen WR, Cusanovich MA, Van Beeumen JJ. 2000. Crystal structures of an oxygen-binding cytochrome c from *Rhodobacter sphaeroides*. *J Biol Chem* 275:16050–16056. <https://doi.org/10.1074/jbc.275.21.16050>
 39. Johnson EA, Russo MM, Nye DB, Schlessman JL, Lecomte JTJ. 2018. Lysine as a heme iron ligand: a property common to three truncated hemoglobins from *Chlamydomonas reinhardtii*. *Biochim Biophys Acta Gen Subj* 1862:2660–2673. <https://doi.org/10.1016/j.bbagen.2018.08.009>
 40. Girvan HM, Toogood HS, Littleford RE, Seward HE, Smith WE, Ekanem IS, Leys D, Cheesman MR, Munro AW. 2009. Novel haem co-ordination variants of flavocytochrome P450BM3. *Biochem J* 417:65–76. <https://doi.org/10.1042/BJ20081133>
 41. Safarian S, Rajendran C, Müller H, Preu J, Langer JD, Ovchinnikov S, Hirose T, Kusumoto T, Sakamoto J, Michel H. 2016. Structure of a *bd* oxidase indicates similar mechanisms for membrane-integrated oxygen reductases. *Science* 352:583–586. <https://doi.org/10.1126/science.aaf2477>
 42. Murali R, Gennis RB. 2018. Functional importance of Glutamate-445 and Glutamate-99 in proton-coupled electron transfer during oxygen reduction by cytochrome *bd* from *Escherichia coli*. *Biochim Biophys Acta Bioenerg* 1859:577–590. <https://doi.org/10.1016/j.bbabi.2018.04.012>
 43. Burmester T, Hankeln T. 2007. The respiratory proteins of insects. *J Insect Physiol* 53:285–294. <https://doi.org/10.1016/j.jinsphys.2006.12.006>
 44. Buecher EJ, Popiel I. 1989. Liquid culture of the entomogenous nematode *Steinernema feltiae* with its bacterial symbiont. *J Nematol* 21:500–504.
 45. Hieb WF, Stokstad EL, Rothstein M. 1970. Heme requirement for reproduction of a free-living nematode. *Science* 168:143–144. <https://doi.org/10.1126/science.168.3927.143>
 46. Weber JJ, Kanost MR, Gorman MJ. 2020. Iron binding and release properties of transferrin-1 from *Drosophila melanogaster* and *Manduca sexta*: implications for insect iron homeostasis. *Insect Biochem Mol Biol* 125:103438. <https://doi.org/10.1016/j.ibmb.2020.103438>
 47. Goodrich-Blair H. 2007. They've got a ticket to ride: *Xenorhabdus nematophila*-*Steinernema carpocapsae* symbiosis. *Curr Opin Microbiol* 10:225–230. <https://doi.org/10.1016/j.mib.2007.05.006>
 48. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13:2498–2504. <https://doi.org/10.1101/gr.1239303>
 49. Bastian M, Heymann S, Jacomy M. 2009. Gephi: an open source software for exploring and manipulating networks. *Proc Int AAAI Conf Web Soc Media* 3:361–362. <https://doi.org/10.1609/icwsm.v3i1.13937>
 50. Fruchterman TMJ, Reingold EM. 1991. Graph drawing by force-directed placement. *Softw Pract Exp* 21:1129–1164. <https://doi.org/10.1002/spe.4380211102>
 51. Jacomy M, Venturini T, Heymann S, Bastian M. 2014. ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLoS One* 9:e98679. <https://doi.org/10.1371/journal.pone.0098679>
 52. Miroux B, Walker JE. 1996. Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J Mol Biol* 260:289–298. <https://doi.org/10.1006/jmbi.1996.0399>
 53. Dumon-Seignovart L, Cariot G, Vuillard L. 2004. The toxicity of recombinant proteins in *Escherichia coli*: a comparison of overexpression in BL21(DE3), C41(DE3), and C43(DE3). *Protein Expr Purif* 37:203–206. <https://doi.org/10.1016/j.pep.2004.04.025>
 54. Koontz L. 2014. Chapter one - TCA precipitation, p 3–10. In Lorsch JBT-M (ed), *Laboratory methods in enzymology: protein part C*. Academic Press.
 55. Tukey JW. 1949. Comparing individual means in the analysis of variance. *Biometrics* 5:99–114. <https://doi.org/10.2307/3001913>
 56. Shevchenko A, Tomas H, Havlis J, Olsen JV, Mann M. 2006. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat Protoc* 1:2856–2860. <https://doi.org/10.1038/nprot.2006.468>
 57. Orsburn BC. 2021. Proteome discoverer-A community enhanced data processing suite for protein informatics. *Proteomes* 9:15. <https://doi.org/10.3390/proteomes9010015>
 58. Eng JK, McCormack AL, Yates JR. 1994. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* 5:976–989. [https://doi.org/10.1016/1044-0305\(94\)80016-2](https://doi.org/10.1016/1044-0305(94)80016-2)
 59. Käll L, Canterbury JD, Weston J, Noble WS, MacCoss MJ. 2007. Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat Methods* 4:923–925. <https://doi.org/10.1038/nmeth.1113>
 60. Ascoli F, Rossi Fanelli MR, Antonini EBT-M. 1981. Preparation and properties of apohemoglobin and reconstituted hemoglobins, p 72–87. In *Hemoglobins*. Academic Press.

61. van den Berg S, Löfdahl P-A, Härd T, Berglund H. 2006. Improved solubility of TEV protease by directed evolution. *J Biotechnol* 121:291–298. <https://doi.org/10.1016/j.jbiotec.2005.08.006>
62. Dawson RMC, Elliott DC, Elliott WH, Jones KM. 1989. *Data for biochemical research*. Third Edition. Clarendon Press, Oxford.
63. Fuhrhop JH, Smith KM. 1975. *Laboratory methods in porphyrin and metalloporphyrin research*. Elsevier Scientific Publishing Company.
64. Yonetani T, Yamamoto H, Woodrow III GV. 1974. Studies on cobalt myoglobins and hemoglobins. I. Preparation and optical properties of myoglobins and hemoglobins containing cobalt proto-, meso-, and deuteroporphyrins and thermodynamic characterization of their reversible oxygenation. *J Biol Chem* 249:682–690. [https://doi.org/10.1016/S0021-9258\(19\)42984-0](https://doi.org/10.1016/S0021-9258(19)42984-0)
65. Motulsky H, Christopoulos A. 2004. *Fitting models to biological data using linear and nonlinear regression: a practical guide to curve fitting*. Oxford University Press.
66. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H. 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol* 37:420–423. <https://doi.org/10.1038/s41587-019-0036-z>
67. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215:403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
68. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7:539. <https://doi.org/10.1038/msb.2011.75>
69. Brown NP, Leroy C, Sander C. 1998. MView: a web-compatible database search or multiple alignment viewer. *Bioinformatics* 14:380–381. <https://doi.org/10.1093/bioinformatics/14.4.380>