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

Short title: Specific secretion of hemophilin proteins with diverse substrates

Authors: Alex S. Grossman[1], David A. Gell[2‡], Derek G. Wu[1,3], Heidi Goodrich-Blair[1†]

Corresponding authors: †: hgblair@utk.edu ‡: david.gell@utas.edu.au

Affiliations: 1. Department of Microbiology, University of Tennessee Knoxville, Knoxville, TN, USA. 2. School of Medicine, University of Tasmania, 17 Liverpool Street, Hobart, TAS 7000, Australia. 3. Department of Plant and Soil Sciences, University of Delaware, Newark, DE, USA.

Author contributions: Alex S. Grossman performed, and generated figures for all network analyses and co-expression experiments. Derek G. Wu assisted Alex S. Grossman with the chimeric hemophilin secretion experiment. David A. Gell performed all physicochemical characterization experiments on the identified hemophilin homologs, including ligand binding and structural modelling, and generated the associated figures. Alex S. Grossman, David A. Gell, and Heidi Goodrich-Blair wrote the text of this manuscript and Derek G. Wu assisted in proofreading.

Abstract:

Cellular life relies on enzymes that require metal cofactors, 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,

23 and nutritional immunity evasion. This Sec-dependent, Gram-negative secretion system is
24 encoded by organisms throughout the phylum Proteobacteria, including human pathogens
25 *Neisseria meningitidis, Proteus mirabilis, Acinetobacter baumannii,* and *Haemophilus*
26 *influenzae*. Experimentally verified T11SS-dependent cargo include host metal acquisition
27 proteins transferrin binding protein B (TbpB) and lactoferrin binding protein B (LbpB), as well as
28 the hemophilin homologs heme receptor protein C (HrpC) and hemophilin A (HphA), the
29 complement immune evasion protein factor-H binding protein (fHbp), and the host symbiosis
30 factor nematode intestinal localization protein C (NilC). Secretion of each of these cargo
31 proteins relies on a specific T11SS. Here, we examined the specificity of T11SS systems for their
32 cognate cargo proteins using taxonomically distributed homolog pairs of T11SS and hemophilin
33 cargo and explore the ligand binding ability of those hemophilin homologs. Our comparative
34 ligand binding analysis of four hemophilin family proteins identified previously unknown ligand
35 binding diversity within this protein family, which informed our description of structural
36 features that are likely to contribute to heme/porphyrin binding specificity. *In vivo* expression
37 of hemophilin homologs revealed that each was secreted in a specific manner by its cognate
38 T11SS protein. Furthermore, secretion assays of chimeric hemophilin proteins revealed that
39 specificity is predominantly dictated by the C-terminal domain of the cognate cargo.
40 Meanwhile, the N-terminal effector domains of these T11SS-dependent cargo proteins feature
41 porphyrin binding pockets that drive ligand binding affinity and specificity. In light of these
42 results, we have termed this N-terminal domain the hemophilin ligand binding domain (Hlb)
43 after its first characterized representative.
44

45 **Introduction:**

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

64 Several recent studies have linked the type eleven secretion system (T11SS) outer
65 membrane proteins (OMPs) and their cargo proteins to iron uptake strategies in Gram-negative
66 bacteria. In *Neisseria*, the T11SS proteins Slam1 and Slam2 surface expose cargo proteins that

67 are responsible for binding host-metal carriers: Transferrin binding protein B (TbpB) and
68 Lactoferrin binding protein B (LbpB) are surface exposed by Slam1, and
69 Hemoglobin/Haptoglobin binding protein A (HpuA) is surface exposed by Slam2 [7–10]. These
70 surface exposed outer membrane lipoproteins facilitate bacterial colonization by capturing
71 their respective host factors (transferrin, lactoferrin, or hemoglobin/haptoglobin) and
72 complexing with a TonB-dependent uptake channel capable of importing the iron cofactor.
73 Since surface exposure is essential for the function of these lipoproteins, genetic inactivation of
74 the T11SS OMP, Slam1 prevents effective colonization and pathogenesis by *Neisseria* [11].
75 While *Neisseria* Slam1 and Slam2 have specificity for their respective cargo [7], no underlying
76 mechanism for specificity has been proposed yet and it is unknown if all T11SS have specificity
77 for their cognate cargo. Bioinformatic analyses revealed a large number of potential T11SS-
78 dependent cargo, lipid anchored and unanchored, which frequently exist in cognate
79 pairs/groupings according to genomic co-occurrence analyses [11,12]. To date, all verified or
80 predicted T11SS-dependent cargo have two distinct domains: an N-terminal domain that varies
81 in predicted structure and function across cargo, and a C-terminal, 8-stranded β-barrel domain
82 from the TbpBBD or lipoprotein C families, though the latter is not exclusive to lipoproteins but
83 also occurs in non-lipidated proteins [7,11,12].
84 T11SS are capable of secreting unlipidated cargo proteins, such as the soluble
85 hemophores heme receptor protein C from *X. nematophila* (HrpC) and hemophilin A from *A.*
86 *baumanii* (HphA) [12,13]. HphA likely captures heme from hemoglobin and other host
87 hemoproteins, and contributes to the virulence of *Acinetobacter baumannii* in a murine
88 infection model through its role as a co-receptor to the TonB-dependent heme receptor HphR

89 [13]. Thus, hemophilin proteins represent a high-affinity heme acquisition system comparable
90 to HasA from *Serratia marcescens* [14] or IsdB from *Staphylococcus aureus* [15]. Known
91 members of the hemophilin protein family function to import heme from a host environment
92 as depicted in Figure 1. Compiling the results of published data from multiple organisms into a
93 single model suggests that that hemophilin crosses the inner membrane through the Sec
94 translocon to reach the periplasm [12], may interact with chaperones such as Skp to traverse
95 the periplasm [16], and then crosses the outer membrane in a T11SS-dependent manner to
96 reach the extracellular milieu [7,12,13,17]. From here, hemophilin captures heme that is
97 released from hemoglobin or other host hemoproteins using a high affinity binding domain
98 [13,18]. Holo-hemophilin is predicted to interact with a TonB-dependent outer membrane co-
99 receptor which imports the heme molecule into the periplasm and releases apo-hemophilin.
100 Finally, the hemin utilization system delivers periplasmic heme into the cytoplasm for
101 incorporation into cellular processes or digestion by heme oxygenase to free the iron cation
102 [19].
103 Within this overall framework, hemophilin can have diverse functional roles within
104 different organisms and environments. For example, in *Haemophilus haemolyticus* hemophilin
105 can act as a probiotic factor by making bioavailable iron inaccessible to nontypeable
106 *Haemophilus influenzae* [20]. Hemophilin producing strains of *H. haemolyticus* inhibit the
107 growth of *H. influenzae* significantly more than non-hemophilin encoding strains in co-cultured
108 media [18] and cell cultures [20]. Additionally, oropharyngeal sampling of human subjects
109 indicates that individuals who carry hemophilin encoding *H. haemolyticus* are between 2.43 and
110 2.67 times less likely to carry nontypeable *H. influenzae* [21]. Conversely, within the

111 opportunistic pathogen *A. baumannii,* hemophilin can act as a virulence factor by facilitating
112 systemic infection in a murine model [13]. Furthermore, predicted hemophilin homologs found
113 in sequence databases display sequence variation within the heme-binding handle domain,
114 suggesting possible variability in ligand binding. To better understand the fundamental
115 biochemical and biological functions of hemophilin family proteins we sought to expand our
116 understanding of hemophilin secretion and activities across a broader taxonomic range.
117

118 **Results:**

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

121 To explore the relatedness of hemophilin family proteins and to identify subcluster
122 divisions that may reflect divergent function, a sequence similarity network generated through
123 Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST) was overlaid with a taxonomic
124 framework [22] (Fig. S1 and Supplemental File 1). The network was populated with the
125 previously published dataset of T11SS-associated cargo that were not predicted to be lipidated
126 or membrane anchored [12]. This analysis revealed a single major cluster containing all
127 previously described hemophilin proteins (88/107 nodes), one smaller cluster containing
128 uncharacterized proteins from predominantly *Pseudomonas* and *Neisseria* species (9/107
129 nodes), and a few unassociated doublets and singletons (10/107 nodes). To focus this study
130 specifically on hemophilin and its direct evolutionary relatives, all nodes not within the central
131 cluster were removed. The remaining nodes were labeled according to taxonomic family and
132 separated using a force-directed separation algorithm, resulting in five subclusters

133 predominantly populated by seven families from Alphaproteobacteria, Betaproteobacteria, and
134 Gammaproteobacteria (Fig. 2AC). The three subclusters that contained Hpl, HrpC, and HphA,
135 respectively were named Hpl-like, HrpC-like, and HphA-like after their respective characterized
136 member. One novel subcluster was termed 'Cobalt/molybdenum associated' due to its genomic
137 co-occurrence with genes predicted to encode cobalt- or molybdenum-dependent enzymes.
138 Another novel subcluster was termed 'Plant/Environmental' due to the dominant presence of
139 homologs encoded by microbes found in soil, water, and plant-associated environments. The
140 subclusters did not fall exclusively along taxonomic lines. For instance, the Hpl-like subcluster
141 includes *Neisseria* and *Pasteurella* hemophilin homologs that clustered closely together,
142 possibly indicating that these genes may have been horizontally exchanged. Cross-taxonomic-
143 class sequence similarity also was observed in the Plant/Environmental cluster of hemophilin
144 homologs.

145 To identify potential distinguishing features among the five subclusters, we examined
146 the identities of genes that commonly occur within genomic neighborhoods (± 6 genes) with
147 *hrpB*. All sequences from the network were submitted to the Rapid ORF Detection & Evaluation
148 Online (RODEO) web tool that uses profile hidden Markov models to identify co-occurring
149 protein domains (Table S1) [23]. Consistent with previous observations, 92.3% (240/260
150 sequences) of hemophilin homologs, regardless of subcluster, were encoded in association with
151 genes predicted to encode TonB and TonB-dependent receptors, suggesting a strong and
152 consistent link between T11SS cargo and TonB-dependent uptake across the outer membrane
153 [11,12]. However, some subclusters had further informative co-occurrences. For example,
154 hemophilin homologs in the HphA-like subcluster showed nearly universal co-occurrence with

155 genes predicted to encode heme oxygenase (51/53) and the iron-sensing regulator FecR
156 (48/53). Similarly, homologs in the Plant/Environmental subcluster typically were encoded near
157 genes predicted to encode metal responsive regulatory proteins (FecR or Fur) (64/72) and
158 occasionally near genes encoding heme oxygenase (19/72). Additionally, 69/72
159 Plant/Environmental-subcluster loci encoded additional regulatory genes such as RpoE, IscR-
160 family regulators, and LysR-family regulators. RpoE is a sigma factor that responds to extra-
161 cytoplasmic stress and is essential for metal resistance in *E. coli* [24]. IscR regulates iron-sulfur
162 cluster biosynthesis according to cellular demand [25], and LysR-family regulators drive diverse
163 pathways by binding DNA directly in response to co-inducing/co-repressing ligands [26].
164 Homologs in the Cobalt/molybdenum-associated subcluster occurred alongside other predicted
165 T11SS-dependent cargo (14/18) and were either located adjacent to a B12 biosynthetic locus
166 (3/18) or near a formate dehydrogenase locus (13/18). Formate dehydrogenase activity relies
167 on molybdenum metal cofactors, and their co-occurrence with T11SS OMPs may hint at a role
168 for the latter in molybdenum acquisition. The HrpC-like subcluster includes genes predicted to
169 encode redox enzymes, such as formate dehydrogenase (48/81) and NADPH:quinone
170 oxidoreductase (12/81), tRNA synthases and modification systems (30/81), and regulatory
171 proteins including TetR family regulators, FaeA family regulators, FecR, and FecI (32/81).
172 Homologs in the Hpl-like subcluster had few unifying co-occurrences, however many co-
173 occurred with tRNA synthases/modification systems (14/36), specifically selenocysteine tRNA
174 synthases (6/36) (Fig. 2B). Overall, network analysis of hemophilin homologs indicate that the
175 family is diversifying into distinct subfamilies independent of their taxonomic lineage, and that

176 those subfamilies are genomically associated with metal-related cellular activities such as ion
177 uptake, metal responsive regulatory proteins, and metal dependent enzymes.
178
179 **Hemophilin cargo proteins have specificity for their cognate T11SS**
180 Hemophilin homologs in each of the network subfamilies are predicted to be secreted
181 by a T11SS, and many were encoded adjacent to a T11SS OMP. We considered the possibility
182 that co-diversification of hemophilins and their T11SS OMPs has resulted in specificity between
183 cognate pairs. To test the extent of specificity between hemophilin cargo proteins and their
184 paired TXISS OMP secretor, and to further establish the role of TXISS OMPs in secretion of
185 hemophilin cargo, we conducted secretion assays on different cognate or non-cognate pairs,
186 chosen from each of the different subclusters of the hemophilin family network (except for the
187 Plant/Environmental subcluster): $HrpB_{X.nem}$/HrpC from *Xenorhabdus nematophila*,
188 $HrpB_{H.haem}$/Hpl from *Haemophilus haemolyticus*, HsmA/HphA from *A. baumannii*, and
189 CrpB/CrpC from *Xenorhabdus cabanillasii* were cloned into pETDuet-1 based expression vectors
190 to perform co-expression and secretion experiments in *E. coli* BL21 C43. Additionally, plasmids
191 were constructed to co-express $HrpB_{X.nem}$ alongside the non-cognate hemophilin homologs Hpl,
192 HphA, and CrpC. Western blotting of supernatant and cellular lysates was performed to
193 monitor cargo localization to the extracellular milieu. Extracellular HrpC and HphA both ran as
194 single protein bands somewhat larger than would be predicted from mature protein sequence
195 alone, with HrpC appearing at ~30 kD (~24 kD expected) and HrpA appearing at 29 kD (~25 kD
196 expected) (Supplemental file 2: sheets 2 and 4) [27]. In both cases this increase in apparent size
197 is too large to be explained by the presence of the signal peptide, though this larger apparent

198 size is consistent with our previous studies of HrpC [12]. Extracellular Hpl and CrpC were both
199 observed running as two different bands. Hpl had a predominant band at ~32 kD and a minor
200 one at the expected size of ~26 kD, while CrpC had a predominant band at ~19 kD and a minor
201 band at the predicted size of ~27 kD (Supplemental file 2: sheets 3 and 5). Interestingly, the
202 extracellular ~19-kD CrpC band was absent or reduced when the protein was expressed alone
203 or with the non-cognate $HrpB_{X.nem}$, suggesting potential protein modification (e.g., cleavage)
204 prior to, or during, secretion by its cognate T11SS. For enumeration of Hpl and CrpC localization
205 we opted to sum both protein bands to fairly represent all of the secreted protein.

206 Co-expression of hemophilin homologs with a cognate T11SS protein always significantly
207 increased the concentration of cargo protein found in the supernatant, though the level of
208 secretion varied greatly among T11SS proteins. At one hour after induction, the presence
209 relative to the absence of the cognate T11SS OMP increased the average level of extracellular
210 hemophilin homolog protein ($HrpB_{X.nem}$/HrpC: 34.2-fold; $HrpB_{H.haem}$/Hpl: 59.3-fold, HsmA/HphA:
211 4.6-fold; CrpB/CrpC: 56.9-fold) (Fig. 3; Supplemental File 2). $HrpB_{X.nem}$ significantly increased
212 (4.8-fold) the average extracellular levels of the non-cognate cargo Hpl, though it was
213 significantly less effective at doing so than the cognate T11SS $HrpB_{H.haem}$. $HrpB_{X.nem}$ did not
214 significantly impact average extracellular levels of HphA, (1.1 fold), or CrpC (1.7 fold).

215

216 **The hemophilin C-terminal β-barrel domain contributes to specificity for T11SS**

217 The data described above indicates that individual hemophilin family cargo proteins and
218 their cognate T11SS OMPs have specificity for each other. Published literature has implicated
219 the C-terminal β-barrel domain of cargo proteins in directing secretion by T11SS OMPs [11]. To

220 assess the role of the hemophilin C-terminal domain on T11SS specificity, two chimeric
221 hemophilin cargo were engineered. The first chimeric cargo had the N-terminal handle domain
222 from HrpC and the C-terminal β-barrel domain from Hpl (henceforth HrpC-Hpl), while the
223 second had the N-terminal handle from Hpl and the C-terminal β-barrel domain from HrpC
224 (henceforth Hpl-HrpC). pETDuet-1 constructs were assembled to independently co-express
225 $HrpB_{H.haem}$ and $HrpB_{X.nem}$ alongside both chimeric cargo proteins (Fig. 4A) and Western blotting
226 of supernatant and cellular lysates was again performed to monitor cargo secretion
227 (Supplemental file 2: sheet 6). The HrpC-Hpl chimera ran as a single band at ~27 kD, which is
228 larger than the predicted size of ~25 kD. The Hpl-HrpC chimera ran as two bands with apparent
229 sizes of ~33 kD and ~24 kD, which straddle the expected size of ~27 kD. Since Hpl-HrpC ran as
230 two bands, similarly to Hpl, we opted to sum both bands for the purpose of enumeration. Co-
231 expression with $HrpB_{H.haem}$ increased on average, the extracellular levels of its cognate cargo
232 Hpl by 65.0-fold, the HrpC-Hpl chimera by 51.0-fold, and the Hpl-HrpC chimera by 7.8-fold,
233 indicating that the presence of the cognate C-terminal domain of Hpl was optimal for secretion
234 by $HrpB_{H.haem}$. Co-expression with $HrpB_{X.nem}$ increased on average, the extracellular levels of the
235 cognate HrpC by 15.6-fold, the HrpC-Hpl chimera by 5.1-fold, and the Hpl-HrpC chimera by 8.1-
236 fold on average (Fig. 4B; Supplemental File 2). Overall, both the presence of the C-terminal
237 domain of the cognate cargo, relative to the non-cognate cargo, resulted in higher levels of
238 secreted chimeric hemophilin homolog protein. This finding reinforces the concept that T11SS
239 cargo are two-domain proteins: An N-terminal effector domain and a C-terminal β-barrel
240 domain that directs cargo for T11SS-mediated secretion [28]. Our findings additionally implicate
241 the C-terminal β-barrel domain in mediating specificity of cargo for particular T11SS OMPs.

242

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

244 To begin to investigate if the heme ligand binding properties of Hpl and HphA are
245 conserved in other hemophilin homologs, Hpl from *H. haemolyticus*, HrpC from *X. nematophila*,
246 HphA from *A. baumannii*, and CrpC from *X. cabanillasii* were expressed without signal peptides
247 in the cytoplasm of *E. coli* and purified. As expected, Hpl from *H. haemolyticus* was recovered
248 from *E. coli* cytoplasm as an approximately 50:50 mix of heme-bound and heme-free protein,
249 with the level of heme saturation likely reflecting competition for heme binding and limitations
250 of heme biosynthesis *in vivo* [18]. Preparations of HrpC (*X. nematophila*) and HphA (*A.*
251 *baumannii*) had a brownish appearance and an absorbance peak at ~413 nm, possibly
252 indicating the presence of sub-saturating levels of a porphyrin ligand, whereas CrpC (*X.*
253 *cabanillasii)* was colorless with no peaks in the visible absorption spectrum, indicating the lack
254 of a porphyrin. Heme-free (apo-protein) preparations of Hpl, HphA, and HrpC were produced by
255 acid- acetone extract and reversed-phase HPLC, and heme-binding activities were investigated
256 by titration (Fig. 5; Table 1). Large changes in the UV-visible spectrum of hemin occurred upon
257 titration with *H. haemolyticus* Hpl, *A. baumannii* HphA or *X. nematophila* HrpC (Fig. 5A–C).
258 Similarities in the Soret (412–414 nm) and Q-band regions (500–600 nm) between HrpC and
259 HphA suggest that the heme coordination structure of HrpC is similar to that in HphA [13]. The
260 binding curves for Hpl, HphA, and HrpC (Fig. 6A) yielded K_d values of 9, 7 and 20 nM
261 respectively; however, the curves were close to linear, indicating that binding might be too
262 strong to reliably extract binding constants. Using simulated data with added Gaussian noise we
263 determined that data generated with K_d values < 15 nM produced fits that did not differ

264 significantly (at P ≤ 0.05 by F test) from fits where the K_d parameter was fixed at an extreme low
265 value. Thus, we suggest that Hpl and HphA bind heme with K_d values ≤ 15 nM and HrpC binds
266 with K_d = 20 (10–50) μM. In contrast, spectral changes upon addition of CrpC to heme were
267 more gradual (Fig. 5D) and were fit with a one-to-one binding model with apparent K_d ~5 μM
268 (Fig. 6A). These values are similar to the binding affinity we determined for BSA with apparent
269 K_d ~2 μM (Fig. S2).

270 To further distinguish between the porphyrin binding affinities of Hpl, HphA, and HrpC,
271 we assayed binding to Zn(II)PPIX, a fluorescent heme analog, in the presence of competition
272 from excess BSA (Fig. 7). We reasoned that fluorescence detection and competition binding
273 would shift the useful detection range to a high-affinity regime, compared to the previous
274 absorption design. These experiments (Fig. 7) yielded IC_{50} values for hemophilin binding Zn(II)-
275 PPIX in competition with BSA as shown in Table 1, together with calculated affinities based the
276 affinity of BSA for Zn(II)PPIX. The IC_{50} values for Hpl and HphA did not differ within error of the
277 measurements, suggesting that Hpl and HphA bind Zn(II)-PPIX with similar affinity. In
278 comparison, the IC_{50} value for *X. nematophila* HrpC was approximately 5–10 fold higher,
279 indicating weaker binding to Zn(II)-PPIX. Using the same fluorescence competition assay, the
280 interaction of CrpC with Zn(II)-PPIX was undetectable, consistent with the finding that CrpC and
281 BSA have similar propensities to bind heme. In summary, the above results suggest that heme
282 binding affinities of the hemophilin homologs proceed from higher to lower affinity in the order
283 Hpl ≈ HphA > HrpC >> CrpC.

284 To investigate the importance of the porphyrin metal to hemophilin binding, we
285 performed spectroscopic titrations with unmetalated PPIX. Because the CrpC gene is frequently

286 located close to cobalt/molybdenum associated genes, we also screened for binding to
287 Co(III)PPIX. Titration, monitored by UV-visible spectroscopy, indicated that Hpl, HphA and HrpC
288 bound PPIX with nM affinities, whereas CrpC binding was much weaker (K_d ~7 μM), in a pattern
289 (affinity HphA > Hpl ≈ HrpC >> CrpC) similar to that seen for a hemin ligand (Fig. 6B and S3). A
290 different pattern was seen with Co(III)-PPIX. Titration experiments with Co(III)-PPIX indicated
291 affinities ranked in the order Hpl > HphA ≈ HrpC ≈ CrpC, corresponding to a dramatic increase in
292 the relative affinity of CrpC for Co(III)PPIX compared to other porphyrins we tested (Fig. 6C and
293 S4). To understand the potential significance of this we looked at the interaction of BSA with
294 Co(III)-PPIX. We found that Co(III)-PPIX, and to a lesser degree Zn(II)-PPIX, bound more strongly
295 to BSA than did PPIX or hemin, suggesting that the Co(III) and Zn(II) metals might contribute to
296 binding interactions that are not available to Fe(III). We tested several other porphyrin-related
297 molecules, including coproporphyrin III, biliverdin and cobalamin and found no evidence for
298 CrpC binding (Fig. S5). These results suggest that Hpl, HphA, and HrpC might effectively
299 scavenge metalated or unmetalated porphyrins from the environment. *X. nematophila* HrpC
300 appeared to bind metalated and unmetalated porphyrin with similar affinities (10–20 nM). In
301 contrast, CrpC binds very weakly to PPIX and hemin. It is interesting to consider whether this
302 weak binding could allow CrpC to preferentially bind cobalt porphyrins in a biologically setting.
303

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

305 Our observations that hemophilin homologs vary in their binding capacities for
306 porphyrins prompted us next to investigate the structure function relationships among them.
307 The Hpl and HphA crystal structures [13,18] were compared against the protein structure

308 database (DALI server [29]) and manually inspected within the protein fold classification
309 databases (CATH [30] and SCOP [31]). We observed the expected C-terminal β-barrel domain
310 structure (Fig. 8A and 8C) that appears to be characteristic of T11SS cargo proteins (CATH
311 superfamily 2.40.160.90, SCOP superfamily 3002098) [18]. More limited similarity was detected
312 between the N-terminal ligand-binding domains of Hpl or HphA (Fig. 8B and 8D) and the N-
313 terminal handle domain of TbpB proteins (CATH superfamilies 2.40.128.240/250; SCOP fold
314 2001281). This region also appears to be characteristic of T11SS cargo proteins, including Hpl
315 and HphA, and forms a β-sheet (strands 1, 8, 7, 6, and 5 in Hpl and HphA) that packs against the
316 β-barrel domain. The remainder of the N-terminal region of the Hpl and HphA are variable in
317 structure compared to each other and the other known T11SS cargo proteins. Whilst the N-
318 terminal domains of Hpl and HphA share features with the handle domain of TbpB, the
319 insertion of one or more α-helical elements between β-strands 3 and 4, and a change in
320 hydrogen-bond connectivity within β-strands 1–4 suggest that the hemophilin ligand binding
321 domain belongs to a separate domain family, as a subgroup of the TbpB handle domain
322 topology.
323 We also considered the predicted structures of HrpC and CrpC as representatives of the
324 two other subclusters of hemophilin family proteins. Overall, sequence similarity between
325 hemophilins is higher in the β-barrel domain (40% identity between Hpl and HrpC; 36% identity
326 between Hpl and CrpC) it is lower in the N-terminal domain (24% identity between Hpl and
327 HrpC; 20% identity between Hpl and CrpC). We made structural models using the program
328 MODELLER [32] with the crystal structures of Hpl and HphA [13,18] as templates (see methods)

329 and found that HrpC (Fig. 8E-F) and CrpC (Fig. 8G-H) are predicted to adopt a similar overall
330 structure as Hpl and HphA.
331
332 **Hemophilin homologs share a signature N-terminal ligand binding domain**
333 We next compared features of the known or predicted N-terminal ligand binding
334 domain folds of the hemophilin subcluster representatives, including from the
335 Plant/Environmental subcluster. Alignments were used to model the structures of
336 representative members of each subcluster, including *X. nematophila* HrpC (Fig. 8E-F) and *X.*
337 *cabanillasii* CrpC (Fig. 8G-H) for which we had measured ligand binding. The ligand-binding
338 domains of Hpl and HphA have a conserved hydrophobic core that was preserved in alignments
339 with HrpC and CrpC, as well as other members of each subcluster, including the
340 Plant/Environmental subcluster (e.g., *Lysobacter enzymogenes* ALN55974 and
341 *Stenotrophomonas maltophila* KUJ02124) (Fig. 9). We also looked for conserved pair-wise
342 contacts between side chains that are distant in the primary sequence and that might provide
343 additional confidence in modelling local structures. For example, we identified a DXNG[V/I]
344 motif corresponding to a β-hairpin in the ligand binding domain of *H. haemolyticus* Hpl that
345 makes a bifurcated hydrogen bond with a Tyr side chain in the β-barrel domain (Fig. 10A) and is
346 conserved in Hpl-like, HrpC-like and CrpC-like sequences, but is absent in the HphA-like cluster
347 (Fig. 9). With respect to this motif, the Plant/Environmental subcluster comprises two subsets,
348 one containing the DXNG motif (represented by ALN55974 from *L. enzymogenes*), and one
349 lacking the DXNG motif and with a divergent sequence/structure in β4-β5 hairpin (represented
350 by KUJ02124 from *S. maltophilia*) Fig. 9). Several groups of sequences in the Hpl-like subcluster

351 could not be aligned or modelled against the hemophilin domain due to greater sequence (and
352 therefore presumably structural) diversity) within this subcluster. With the exception of this
353 group, our alignment and structural motif analyses indicate that hemophilin homologs (e.g.,
354 HrpC, CrpC, HphA, Plant/Environmental subclusters, and a subset of the Hpl subcluster) have an
355 N-terminal domain with the same topology as the ligand-binding domains of *H. haemolyticus*
356 Hpl and *A. baumannii* HphA and that this hemophilin ligand binding domain is the signature
357 feature of an extensive family of hemophilin homologs.
358
359 **Two distinct heme binding domains occur among hemophilin homologs**
360 Having identified another hemophilin homolog capable of coordinating heme, we used
361 models to investigate possible structural reasons for common and distinct ligand binding
362 properties of hemophilin homologs. The available crystal structures of *H. haemolyticus* Hpl and
363 *A. baumannii* HphA, both of which bind heme, show that residues in loops β2-β3, β3-β4
364 (including the β3-β4 α-helix) and β5-β6 contribute majority of heme-binding contacts and that
365 the mode of heme binding in Hpl and HphA differs in the relative contributions of these loops
366 and in the heme-iron coordination structure (Fig. 8B and 8D). Of particular interest, heme in Hpl
367 is coordinated through a single His side chain in the β5-β6 loop, whereas heme in HphA is
368 coordinated by two His side chains, one from the β5-β6 loop and a second donated by the β2-
369 β3 loop.
370 The bis-histidyl heme coordination observed for HphA appears to be ubiquitous in the
371 HphA subcluster, based on the conservation of the amino acid sequence of this motif (data not
372 shown). A subset of sequences within the Hpl subcluster had >65% identity to *H. haemolyticus*

373 Hpl, and these are predicted to all share the same single histidinyl coordination heme binding
374 mode. However, the heme coordination pattern for the remaining members of the Hpl
375 subcluster is unclear due to the high level of sequence/structure diversity in this group.
376 Molecular models show that HrpC subcluster and Plant/Environmental subcluster proteins
377 contain His side chains in both the β2-β3 loop and β5-β6 loops, and modelling in the presence
378 of a heme ligand indicates that the position of these His residues is compatible with bis-histidyl
379 heme coordination similar to HphA (Fig. 8F and Fig. S6). Other structural similarities among
380 HphA, HrpC, and Plant/Environmental subclusters, but not for the Hpl subcluster, exist in the
381 β2-β3 loop, including a Gly, Ser, Pro rich sequence, and similar loop length (Fig. 9). The
382 mechanism of heme binding and release by *A. baumannii* HphA is proposed to involve
383 unfolding of the β2-β3 loop [13], and if so, this mechanism is likely to be conserved in
384 hemophilin homologs with this motif. The second heme-coordinating His of HphA occurs in the
385 β5-β6 loop, which contains a conserved PX[S/T]H sequence motif.
386 Like the β2-β3 loop, the PX[S/T]H motif of the β5-β6 loop is conserved across the HphA,
387 HrpC and Plant/Environmental clusters (Fig. 9, Fig. 10B). In *A. baumannii* HphA, the His side
388 chain of this motif is packed against the pyrrolidine ring of Pro and makes a H-bond with the Ser
389 side chain (substitution of Thr for Ser is expected to preserve this interaction). The carbonyl
390 oxygen of Pro also accepts a H-bond from the backbone amide of His. These interactions are
391 expected to stabilize the orientation of the heme-ligating His, and conservation of this motif in
392 HrpC and Plant/Environmental cluster hemophilin homologs further implies that the mode of
393 heme binding in HphA, HrpC and Plant/Environmental proteins is similar. The PX[S/T]H motif is

394 absent in Hpl, which has a substantially longer β5-β6 loop (20 residues in Hpl vs 14 residues in
395 HphA) and more completely covers the face of the porphyrin.
396
397 **The N-terminal ligand binding domain regions of hemophilin homologs are diversifying**
398 In addition to iron-ligating side chains, other amino acids contribute to ligand binding
399 through non-covalent interactions with apolar and polar sites on the porphyrin. Models of *X.*
400 *nematophila* HrpC show multiple hydrophobic contacts with the porphyrin ring contributed by
401 side chains in the β3-β4 α-helix, β5-β6 loop and underlying β sheet, and a hydrogen bond
402 network around the charged porphyrin 17-propionate comprising Tyr64, Ser96 and His97 side
403 chains – features similar to the HphA heme site (Fig. 8, compare D, F). Differences include a
404 shorter β3-β4 α-helix in HrpC than either HphA or Hpl, and a β4-β5 hairpin in HrpC that is more
405 similar to Hpl than to HphA. The pyrrole rings B and C appear to be more exposed to solvent in
406 HrpC than in either Hpl or HphA, which may partly explain a relatively lower affinity for the
407 metalated porphyrins, hemin and Zn(II)PPIX, because hydration is expected to increases the
408 rate of scission of the Fe-metal bond (although the lack of effect on Co(III)PPIX binding is not
409 explained).
410 Our porphyrin binding assays revealed that *X. cabanillasii* CrpC does not bind heme.
411 Consistent with this observation, while *X. cabanillasii* CrpC is predicted to adopt the hemophilin
412 ligand binding domain fold, it lacks His or other residues that typically coordinate heme-iron, in
413 either the β2-β3 or β5-β6 loops (Fig. 8G-H and Fig. S6). The β2-3 loop is reduced to a three-
414 residue hairpin and is too small to contribute to a porphyrin pocket of the kind seen in HphA
415 and HrpC. Neither does CrpC look like the Hpl heme-binding site, which employs a much longer

416 β2-β3 α-helix and longer β5-β6 loop. Nevertheless, the distribution of apolar and polar side
417 chains in Hpl/HphA that contact the porphyrin skeleton and ionizable propionates, respectively,
418 is largely conserved in CrpC, although we cannot model the precise conformation of the β5-6
419 loop in CrpC due to lack of a suitable template. Thus, CrpC could potentially accommodate a
420 porphyrin-like ligand in a more solvent exposed, and therefore presumably lower affinity,
421 binding site. This *X. cabanillasii* distinct non-histidine ligand binding domain is shared among
422 the other CrpC-like subcluster homologs, except for a small number of sequences that have a
423 large deletion in the hemophilin domain that removes the porphyrin binding site altogether.
424
425 **Discussion:**
426 A cornerstone of microbial existence is the extracellular deployment of metal-binding
427 molecules, such as hemophilins which facilitate the competitive uptake of metals essential for
428 cellular physiology. Here we investigated the structural features, secretion, and ligand binding
429 of representative hemophilins, a family of proteins in five distinct sequence similarity
430 subclusters encoded by Proteobacteria from human, animal, plant, built and free-living
431 environments. We found that the entire hemophilin family shares common motifs present in
432 the solved structures of the hemophilins *H. haemolyticus* Hpl and *A. baumannii* HphA [13,18]: a
433 central β-sheet domain and a C-terminal β-barrel domain, both of which are previously
434 described characteristics of T11SS cargo proteins [11,12]. In contrast, the N-terminal ligand
435 binding region was conserved in some, but not all hemophilin homologs. We established that,
436 like the hemophilins *X. nematophila* HrpC and *A. baumannii* HphA [12,33,34], two other
437 hemophilin homologs, *H. haemolyticus* Hpl and *X. cabanillasii* CrpC rely on a T11SS secretor to

438 reach the extracellular milieu. This establishes T11SS-dependence across all four hemophilin
439 sequence subclusters tested, and indicates T11SS-dependent secretion is a hallmark of the
440 entire family. Further, the results presented here support the concept that the β-sheet-β-barrel
441 domains are an ancestral scaffold for T11SS-dependent secretion, onto which new effector
442 variants are emerging through evolution of the N-terminal domain.

443 Despite the overall conservation of the core β-sheet-β-barrel motif among T11SS-
444 dependent cargo, there must be inherent differences among these domains, since different
445 cargo proteins show specificity for different T11SS secretors. To explore the underlying
446 contributors to such specificity, we capitalized on the fact that each hemophilin homolog is
447 associated with a cognate T11SS secretor, enabling different combinations of related cargo-
448 secretor pairs to be assessed for secretion activity. We found that, despite sequence, structural,
449 and functional similarity among hemophilin homologs and among their cognate T11SS
450 secretors, individual cargo-secretor pairs have specificity for each other, indicating co-
451 adaptation and co-diversification. These results establish that T11SS specificity is not shared
452 between cargo, even between closely-related homologs, and that the efficacy or rate of
453 secretion varies between T11SS/cargo pairings. By testing the secretion of hemophilin chimeras
454 with non-native C-terminal β-barrel domains we determined that this domain is largely
455 responsible for specificity of targeting to its cognate T11SS secretor. The effectiveness of the
456 tested secretors (*H. haemolyticus* HrpB and *X. nematophila* HrpB) was highest for the cognate
457 hemophilin cargo protein, followed by the chimeric hemophilin with the cognate C-terminal β-
458 barrel domain. The chimeric hemophilin with the non-cognate C-terminal β-barrel domain was
459 not secreted. Chimera secretion did not reach the same levels as native hemophilin cargo

460 protein-T11SS secretor pairings, which may indicate that the conserved β-sheet, which was not
461 substituted within the chimeric proteins, might also contribute to specificity. Alternatively, it
462 may indicate that the chimeric proteins generated have compromised stability relative to the
463 native cargo sequences that slows transposition or speeds up protein degradation.

464 The ability of the tested T11SS secretors to distinguish their cognate C-terminal domain
465 is striking given that their cargo proteins, Hpl and HrpC are closely related in sequence (39%
466 identity) and function, and reveals a previously underappreciated resolution of T11SS cargo
467 selection. Further, we show here that when expressed in a common *E. coli* background under
468 identical conditions, the T11SS secretors vary in their magnitude or efficacy. For example, HrpB
469 increased extracellular Hpl by 59.3-fold while HsmA increased extracellular HphA 4.6-fold.
470 These aspects of secretion specificity and efficacy are important to consider for potential T11SS
471 bioengineering applications. For instance, our findings raise the possibility that novel T11SS-
472 dependent cargo could be engineered for specific secretion by fusing novel N-terminal
473 functional domains to the C-terminal domain from a characterized cargo protein. Similarly,
474 Slam proteins, the subset of T11SS responsible for surface exposure of lipoproteins, have been
475 proposed as a novel mechanism for surface presentation of immunogenic antigens as a
476 potential vaccination strategy [35], and our discovery that different T11SS may have different
477 levels of secretion efficacy could be exploited to fine-tune the levels of surface exposure or
478 secretion. An important next step towards achieving this goal will be to identify the individual
479 sequence and structural motifs responsible for the specificity between cargo and T11SS and any
480 T11SS-dependent cargo modifications that may occur, to enable automatic annotation of T11SS
481 cognate pairs and informed engineering of novel pairings.

482 The results presented here provide a comparative framework in which to consider the
483 evolution and relatedness of T11SS cargo protein structures, which can be described as
484 comprising three domains: an N-terminal variable effector domain, a mid-protein β-sheet
485 domain, and a C-terminal β-barrel domain. The C-terminal β-barrel domain, comprising 8 β-
486 strands arranged in a meander topology with a sheer value of 8, and a central core of
487 hydrophobic side chains [18], is a unique and defining feature of T11SS cargo proteins that, as
488 we have established here, contributes substantially to the recognition of cargo by the cognate
489 T11SS secretor. The β-sheet domain is also universally present among known T11SS cargo. It is
490 formed in large part by amino acids in the central portion of the protein sequence, and stacks
491 against the β-barrel forming a physical barrier and bridge between it and the variable N-
492 terminal effector domain. In some cargo, including hemophilins, there are integrated structural
493 interactions between the variable N-terminal region and the β-sheet, with residues in the
494 former contributing one or more β-strands to the β-sheet, and residues in the latter
495 contributing key active site residues.

496 In the case of hemophilins, there is a conserved β-strand topology across the entire
497 protein, comprising the effector strands (β2-β4), β-sheet strands (β1 and β5-8) and β-barrel
498 strands (β9-18). Within this topology, a conserved hydrophobic core is formed by residues in
499 the β1-β7 region, as are interfacial contacts between the β-sheet and β-barrel domains. Indeed,
500 the most highly conserved sequence motif in hemophilins is a YFGEW pentapeptide that maps
501 to β7 and contributes side chains to both the hemophilin hydrophobic core and to interdomain
502 contacts. The relatedness of the hemophilins with other metal-related T11SS cargo is supported
503 by the fact that this conserved hemophilin ligand binding domain (Hlb) shares a parent fold

504 (SCOP fold 2001281) with the TbpB handle domain (SCOP family 4006246; TbpB handle
505 domain-like) and with HpuA (SCOP family 4005058, hemoglobin receptor HphA). However, it is
506 distinguished from these by differences in connectivity in the β1-β3 strands and inclusion of
507 one or more α-helical elements in the β3-β4 loop. These results imply that β-strand topology
508 and the hydrophobic core of the Hlb comprise an ancestral scaffold which has been elaborated
509 on in evolution to bind a diverse range of ligands: heme in the case of Hpl, HphA, and HrpC, a
510 variety of protein targets in the cases of HpuA, TbpB, and fHbp, and unknown ligands in the
511 case of CrpC and NilC.

512 When considering active site residues and topology, the hemophilin homologs fell into
513 three general classes: the Hpl type with a single histidinyl coordination binding site, the
514 HphA/HrpC type, which included representatives from the Plant/Environmental subcluster,
515 with a bi-histidinyl coordination binding site, and the CrpC type which lacked any histidines in
516 the predicted active site. The *A. baumannii* HphA might be considered closer to a canonical
517 heme-binding structure, whereas the Hpl-like subcluster appears to be more diverse in
518 structure and arrangement of heme ligands.

519 In the HphA/HrpC type class, the heme is coordinated between two histidines, one in
520 the β2-β3 loop and a second in the β5-β6 loop. In all members of this binding class, the β2-β3
521 loop is rich in Gly, Pro, and polar residues. This low complexity sequence suggests that the
522 conformational heterogeneity that is seen in HphA may be a general feature; in heme-bound
523 HphA the loop and heme-coordinating His cover one face of the porphyrin, whereas the same
524 loop points out away from the binding site in the apo protein [13]. The structure of the β5-β6
525 loop does not differ between holo and apo structures of HphA and is predicted to be conserved

526 across the HphA, HrpC and Plant/Environmental subclusters; a characteristic PX[S/T]H motif
527 that features the second heme-ligating His side is stabilized by π-stacking and H-bonding
528 interactions. A search of the protein database using RASMOT-3D [36] identified only one other
529 heme binding protein (nine-heme cytochrome c, pdb 19hc) with a similar 3D arrangement of
530 Pro, Ser, and heme-coordinating His side chains, but different primary sequence arrangement,
531 thus the PX[S/T]H appears to be specific feature of the hemophilin domain. The modelled CrpC
532 structure is highly similar to HrpC except that the β2-β3 loop is reduced to a hairpin and
533 PX[S/T]H motif is gone, suggesting an evolutionary trajectory for loss of heme binding in CrpC.
534 Whereas *H. haemolyticus* Hpl clearly adopts the same topology as HphA, HrpC,
535 Plant/Environmental, and CrpC subcluster representatives (with addition of a second helical
536 segment in the β3-β4 loop of Hpl), the Hpl subcluster exhibits more extensive sequence (and
537 likely structural) variability. One common feature within the Hpl subcluster is a D[R/K/S]NGV
538 motif in the β4-β5 hairpin that is predicted to make conserved interactions at the interface
539 between the β-sheet and the β-barrel domains. A similar motif appears in HrpC, CrpC, and a
540 subset of the Plant/Environmental subcluster, but it is absent in HphA and the remaining
541 Plant/Environmental cluster sequences, indicating that some structural features do not
542 conform to a simple division between HphA-like and Hpl-like sequences. The heme-associated
543 α-helix in both HphA and Hpl is longer than predicted in HrpC and Plant/Environmental cluster
544 proteins, possibly contributing to higher heme affinity in HphA/Hpl by protecting a larger
545 surface area of the porphyrin from solvent. Although most sequences within the Hpl subcluster
546 carry His in the putative β5-6 loop, one group (sequence OFR67839 is one example) lacks this
547 residue and other subgroups potentially have two His ligands.

548 The distinguishing active site characteristics among the hemophilin homologs
549 contextualized our findings that purified hemophilin homologs displayed variable affinity for
550 hemin (Fig. 5; Table 1). Intriguingly, this variable affinity for hemin may reflect differences in the
551 ecological roles of the hemophilin homologs. Hpl and HphA, which have established roles in
552 competing for heme under stringent conditions, had the highest heme affinity. *H. haemolyticus*
553 Hpl is a heme chelating protein for nutritional heme uptake and can prevent the related
554 organism, *H. influenzae*, from accessing Plant/Environmental heme [18]. This role for *H.*
555 *haemolyticus* Hpl as a tool for competing against other organisms by sequestering a limiting
556 nutrient (a form of nutritional immunity) would naturally favor the evolution of high hemin
557 affinity. In a similar vein, *A. baumannii* HphA competes with mammalian host nutritional
558 immunity for heme. In contrast to these two hemophilins, HrpC, encoded by *X. nematophila*,
559 displayed lower heme binding affinity. This may be a reflection of the distinctive animal host
560 niches occupied by *X. nematophila* relative to *H. haemolyticus* and *A. baumannii*. As a
561 pathogen, *Xenorhabdus* can infect diverse insects, which are notoriously heme poor [37] and as
562 a mutualist, it colonizes the intestinal tissues of its nematode host, which is a heme auxotroph
563 [38,39]. The insect *Drosophila melanogaster* transferrin-1 binds heme with lower affinity than
564 mammalian transferrin and is susceptible to low pH conditions [40]. Therefore, to access heme
565 in an insect environment, *X. nematophila* HrpC may not need high affinity to overcome host
566 chelation. In turn, lower affinity may offer a selective advantage by enabling resource sharing
567 with its mutualistic nematode host, *S. carpocapsae,* with the resulting improved fitness
568 benefitting both mutualistic partners. Dissociation of heme from HrpC would be essential for
569 heme to be transferred to nematode iron chelating proteins. Entomopathogenic nematodes

570 can live for months at a time without feeding while in their free-living infective stage. During
571 this time they have a sealed intestine and no access to exogenous nutrients and they may rely
572 on their *Xenorhabdus* symbionts to provide heme chaperoned by hemophilin [41]. While this
573 study did not test the heme-binding affinity of hemophilin homologs from the
574 Plant/Environmental subcluster, these homologs encode His residues in the β2-β3 loop and the
575 β5-β6 loop like HphA and HrpC (Fig. S6), indicating the potential for heme-binding activity.
576 Future studies on the binding affinities of these homologs may shed light on additional
577 functions of hemophilins and their roles in facilitating bacterial lifestyles within plants, built
578 environments, and soil ecosystems.

579 Our work has revealed that not all hemophilins have heme-binding activity, which opens
580 the possibility that this family is diversifying and potentially gaining new ligand binding
581 activities. CrpC, a member of the Cobalt/molybdenum associated subcluster of hemophilin
582 homologs, did not bind heme, consistent with the fact that the predicted active site had neither
583 the heme-coordinating His residues noted for the other hemophilin classes, nor any other
584 typical heme ligands such as Cys, Met, or Tyr. There are residue types in the CrpC β5-β6 loop
585 (Asn, Gln, Glu) and β3-β4 α-helix (Lys), that conceivably could act as metal ligands; a small
586 number of protein crystal structures show Asn and Gln as a second heme ligand together with
587 His [42] or Gln [43], respectively, and a Glu side chain [44] is in within coordinate bonding
588 distance of heme iron in the terminal cytochrome bd oxidase, though it does not contribute
589 substantially to heme binding affinity [45]. Lys residues occur in several algal globins, but only
590 as a second Fe-ligand together with His [46]. However, none of these possibilities fit with the
591 predicted structure of CrpC, which does not have His or Cys residues close to the β5-β6 loop.

592 Thus the absence of a strong heme-iron ligand provides a structural explanation for the failure
593 of CrpC to bind to heme or Zn(II)PPIX.
594 In this context, the binding of Co(III)PPIX to CrpC with comparable affinity to other
595 hemophilins was unexpected. However, it is intriguing given the genomic context of the genes
596 encoding CrpC and its T11SS partner, CrpB, which are found adjacent to an anaerobic
597 B12/cobalamin biosynthesis locus in at least three strains of *Xenorhabdus* (Fig. 2B;
598 Supplemental file 2), which hints at a role for the CrpC/CrpB cargo/secretion pair in cobalt or
599 cobalamin binding. We found that the affinity of BSA for Co(III)PPIX was similarly about 2 orders
600 of magnitude stronger than binding to unmetalated PPIX suggesting that Co(III)PPIX makes
601 useful contacts with a wider range of side chains, perhaps those present in the CrpC active site,
602 than does either heme or Zn(II)PPIX. CrpC binds to Co(III)PPIX with similar affinity to other
603 hemophilin homologues, but has a dramatically lower affinity for heme and PPIX compared to
604 other hemophilin homologues. Based on these data, we suggest that CrpC has evolved to lose a
605 heme binding function while maintaining the ability to bind an alternative cobalt-metalated
606 porphyrin-related ligand, such as an intermediate in cobalamin biosynthesis or metabolism.
607 Overall, our findings suggest that bacteria can encode multiple members of the hemophilin
608 family, both bona-fide hemophilins, such as HrpC, and paralogs like CrpC, that might have
609 evolved new ligand binding activities by divesting themselves of their affinity for heme. An
610 exciting avenue for the discovery of new metal-, or other ligand-binding proteins will be to
611 examine the activities of additional members of the hemophilin family of proteins.
612
613 **Materials and Methods:**

614 **Sequence similarity network analysis**

615 Protein sequence similarity networks were generated using the Enzyme Function
616 Initiative-Enzyme Similarity Tool (EFI-EST) [22]. As an input we used the previously reported
617 database of soluble TbpBBD cargo encoding genes which co-occurred with T11SS encoding
618 genes [12]. EFI-EST performs an all-by-all BLAST of query sequences to assess relatedness, and
619 then generates a network where each node represents a protein sequence, and the color of the
620 edges indicates relatedness between nodes. A minimum alignment score of 35 was chosen to
621 reduce total network edges enough to visualize protein subclusters. To simplify visualization
622 proteins sharing ≥ 80% identity were compressed into representative nodes. Networks were
623 visualized and interpreted using Cytoscape v3.7.1 [47] and Gephi v0.9.5 [48]. For the complete
624 network of soluble TbpBBD domain proteins, nodes were organized with the Fruchterman-
625 Reingold force-directed algorithm [49]. For the network containing only hemophilin family
626 proteins, nodes were organized with the ForceAtlas2 algorithm for continuous force-directed
627 arrangement [50].

628

629 **Bacterial culture conditions**

630 All strains, plasmids, and primers utilized in this study are described in supplemental file
631 3. All cultures were grown in either glucose minimal media [33], LB stored in the dark to
632 prevent formation of oxidative radicals (henceforth dark LB), or glucose minimal media
633 supplemented with 1% dark LB. Plate based cultures were grown on either LB supplemented
634 with pyruvate to prevent the formation of reactive oxygen radicals (henceforth LBP or glucose
635 minimal plates [33]. For plasmid-based expression, chemically competent *Escherichia coli* strain

636 BL21-DE3 (C43) were chosen for ease of transformation and their ability to non-toxically
637 express membrane proteins [51,52]. Strains of *Escherichia coli* were grown at 37°C. Where
638 appropriate, media was supplemented with ampicillin at a concentration of 150 µg/ml,
639 chloramphenicol at 15 µg/ml or kanamycin at 50 µg/ml unless another concentration is stated.
640 Protein expression was induced at the midlog point of bacterial growth via addition of isopropyl
641 β-d-1-thiogalactopyranoside (henceforth IPTG) at a concentration of 0.5mM.
642
643 **Construction of T11SS and T11SS-dependent cargo expression plasmids**
644 Expression plasmids for HrpC alone (HGB2531) and HrpBC co-expression (HGB2530)
645 were previously generated and reported on [12]. FLAG-*hrpB* was amplified from pETDuet-
646 1/$hrpBC_{X.nem}$ using primers 1-2. *hpl*-FLAG and its adjacently encoded T11SS neighbor, FLAG-
647 $hrpB_{H.haem,}$ were amplified from the purified genome of *Haemophilus haemolyticus* BW1 using
648 primers 3-6. *crpC*-FLAG and its adjacently encoded T11SS neighbor, FLAG-*crpB*, were amplified
649 from the purified genome of *Xenorhabdus cabanillasii* (HGB2490) using primers 7-10. *hphA*
650 (ACJ40780.1) and *hsmA* (ACJ40781.1) were generated via gene synthesis by Genscript. To make
651 cargo-only expression plasmids, pETDuet-1, *hpl*-FLAG, *crpC*-FLAG, and *hphA*-FLAG were digested
652 with NcoI and NotI. Each cargo protein was independently ligated into MCS1 of pETDuet-1 via
653 T4 DNA ligase, resulting in pETDuet-1/*hpl* (HGB2526), pETDuet-1/*crpC* (HGB2525), and
654 pETDuet-1/*hphA* (HGB2532). Integration of each T11SS-dependent cargo was confirmed via
655 digestion with NcoI and NotI as well as Sanger sequencing using primers 11-12 at the University
656 of Tennessee (UT) Genomics Core.

657 To make T11SS/cargo co-expression plasmids, each of the above cargo-only expression
658 plasmids were digested with KpnI and NdeI, alongside the PCR products for FLAG-*hrpB*$_{H.haem,}$
659 FLAG-*crpB*, and FLAG-*hsmA*. Each T11SS protein was then independently ligated into MCS2 of
660 the plasmid containing its cognate cargo via T4 DNA ligase, resulting in pETDuet-
661 1/*hpl*/*hrpB*$_{H.haem}$ (HGB2523), pETDuet-1/*crpC*/*crpB* (HGB2524), and pETDuet-1/*hphA*/*hsmA*
662 (HGB2533). Additionally, the PCR product for FLAG-*hrpB* from *X. nematophila* was digested with
663 KpnI and NdeI and ligated into MCS2 of all the cargo-only expression plasmids, resulting in
664 pETDuet-1/*hpl*/*hrpB*$_{X.nem}$ (HGB2529), pETDuet-1/*crpC*/*hrpB*$_{X.nem}$ (HGB2528), and pETDuet-
665 1/*hphA*/*hrpB*$_{X.nem}$ (HGB2527). Integration of each T11SS protein was confirmed via digestion
666 with KpnI and NdeI as well as Sanger sequencing using primer 13 the at the University of
667 Tennessee (UT) Genomics Core.
668 To construct chimeric hemophilin homologs, *hrpC* and *hpl* were split into two domains
669 based on multiple sequence alignment and the NCBI conserved domain database. *hrpC* was
670 split between position 402 and 403, while *hpl* was split between position 474 and 475. Primers
671 14-15 were used to amplify the hemophilin handle domain from pETDuet-1/*hrpC*/*hrpB*$_{X.nem}$
672 (HGB2530). Primers 16-17 were used to amplify *hrpB* and the hemophilin β-barrel domain from
673 pETDuet-1/*hrpB*$_{X.nem}$/*hpl* (HGB2529). These two products were assembled into pETDuet-
674 1/Chimeric hemophilin(*hrpC*-*hpl*)/*hrpB*$_{X.nem}$ (HGB2595). Primers 18-19 were used to amplify the
675 hemophilin handle domain from pETDuet-1/*hrpB*$_{X.nem}$/*hpl* (HGB2529). To amplify *hrpB*$_{X.nem}$ and
676 the hemophilin β-barrel domain from pETDuet-1/*hrpC*/*hrpB*$_{X.nem}$ (HGB2530). These two
677 products were assembled into pETDuet-1/Chimeric hemophilin(*hpl*-*hrpC*)/*hrpB*$_{X.nem}$ (HGB2596).
678 Primers 14-15 were used to amplify the hemophilin handle domain from pETDuet-

679 1/*hrpC*/*hrpB*$_{X.nem}$ (HGB2530). Primers 16-17 were used to amplify *hrpB*$_{H.haem}$ and the hemophilin
680 β-barrel domain from pETDuet-1/*hpl*/*hrpB*$_{H.haem}$ (HGB2523). These two products were
681 assembled into pETDuet-1/ Chimeric hemophilin (*hrpC-hpl*)/*hrpB*$_{H.haem}$ (HGB2597). Finally,
682 pETDuet-1/Chimeric hemophilin (*hpl-hrpC*)/*hrpB*$_{X.nem}$ (HGB2596) and pETDuet-1/*hpl*/*hrpB*$_{H.haem}$
683 (HGB2523) were digested with NotI and NcoI, liberating the hemophilin homolog from each
684 vector. The chimeric hemophilin from HGB2596 was isolated via gel electrophoresis and then
685 ligated into MCS1 of the vector isolated from HGB2523, resulting in pETDuet-1/Chimeric
686 hemophilin (*hpl-hrpC*)/*hrpB*$_{H.haem}$ (HGB2598). Integration of each T11SS protein and chimeric
687 cargo was confirmed via Sanger sequencing using primers 12-13 at the University of Tennessee
688 (UT) Genomics Core.
689
690 **Protein expression and Immunoblotting**
691 *E. coli* strains used for expression experiments were taken fresh from storage at -80°C
692 for each experiment. Strains were cultured on glucose minimal media plates + ampicillin
693 overnight. For each biological replicate, 10 colonies were pooled and inoculated into 5ml of
694 fresh minimal media glucose + ampicillin broth and incubated rotating overnight. Each replicate
695 of each strain was rinsed 2x in PBS and normalized to an OD_{600} of 0.05 in 60ml of glucose
696 minimal media + 1% LB + ampicillin. These were grown shaking at 225rpm until they reached
697 mid log growth ($OD_{600} \approx 1$), typically between 5 and 8.5 hours. Upon reaching midlog growth,
698 25mls of each culture was removed and used as an uninduced T0 control. The remaining 35mls
699 were supplemented with IPTG to a concentration of 0.5mM. One milliliter supernatant samples
700 were taken at 1 and 2.5 hour(s) post induction. Supernatant samples were clarified via

701 centrifugation and filter sterilized. At 2.5 hours post induction the remaining cultures were
702 concentrated via centrifugation, rinsed 2x in PBS, and lysed via sonication (30 s at ~500-rms
703 volts). Supernatant samples and cellular lysate samples were supplemented with PMSF (1.7
704 µg/ml), Leupeptin (4.75 µg/ml), and Pepstatin A (0.69 µg/ml) to inhibit proteinase activity. The
705 no plasmid control was performed identically except without the presence of ampicillin in the
706 media.

707 The protein concentration of cellular lysates was normalized via the Pierce 660nm
708 Protein Assay (REF22660). For supernatant samples 600µl of each filtered sample was
709 precipitated via 10% Trichloroacetic acid precipitation as previously described [12,53]. Samples
710 were boiled for 10-25 min. prior to performing SDS-PAGE to ensure complete unfolding in the
711 protein sample buffer. SDS-PAGE was performed in duplicate using 10% polyacrylamide gels.
712 The first gel was used to perform Coomassie staining for total protein content while the second
713 gel was transferred to a PVDF membrane for Western immunoblotting. Immunoblots were
714 incubated in 50% Ly-cor blocking buffer:50% Tris buffered saline (TBS) for 1 hour to block.
715 Immunoblots were then incubated in 50% Ly-cor blocking buffer:50% TBS supplemented 0.1%
716 Tween20 and 1:5000 rat α-FLAG antibody for 1 hour. Subsequently the blots were incubated in
717 50% Ly-cor blocking buffer:50% TBS supplemented 0.1% Tween20 and 1:5000 goat α-rat
718 antibody bound to a 680CW fluorophore for 1 hour. Finally, immunoblots were visualized using
719 a Li-cor odyssey imaging the 700nm wavelength. The intensity of supernatant samples was
720 normalized to a clearly visible, non-target protein band in the Coomassie stain to control for
721 protein concentration. Efficacy of secretion was measured as the fold change of cargo protein
722 present in the supernatant when co-expressed with a T11SS protein relative to cargo protein

723 present in the supernatant when expressed alone. Fold changes were not normally distributed
724 initially, so they were $\log_{10}$ transformed prior to analysis. The resulting data were analyzed via a
725 one-way ANOVA and a Tukey's HSD Test [54].
726
727 **Purification of hemophilin homologs**
728 Hemophilin from *H. haemolyticus* was expressed and purified as previously described to
729 yield low and high heme-content fractions after anion exchange chromatography [18]. Heme
730 was removed by cold acid acetone treatment to yield an apo hemophilin fraction, as previously
731 described [55]. Residual Fe(III) heme was estimated at 1.8% of sites, based on extinction
732 coefficients of met-hemophilin being 96,100 M^{-1} cm^{-1} and 38,600 M^{-1} cm^{-1} at 414 nm and 280
733 nm, respectively, and extinction coefficient of the apo-protein being 25,900 M^{-1} cm^{-1} at 280
734 nm.
735 Expression constructs encoding the hemophilin homologs from *X. nematophila* (residues
736 23–247), *X. cabanillasii* (residues 23–238) and *A. baumannii* (residues 21–264) were
737 constructed in pET28a. In each case, the native N-terminal signal peptide was omitted and
738 replaced with a hexa-histidine tag and engineered tobacco etch virus (TEV) protease cleavage
739 site. Clones were transformed into *E. coli* strain Rossetta-2 (Novagen), grown in LB containing
740 34 µg/ml chloramphenicol and 25 µg/ml kanamycin; expression was induced with 1 mM IPTG
741 for 3 hours shaking at 37°C. Cells were suspended in lysis buffer (0.5 M NaCl, 0.05 M sodium
742 phosphate, 0.02 M imidazole, 100 µM phenylmethylsulfonyl fluoride, pH 7.2) and lysed by
743 sonication (Branson), then hemophilin homologs were captured by Ni-affinity chromatography.
744 TEV protease was expressed and purified as described [56]. The His-tag was cleaved from

745 hemophilin homologs by TEV protease treatment overnight at room temperature, to liberate
746 hemophilin proteins with an additional N-terminal Gly-His-Met tripeptide residual from the TEV
747 cleavage site. TEV protease and His-tag peptides were removed over a second Ni-affinity
748 column. Hemophilin preparations from *X. nematophila* and *A. baumannii* had a brownish
749 appearance and an absorbance peak at ~413 nm characteristic of a porphyrin ligand, as well as
750 less intense absorption peaks at 533 and 659 nm. Ligand was estimated to occupy ~25% of sites
751 based on comparison with spectra of hemophilin from *H. haemolyticus*. In contrast, CrpC from
752 *X. cabanillasii* was colorless. Acid acetone or methyl ethyl ketone extraction was not effective
753 to remove colored contaminants from HrpC of *X. nematophila* or *A. baumannii*. Apo-protein
754 fractions of these proteins were prepared by reversed-phase HPLC over a C4 stationary phase
755 (Waters Symmetry) developed with a CH_3CN:water mobile phase gradient containing 0.1%
756 trifluoroacetic acid. Solvent was removed by lyophilisation. Apo-CrpC from *X. cabanillasii* was
757 applied to a strong anion exchange resin (Q sepharose, Pharmacia) in 25 mM Tris-HCl buffer (pH
758 8.25 at 21°C) and collected in the flow-through. All apo-proteins were dialyzed into 20 mM Tris-
759 HCl buffer (pH 7.9 at 21°C) prior to storage at –80°C. Apo-protein concentrations were
760 determined by absorption extinction coefficient at 280 nm calculated from amino acid
761 composition.
762
763 **UV-visible absorption and fluorescence spectroscopy**
764 UV-visible spectra were recorded on a Jasco V-630 spectrophotometer fitted with a
765 temperature-controlled sample holder (Jasco) and spectrosil quartz cuvettes with a path length
766 of 1.0 cm (Starna, Baulkham Hills, Australia). Porphyrin concentrations were determined

767 according to the following molar extinction coefficients and solvent conditions: hemin chloride,
768 ε_{385} = 58,400 $M^{-1}cm^{-1}$ in 0.1 M NaOH [57]; PPIX ε_{554} = 13,500 $M^{-1}cm^{-1}$ in 2.7 M HCl [57];
769 coproporphyrin III, ε_{548} = 16,800 $M^{-1}cm^{-1}$ in 0.1 M HCl [58]; Co(III)PPIX ε424 = 180,000 $M^{-1}cm^{-1}$
770 in NaOH(0.1 M):pyridine:H_2O 3:10:17 [59]. Zn(II)PPIX was determined as free PPIX after
771 decomposition in 2.7 M HCl. To prepare Zn-PPIX, 0.5 g PPIX (Frontier Scientific) was dissolved in
772 boiling chloroform (100 mL), to which a saturated solution of Zn acetate in MeOH (1 mL) was
773 added. The mixture was refluxed for 20 min and then a small amount of MeOH was added and,
774 after cooling, the dark red solid was filtered off (50:50 Zn-PPIX:PPIX by HPLC). Zn-PPIX was
775 purified by RP-HPLC over a C18 solid phase (Phenomenex) with isocratic
776 acetone:MeOH:water:formic acid (280:120:100:1) mobile phase, which achieved baseline
777 separation of the Zn-PPIX fraction.

778 Hemin binding measurements in absorbance mode were made by successive additions
779 of apo-protein (~0.4 mM stock) into porphyrin solution (1.5 µM) in 20 mM Tris-HCl, pH 8 at
780 21°C. All binding experiments were performed at 21°C. Data were fitted to a 1:1 binding model
781 accounting for ligand depletion, where F_{obs} is fluorescence signal, F_0 is the starting

782 $$F_{obs} = F_0 - F_{sat}((L+K_d+M) - \mathrm{sqrt}((L+K_d+M)^2 - 4ML))$$

783 fluorescence, F_{sat} is a scaling factor for fluorescence at saturation, L and M are the ligand and
784 macromolecule concentrations, respectively, and K_d is the equilibrium dissociation constant.
785 Data were fitted using GNUPLOT version 4.6. A 95% confidence interval for the K_d parameter was
786 obtained by determining a threshold sum-of-squares for which a fit with all fixed parameters

787 $$SS_{fixed} = SS_{best}\ (\ (F \times Dfn)/Dfd + 1)$$

788 would not be significantly different from the best-fit model at a significance level of P = 0.05
789 [60]. Here, SS_{best} is the sum-of-squares for the fit with all parameters floated; Dfn and Dfd are
790 the degrees of freedom in the numerator and denominator, respectively, for calculation of *F*.
791 The K_d value was then fixed at values above or below the best-fit K_d until fits exceeded the
792 threshold sum-of-squares. To determine a lower limit of Kd that we could expect to fit from
793 absorbance measurements we simulated data for different K_d values with Gaussian noise added
794 to give fits with SS_{best} that matched our experimental data. We determined by F test that K_d
795 values ≤ 15 nM were not significantly different (at P = 0.05) from arbitrarily high K_d
796 (approximating a straight-line fit), thus we decided on 15 nM as a lower cutoff for reporting K_d.
797 Fluorescence measurements for Zn(II)PPIX binding were made in 96-well format in 20
798 mM Tris-HCl, pH 8.0 at 21°C using individually mixed samples covering an appropriate range of
799 hemophilin protein concentrations based on preliminary experiments. IC_{50} for hemophilins
800 binding to Zn(II)PPIX (0.5 or 1.0 µM) in competition with BSA (15 µM) were determined by
801 fitted to the same 1:1 binding model as described above. The Cheng and Prusoff equation [61]
802 was used to convert to K_d, where [BSA] is the molar concentration of competing BSA (15 µM)

803 $$K_d = IC_{50} / (1 + [BSA]/K_{d,BSA})$$

804 and $K_{d,BSA}$ was the dissociation equilibrium constant for BSA binding to Zn(II)PPIX, determined to
805 be 0.3 (0.2–0.4) µM from absorbance titration data.
806

807 **Generation of threading models for hemophilin active sites**

808 Threading models were generated with the program MODELLER version 10.2 [32], using
809 the X-ray crystal structures of *H. Haemolyticus* Hpl (PDB 6om5) and *A. baumannii* HphA (PDB

810 7red) as templates [13,18]. Residues of the N-terminal signal peptide, as identified by SIGNALP
811 5.0 [62], were removed prior to analysis or modeling of all sequences. The threading approach
812 used by MODELLER relies on accurate alignment of template and query sequences. To identify
813 residues likely to be important for structural integrity of the hemophilin ligand binding domain
814 we collected the small number of sequences that shared similarity with Hpl, HphA and *X.*
815 *nematophila* HrpC in the N-terminal region, or with pair-wise combinations of these, based on
816 BLASTP searches. We aligned these sequences with the structure-based alignment of Hpl and
817 HphA produced by MODELLER to generate a profile against which subclusters from the sequence
818 similarity network analysis were then aligned using CLUSTALO. Alignments were inspected against
819 several criteria to see that they were structurally plausible. We specifically examined the
820 preservation of the conserved hydrophobic core ligand binding domains of Hpl and HphA,
821 corresponding to Hpl residues G26, I38, I40, A49, V51, I53, F76, M79, A83, A104, L125, F127,
822 Y136, G138, W140, Y136, A158 (Fig. 9). An initial structure-based alignment of *H. haemolyticus*
823 Hpl and *A. baumannii* HphA was produced using MODELLER, and this pairing was fixed in all
824 subsequent alignments. Amino acid residues with potential importance for the hemophilin fold
825 were identified by sequence homology (with the signal peptides and BBD removed) using BLASTP
826 [63]. Sequences were identified with similarity to the hemophilin domains of Hpl, HphA and the
827 N-terminal region of *X. nematophila* HrpC (WP_019473020, WP_057440571, WP_244182492),
828 or Hpl and HphA only (WP_005758278), or HphA and HrpC only (MBP6115507, WP_228864429,
829 WP_232888613), or Hpl and HrpC only (WP_038256617, WP_092512525, WP_244182492), or
830 HphA only (WP_121975315), or Hpl only (RKV63521). These sequences were aligned with the
831 structure-based alignment of Hpl and HphA using CLUSTAL OMEGA [64] to generate a. 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 [65].

Acknowledgements:

We thank Dr. Brianna Atto for useful discussions and Sarah Kauffman for technical assistance.

Financial Disclosure Statement:

This work was funded in part by funds awarded to HGB from the University of Tennessee-Knoxville 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.

847 **References:**

848 1. Bertini I, Gray H, Stiefel E, Selverstone Valentine J. Biological inorganic chemistry:
849 structure and reactivity. Sausalito: University Science Books; 2007.
850 2. Zhang H-R, Wang Y, Xiu P, Qi Y, Chai F. Roles of iron limitation in phytoplankton dynamics
851 in the western and eastern subarctic pacific. Frontiers in Marine Science. 2021. Available:
852 https://www.frontiersin.org/article/10.3389/fmars.2021.735826
853 3. Hood MI, Skaar EP. Nutritional immunity: transition metals at the pathogen–host
854 interface. Nat Rev Microbiol. 2012;10: 525–537. doi:10.1038/nrmicro2836
855 4. Khan FA, Fisher MA, Khakoo RA. Association of hemochromatosis with infectious
856 diseases: expanding spectrum. Int J Infect Dis. 2007;11: 482–487.
857 doi:https://doi.org/10.1016/j.ijid.2007.04.007
858 5. Posey JE, Gherardini FC. Lack of a role for iron in the Lyme disease pathogen. Science.
859 2000;288: 1651–1653. doi:10.1126/science.288.5471.1651
860 6. Holden VI, Bachman MA. Diverging roles of bacterial siderophores during infection.
861 Metallomics. 2015;7: 986–995. doi:10.1039/c4mt00333k
862 7. Hooda Y, Lai CC-L, Judd A, Buckwalter CM, Shin HE, Gray-Owen SD, et al. Slam is an outer
863 membrane protein that is required for the surface display of lipidated virulence factors in
864 Neisseria. Nat Microbiol. 2016;1. doi:10.1038/nmicrobiol.2016.9
865 8. Chen CJ, Sparling PF, Lewis LA, Dyer DW, Elkins C. Identification and purification of a
866 hemoglobin-binding outer membrane protein from Neisseria gonorrhoeae. Infect
867 Immun. 1996;64: 5008–5014. doi:10.1128/iai.64.12.5008-5014.1996
868 9. Calmettes C, Alcantara J, Yu R-H, Schryvers AB, Moraes TF. The structural basis of
869 transferrin sequestration by transferrin-binding protein B. Nat Struct Mol Biol. 2012;19:
870 358–360. doi:10.1038/nsmb.2251
871 10. Ostan NKH, Yu R-H, Ng D, Lai CC-L, Pogoutse AK, Sarpe V, et al. Lactoferrin binding
872 protein B - a bi-functional bacterial receptor protein. PLoS Pathog. 2017;13: e1006244.
873 doi:10.1371/journal.ppat.1006244
874 11. Hooda Y, Lai CC-L, Moraes TF. Identification of a large family of Slam-dependent surface
875 lipoproteins in Gram-negative bacteria. Front Cell Infect Microbiol. 2017;7.
876 doi:10.3389/fcimb.2017.00207
877 12. Grossman AS, Mauer TJ, Forest KT, Goodrich-Blair H, Comstock L. A widespread bacterial
878 secretion system with diverse substrates. MBio. 2022;12: e01956-21.
879 doi:10.1128/mBio.01956-21
880 13. Bateman TJ, Shah M, Ho TP, Shin HE, Pan C, Harris G, et al. A Slam-dependent
881 hemophore contributes to heme acquisition in the bacterial pathogen Acinetobacter
882 baumannii. Nat Commun. 2021;12: 6270. doi:10.1038/s41467-021-26545-9
883 14. Arnoux P, Haser R, Izadi-Pruneyre N, Lecroisey A, Czjzek M. Functional aspects of the
884 heme bound hemophore HasA by structural analysis of various crystal forms. Proteins
885 Struct Funct Bioinforma. 2000;41: 202–210. doi:https://doi.org/10.1002/1097-
886 0134(20001101)41:2<202::AID-PROT50>3.0.CO;2-8
887 15. Gianquinto E, Moscetti I, De Bei O, Campanini B, Marchetti M, Luque FJ, et al. Interaction
888 of human hemoglobin and semi-hemoglobins with the Staphylococcus aureus
889 hemophore IsdB: a kinetic and mechanistic insight. Sci Rep. 2019;9: 18629.

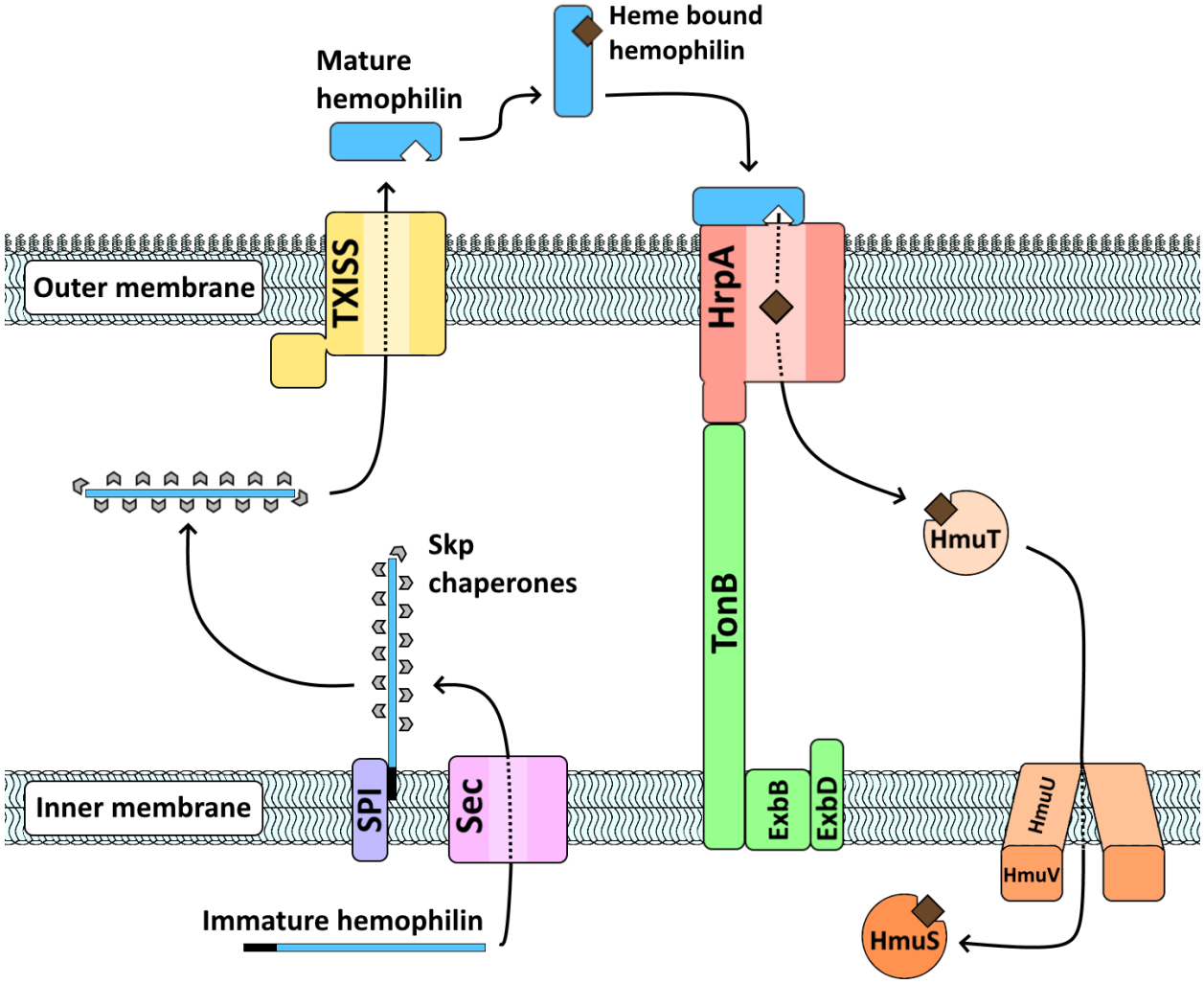
890 doi:10.1038/s41598-019-54970-w
891 16. Huynh MS, Hooda Y, Li YR, Jagielnicki M, Lai CC-L, Moraes TF. Reconstitution of surface
892 lipoprotein translocation through the Slam translocon. Elife. 2022;11.
893 doi:10.7554/eLife.72822
894 17. Fantappiè L, Irene C, De Santis M, Armini A, Gagliardi A, Tomasi M, et al. Some Gram-
895 negative lipoproteins keep their surface topology when transplanted from one species to
896 another and deliver foreign polypeptides to the bacterial surface. Mol & Cell
897 Proteomics. 2017;16: 1348 LP – 1364. doi:10.1074/mcp.M116.065094
898 18. Latham RD, Torrado M, Atto B, Walshe JL, Wilson RK, Guss JM, et al. A heme-binding
899 protein produced by Haemophilus haemolyticus inhibits non-typeable Haemophilus
900 influenzae. Mol Microbiol. 2020;113: 381–398. doi:https://doi.org/10.1111/mmi.14426
901 19. Schwiesow L, Mettert E, Wei Y, Miller HK, Herrera NG, Balderas D, et al. Control of hmu
902 heme uptake genes in Yersinia pseudotuberculosis in response to iron sources. Frontiers
903 in Cellular and Infection Microbiology. 2018. Available:
904 https://www.frontiersin.org/article/10.3389/fcimb.2018.00047
905 20. Atto B, Latham RD, Kunde D, Gell DA, Tristram S. In vitro probiotic potential of
906 hemophilin-producing strains of Haemophilus haemolyticus. bioRxiv. 2020;
907 2020.01.02.893487. doi:10.1101/2020.01.02.893487
908 21. Atto B, Kunde D, Gell DA, Tristram S. Oropharyngeal carriage of hpl-containing
909 Haemophilus haemolyticus predicts lower prevalence and density of NTHi colonisation in
910 healthy adults. Pathog (Basel, Switzerland). 2021;10. doi:10.3390/pathogens10050577
911 22. Zallot R, Oberg NO, Gerlt JA. The EFI web resource for genomic enzymology tools:
912 leveraging protein, genome, and metagenome databases to discover novel enzymes and
913 metabolic pathways. Biochemistry. 2019;58: 4169–4182.
914 doi:10.1021/acs.biochem.9b00735
915 23. Tietz JI, Schwalen CJ, Patel PS, Maxson T, Blair PM, Tai H-C, et al. A new genome-mining
916 tool redefines the lasso peptide biosynthetic landscape. Nat Chem Biol. 2017;13: 470–
917 478. doi:10.1038/nchembio.2319
918 24. Egler M, Grosse C, Grass G, Nies DH. Role of the extracytoplasmic function protein family
919 sigma factor RpoE in metal resistance of Escherichia coli. J Bacteriol. 2005;187: 2297–
920 2307. doi:10.1128/JB.187.7.2297-2307.2005
921 25. Schwartz CJ, Giel JL, Patschkowski T, Luther C, Ruzicka FJ, Beinert H, et al. IscR, an Fe-S
922 cluster-containing transcription factor, represses expression of Escherichia coli genes
923 encoding Fe-S cluster assembly proteins. Proc Natl Acad Sci. 2001;98: 14895–14900.
924 doi:10.1073/pnas.251550898
925 26. Maddocks SE, Oyston PCF. Structure and function of the LysR-type transcriptional
926 regulator (LTTR) family proteins. Microbiology. 2008;154: 3609–3623.
927 doi:10.1099/mic.0.2008/022772-0
928 27. Science Gateway. Protein molecular weight calculator. www.sciencegateway.org;
929 Available: https://www.sciencegateway.org/tools/proteinmw.htm
930 28. Grossman AS, Escobar CA, Mans EJ, Mucci NC, Mauer TJ, Jones KA, et al. A surface
931 exposed, two-domain lipoprotein cargo of a type XI secretion system promotes
932 colonization of host intestinal epithelia expressing glycans. Frontiers in Microbiology.
933 2022. Available: https://www.frontiersin.org/article/10.3389/fmicb.2022.800366

934 29. Holm L, Rosenström P. Dali server: conservation mapping in 3D. Nucleic Acids Res.
935 2010;38: W545-9. doi:10.1093/nar/gkq366
936 30. Sillitoe I, Bordin N, Dawson N, Waman VP, Ashford P, Scholes HM, et al. CATH: increased
937 structural coverage of functional space. Nucleic Acids Res. 2021;49: D266–D273.
938 doi:10.1093/nar/gkaa1079
939 31. Andreeva A, Kulesha E, Gough J, Murzin AG. The SCOP database in 2020: expanded
940 classification of representative family and superfamily domains of known protein
941 structures. Nucleic Acids Res. 2020;48: D376–D382. doi:10.1093/nar/gkz1064
942 32. Webb B, Sali A. Comparative protein structure modeling using MODELLER. Curr Protoc
943 Bioinforma. 2016;54: 5.6.1-5.6.37. doi:https://doi.org/10.1002/cpbi.3
944 33. Bhasin A, Chaston JM, Goodrich-Blair H. Mutational analyses reveal overall topology and
945 functional regions of NilB, a bacterial outer membrane protein required for host
946 association in a model of animal-microbe mutualism. J Bacteriol. 2012;194: 1763–1776.
947 doi:10.1128/JB.06711-11
948 34. Cowles CE, Goodrich-Blair H. The Xenorhabdus nematophila nilABC genes confer the
949 ability of Xenorhabdus spp. to colonize Steinernema carpocapsae nematodes. J Bacteriol.
950 2008;190: 4121–4128. doi:10.1128/JB.00123-08
951 35. Moraes TF, Lai CC-L, Hooda Y, Judd A, Schryvers AB, Gray-Owen SD. Slam polynucleotides
952 and polypeptides and uses thereof. US; US20210379176A1, 2021. pp. 1–36. Available:
953 https://patents.google.com/patent/US20210379176A1/en
954 36. Debret G, Martel A, Cuniasse P. RASMOT-3D PRO: a 3D motif search webserver. Nucleic
955 Acids Res. 2009;37: W459–W464. doi:10.1093/nar/gkp304
956 37. Burmester T, Hankeln T. The respiratory proteins of insects. J Insect Physiol. 2007;53:
957 285–294. doi:10.1016/j.jinsphys.2006.12.006
958 38. Buecher EJ, Popiel I. Liquid culture of the entomogenous nematode steinernema feltiae
959 with its bacterial symbiont. J Nematol. 1989;21: 500–504.
960 39. Hieb WF, Stokstad EL, Rothstein M. Heme requirement for reproduction of a free-living
961 nematode. Science. 1970;168: 143–144. doi:10.1126/science.168.3927.143
962 40. Weber JJ, Kanost MR, Gorman MJ. Iron binding and release properties of transferrin-1
963 from Drosophila melanogaster and Manduca sexta: Implications for insect iron
964 homeostasis. Insect Biochem Mol Biol. 2020;125: 103438.
965 doi:10.1016/j.ibmb.2020.103438
966 41. Goodrich-Blair H. They've got a ticket to ride: Xenorhabdus nematophila–Steinernema
967 carpocapsae symbiosis. Curr Opin Microbiol. 2007;10: 225–230.
968 doi:https://doi.org/10.1016/j.mib.2007.05.006
969 42. Leys D, Backers K, Meyer TE, Hagen WR, Cusanovich MA, Van Beeumen JJ. Crystal
970 structures of an oxygen-binding cytochrome c from Rhodobacter sphaeroides. J Biol
971 Chem. 2000;275: 16050–16056. doi:10.1074/jbc.275.21.16050
972 43. Girvan HM, Toogood HS, Littleford RE, Seward HE, Smith WE, Ekanem IS, et al. Novel
973 haem co-ordination variants of flavocytochrome P450BM3. Biochem J. 2009;417: 65–76.
974 doi:10.1042/BJ20081133
975 44. Safarian S, Rajendran C, Müller H, Preu J, Langer JD, Ovchinnikov S, et al. Structure of a
976 bd oxidase indicates similar mechanisms for membrane-integrated oxygen reductases.
977 Science. 2016;352: 583–586. doi:10.1126/science.aaf2477

978 45. Murali R, Gennis RB. Functional importance of Glutamate-445 and Glutamate-99 in
979 proton-coupled electron transfer during oxygen reduction by cytochrome bd from
980 Escherichia coli. Biochim Biophys Acta - Bioenerg. 2018;1859: 577–590.
981 doi:https://doi.org/10.1016/j.bbabio.2018.04.012
982 46. Johnson EA, Russo MM, Nye DB, Schlessman JL, Lecomte JTJ. Lysine as a heme iron
983 ligand: A property common to three truncated hemoglobins from Chlamydomonas
984 reinhardtii. Biochim Biophys Acta - Gen Subj. 2018;1862: 2660–2673.
985 doi:https://doi.org/10.1016/j.bbagen.2018.08.009
986 47. Shannon P. Cytoscape: A software environment for integrated models of biomolecular
987 interaction networks. Genome Res. 2003;13: 2498–2504. doi:10.1101/gr.1239303
988 48. Bastian M, Heymann S, Jacomy M. Gephi: An open source software for exploring and
989 manipulating networks. Proc Int AAAI Conf Web Soc Media. 2009;3: 361–362. Available:
990 https://ojs.aaai.org/index.php/ICWSM/article/view/13937
991 49. Fruchterman TMJ, Reingold EM. Graph drawing by force-directed placement. Softw Pract
992 Exp. 1991;21: 1129–1164. doi:10.1002/spe.4380211102
993 50. Jacomy M, Venturini T, Heymann S, Bastian M. ForceAtlas2, a continuous graph layout
994 algorithm for handy network visualization designed for the Gephi software. PLoS One.
995 2014;9: e98679. Available: https://doi.org/10.1371/journal.pone.0098679
996 51. Miroux B, Walker JE. Over-production of proteins in Escherichia coli: mutant hosts that
997 allow synthesis of some membrane proteins and globular proteins at high levels. J Mol
998 Biol. 1996;260: 289–298. doi:10.1006/jmbi.1996.0399
999 52. Dumon-Seignovert L, Cariot G, Vuillard L. The toxicity of recombinant proteins in
1000 Escherichia coli: a comparison of overexpression in BL21(DE3), C41(DE3), and C43(DE3).
1001 Protein Expr Purif. 2004;37: 203–206. doi:10.1016/j.pep.2004.04.025
1002 53. Koontz L. Chapter one - TCA precipitation. In: Lorsch JBT-M in E, editor. Laboratory
1003 methods in enzymology. Academic Press; 2014. pp. 3–10.
1004 doi:https://doi.org/10.1016/B978-0-12-420119-4.00001-X
1005 54. Tukey JW. Comparing individual means in the analysis of variance. Biometrics. 1949;5:
1006 99–114. doi:10.2307/3001913
1007 55. Ascoli F, Rossi Fanelli MR, Antonini EBT-M in E. Preparation and properties of
1008 apohemoglobin and reconstituted hemoglobins. Hemoglobins. Academic Press; 1981. pp.
1009 72–87. doi:https://doi.org/10.1016/0076-6879(81)76115-9
1010 56. van den Berg S, Löfdahl P-Å, Härd T, Berglund H. Improved solubility of TEV protease by
1011 directed evolution. J Biotechnol. 2006;121: 291–298.
1012 doi:https://doi.org/10.1016/j.jbiotec.2005.08.006
1013 57. Arnstein HR V. Data for biochemical research (Third Edition) : by R.M.C. Dawson, D.C.
1014 Elliott, W.H. Elliott and K.M. Jones Oxford University Press; Oxford, 1986 xii + 580 pages.
1015 £35.00. FEBS Lett. 1988;234: 506.
1016 58. Fuhrhop JH, Smith KM. Laboratory methods in porphyrin and metalloporphyrin research.
1017 Elsevier Scientific Publishing Company; 1975. Available:
1018 https://books.google.com/books?id=0tpFwgEACAAJ
1019 59. Yonetani T, Yamamoto H, Woodrow GV 3rd. Studies on cobalt myoglobins and
1020 hemoglobins. I. Preparation and optical properties of myoglobins and hemoglobins
1021 containing cobalt proto-, meso-, and deuteroporphyrins and thermodynamic

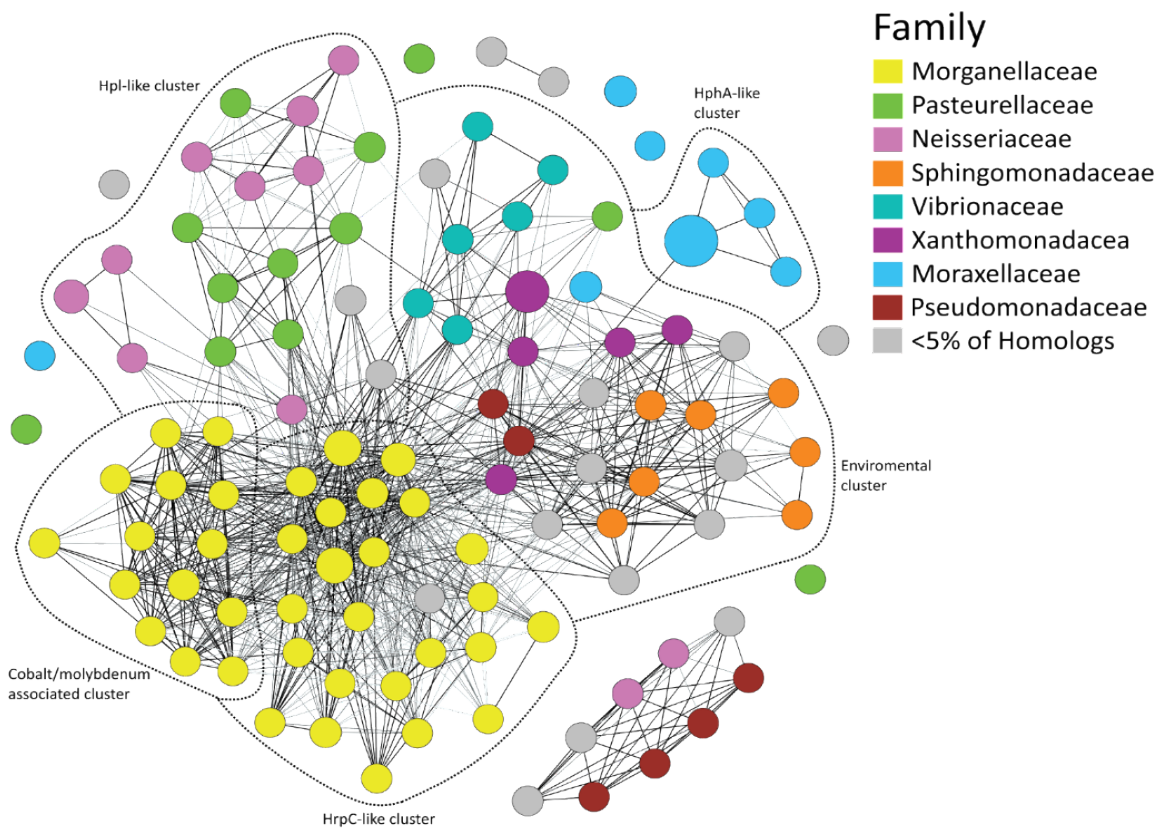
1022 characterization of their reversible oxygenation. J Biol Chem. 1974;249: 682–690.
1023 60. Motulsky H, Christopoulos A. Fitting models to biological data using linear and nonlinear
1024 regression: a practical guide to curve fitting. Oxford University Press; 2004.
1025 61. Cheng Y, Prusoff WH. Relationship between the inhibition constant (K1) and the
1026 concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic
1027 reaction. Biochem Pharmacol. 1973;22: 3099–3108. doi:10.1016/0006-2952(73)90196-2
1028 62. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, et
1029 al. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat
1030 Biotechnol. 2019;37: 420–423. doi:10.1038/s41587-019-0036-z
1031 63. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J
1032 Mol Biol. 1990;215: 403–410. doi:10.1016/S0022-2836(05)80360-2
1033 64. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of
1034 high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol.
1035 2011;7: 539. doi:10.1038/msb.2011.75
1036 65. Brown NP, Leroy C, Sander C. MView: a web-compatible database search or multiple
1037 alignment viewer. Bioinformatics. 1998;14: 380–381.
1038 doi:10.1093/bioinformatics/14.4.380
1039

1040



1041

1042 Figure 1. **Conceptual model of hemophilin mediated heme acquisition.** Described hemophilin
1043 homologs have a signal peptide directing them to the Sec translocon and signal peptidase I.
1044 Based on analogy to the T11SS-dependent lipoprotein TbpB, hemophilin may depend on Skp for
1045 periplasmic transport. Each hemophilin protein has a cognate T11SS which translocates it
1046 through the outer membrane where it acts to bring heme to a TonB-dependent receptor for
1047 uptake. Based on analogy to the *Yersinia pestis* Has hemophore system, heme is transported
1048 through the periplasm and across the inner membrane by the hemin utilization system (Hmu).
1049 Once brought into the cytoplasm, heme can be degraded by heme oxygenase (HmuS), an
1050 enzyme which is frequently encoded adjacent to hemophilin loci.
1051



Supplementary Figure S1. **Distribution and relatedness of soluble T11SS-dependent cargo proteins.** A sequence similarity network of all soluble TbpBBD-domain-containing proteins identified in Grossman and Mauer, et al. (2022). All of these sequences are encoded in the same genomic region as a T11SS-encoding gene (± 6 genes). Sequences were submitted to EFI-EST (see methods) to generate the network. 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 sequence identity, with the darkest edges being the most identical. Dotted lines indicate proposed subclusters as defined using the Fruchterman-Reingold algorithm and the distribution of characterized proteins. Sequences that did not connect to any characterized hemophilin family protein were excluded from further analysis (e.g., Figure 3). Despite having similar domain architectures to hemophilin, none of these divergent sequences contained experimentally characterized proteins.

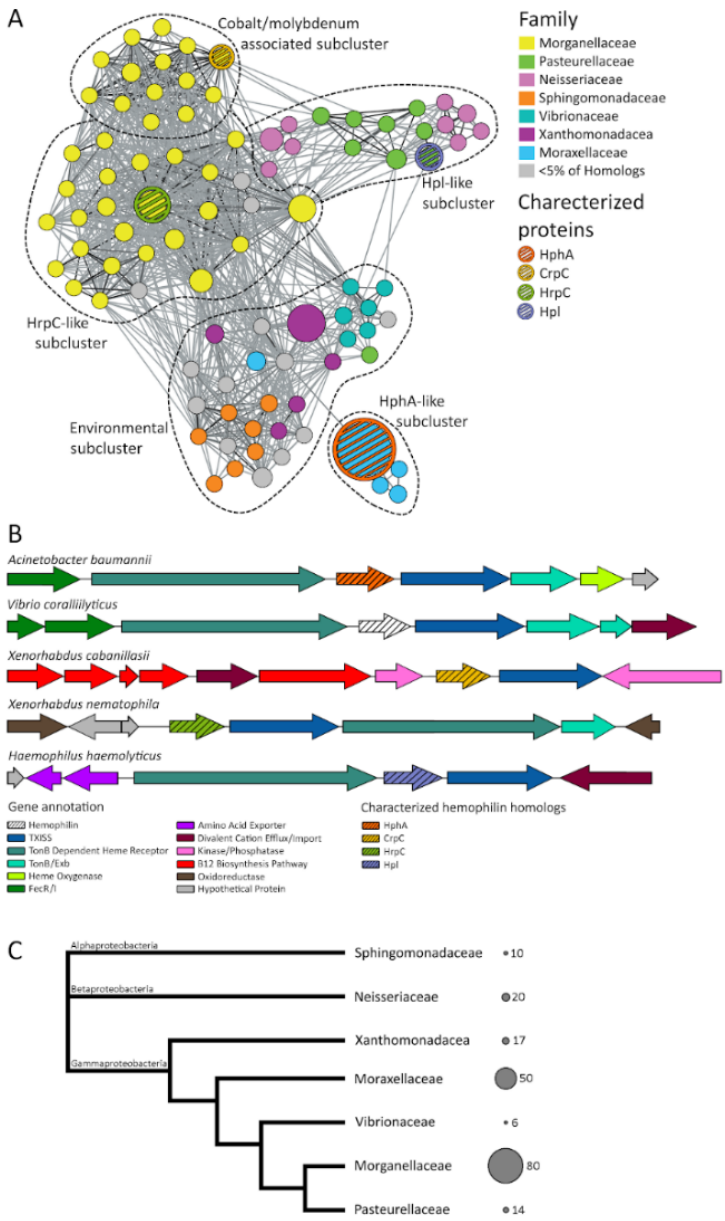


Figure 2. **Distribution and relatedness of hemophilin family proteins. A)** A sequence similarity network of hemophilin homologs generated with EFI-EST. Each node represents one or more protein sequences with 80% or greater identity, the larger the node the more sequences it contains. Edges indicate an alignment score of 35 or greater. Edge darkness indicates shared sequence identity, with the darkest edges being the most identical. Dotted lines indicate proposed subclusters as defined using a 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.

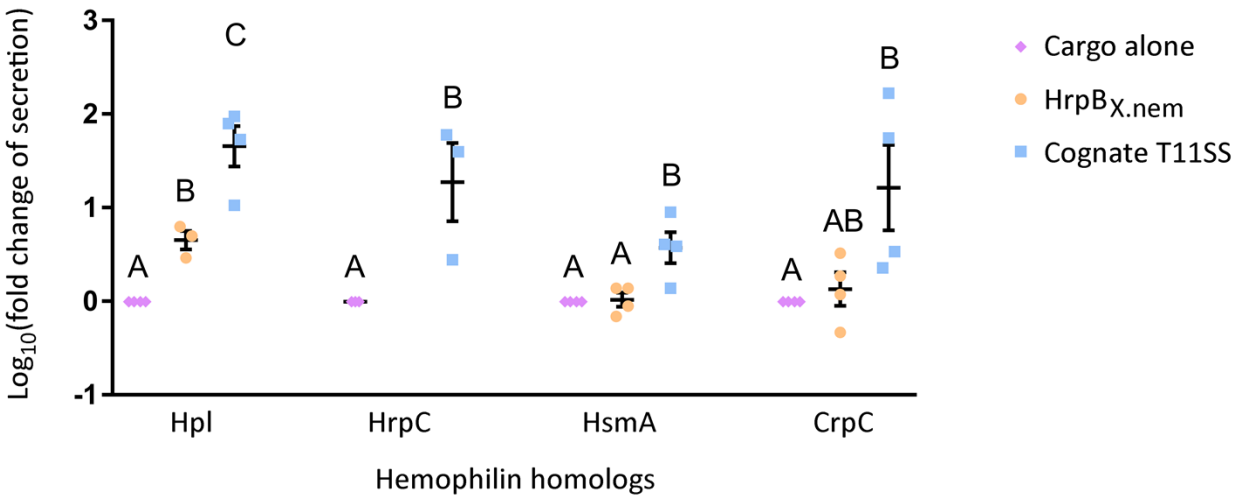


Figure 3. **Secretion of hemophilin proteins by their cognate and non-cognate T11SS proteins.** Each cargo protein was expressed in *Escherichia coli* in isolation (cargo alone; pink diamonds) or co-expressed in with their cognate T11SS (blue squares) or with non-cognate $HrpB_{X.nem}$ (orange circles). 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 log_{10} function prior to performing a Tukey's HSD test for each hemophilin homolog. Letters indicate significance groups. All four T11SS proteins assayed significantly increased the amount of cognate cargo that was secreted into the extracellular milieu. The non-cognate T11SS $HrpB_{X.nem}$ significantly increased secretion of Hpl but less effectively than the cognate T11SS of Hpl. $HrpB_{X.nem}$ did not significantly affect the secretion of HsmA or CrpC.

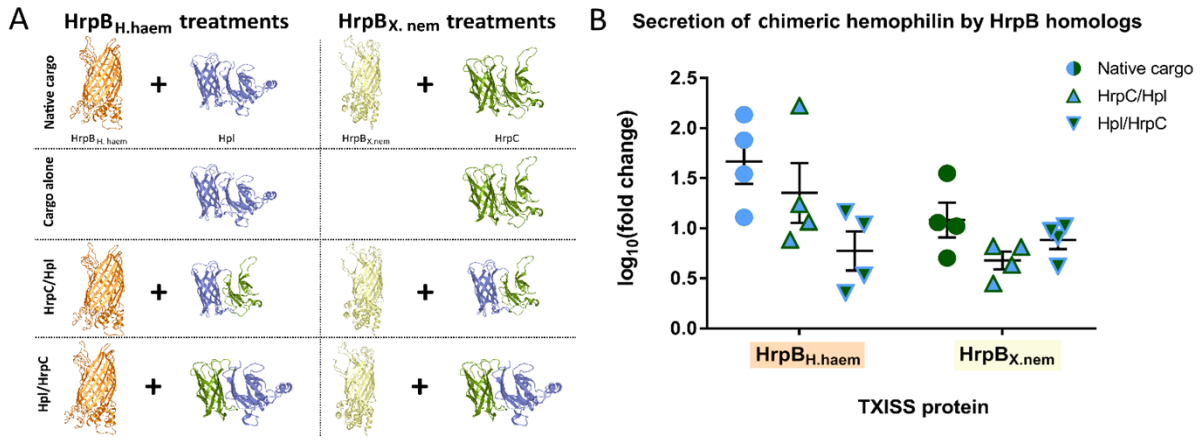


Figure 4. **Secretion of domain-swapped chimeric hemophilin proteins by $HrpB_{X.nem}$ and $HrpB_{H.haem}$. A)** A pictorial depiction of the experimental treatments used to assess secretion of chimeric hemophilin cargo proteins by HrpB T11SS secretors from *H. haemolyticus* ($HrpB_{H.haem}$; orange) 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 Hlp/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)** Secretion data were made normal with a log_{10} transformation. Both 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 proteins.

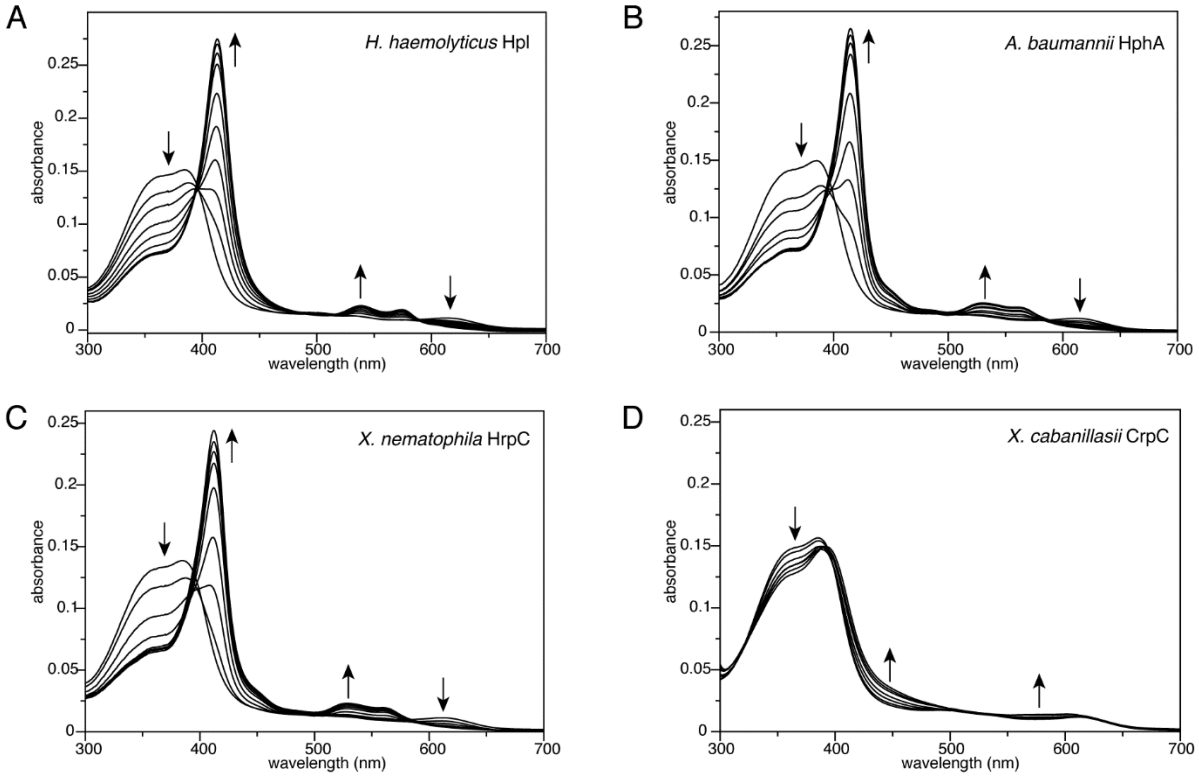


Figure 5. **Titration of hemophilin homologues with hemin.** Titrations of hemophilin proteins into hemin (Fe(III)-PPIX) solution (2.5 µM hemin in 20 mM Tris-HCl, pH 8.0 at 21°C) were monitored by UV-visible absorption spectroscopy. **A)** Spectra recorded after addition of *H. haemolyticus* Hpl to a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

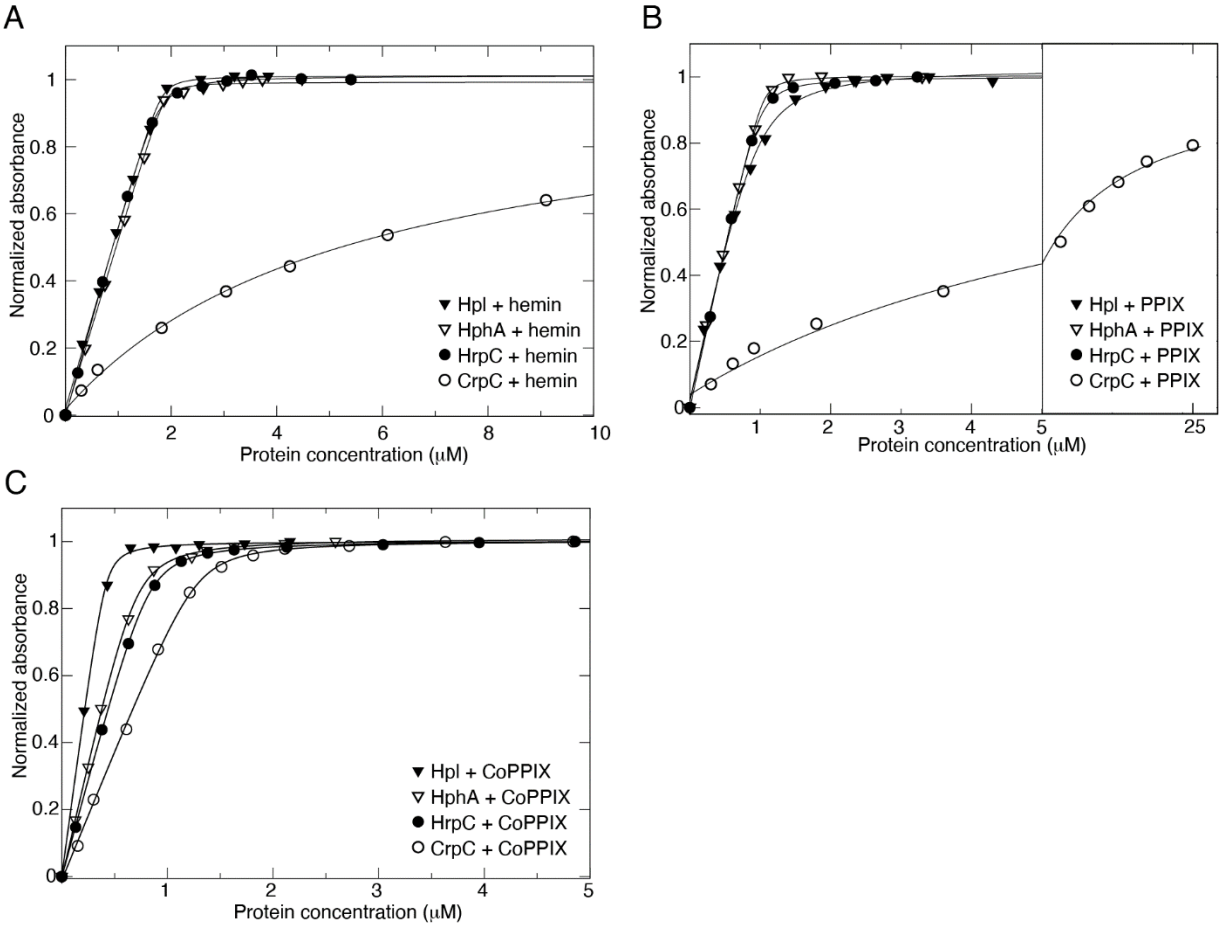
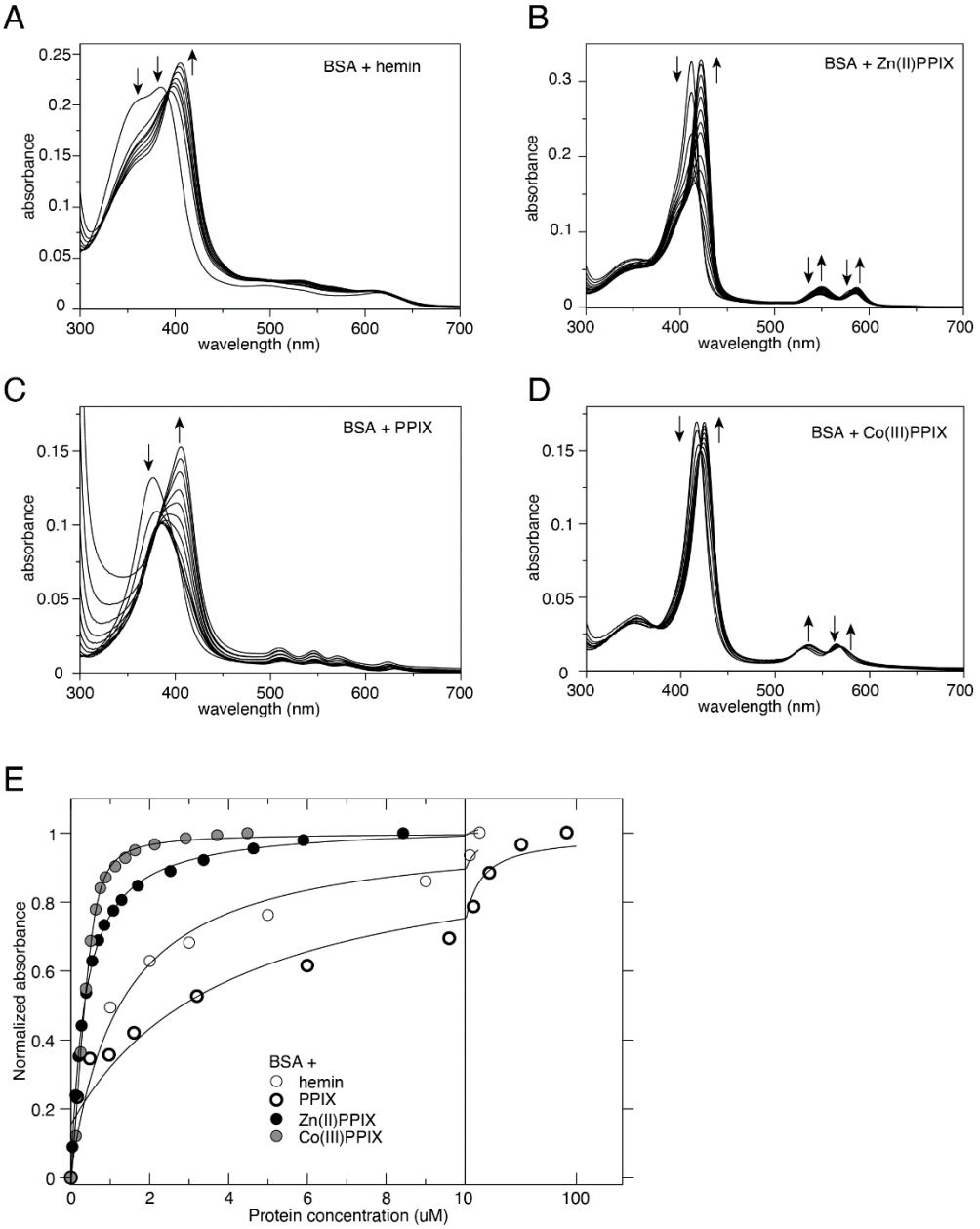
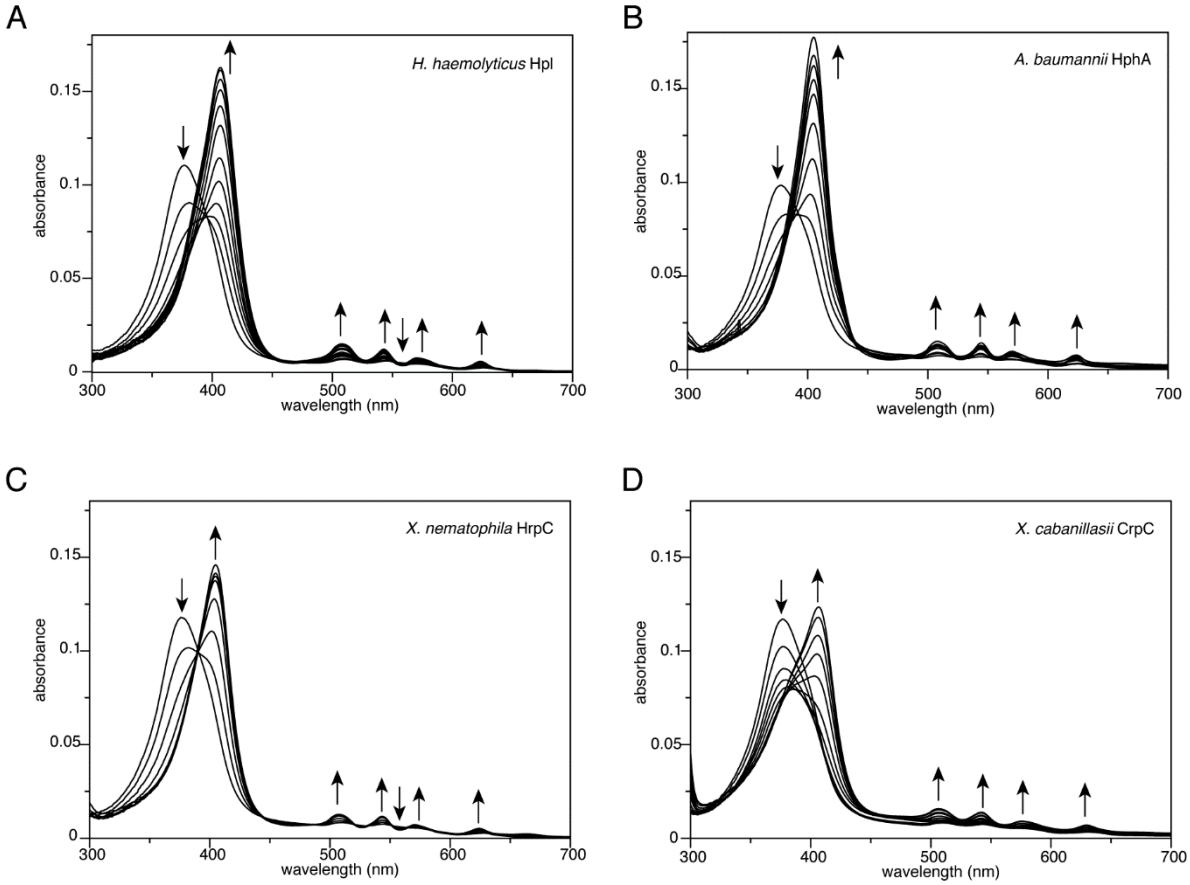


Figure 6. **Binding isotherms for hemophilin homologs binding to porphyrins**. Binding isotherms were generated from absorption data (symbols) by fitting to peak wavelength changes using a 1:1 binding model accounting for ligand depletion (lines). **A)** Binding isotherms for HpI, HphA, HrpC or CrpC with hemin, generated from data in Figure 5. **B)** Binding isotherms for HpI, HphA, HrpC or CrpC with PPIX, generated from data in Supplementary Figure 2. **C)** Binding isotherms for HpI, HphA, HrpC or CrpC with Co(III)PPIX, generated from data in Supplementary Figure 4.



Supplementary Figure 2. **Titration of BSA into porphyrins monitored by UV-vis spectroscopy. A)** Spectra of hemin (2.5 µM) alone or with addition of BSA to a final concentration covering the range 1–21 µM. **B)** Spectra of Zn(II)PPIX (1.5 µM), alone or with addition of BSA to a final concentration in the range 0.05–12 µM. **C)** Spectra of PPIX (1.5 µM) alone or with addition of BSA to a final concentration in the range 0.2–92 µM. **D)** Spectra of Co(III)PPIX (1.5 µM) alone or with addition of BSA to a final concentration in the range 0.13–4.5 µM BSA. Arrows show the direction of spectral changes. All spectra were recorded in 20 mM Tris-HCl, pH 8.0 (at 21°C). **E)** Binding isotherms for BSA with hemin, PPIX, Zn(II)PPIX and Co(III)PPIX generated from the data in **A–D** by fitting to peak wavelength changes using a 1:1 binding model accounting for ligand depletion.



Supplementary Figure 3. **Titration of HrpC proteins in PPIX monitored by UV-vis spectroscopy. A)** Spectra of PPIX (1.5 µM) alone, or with addition of *H. haemolyticus* Hpl to a final concentration in the range 0.2–4.3 µM. **B)** Spectra of PPIX (1.5 µM) alone, or with addition of *A. baumannii* HphA to a final concentration in the range 0.2–3.3 µM. **C)** Spectra of PPIX (1.5 µM) alone, or with addition of *X. nematophila* HrpC to a final concentration in the range 0.3–3.2 µM. **D)** Spectra of PPIX (1.5 µM) alone, or with addition of *X. cabanillasii* CrpC to a final concentration in the range 0.3–24 µM. Arrows show the direction of spectral changes. All spectra were recorded in 20 mM Tris-HCl, pH 8.0 (at 21°C).

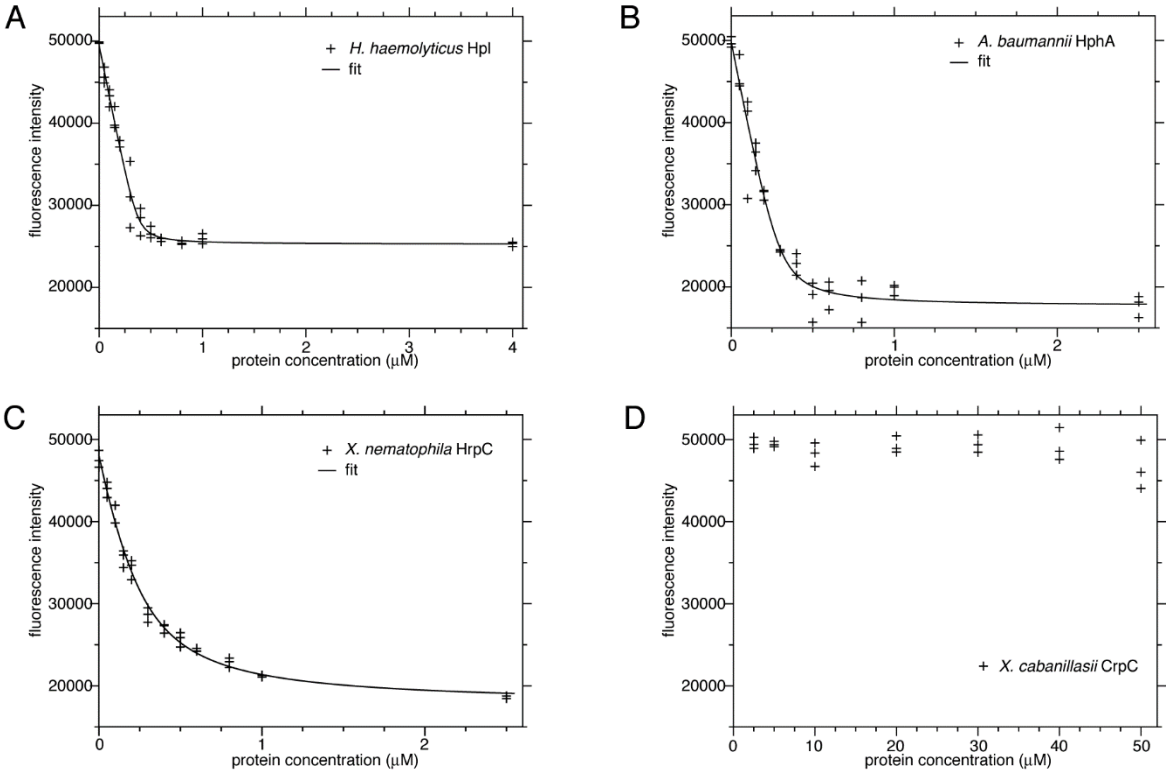
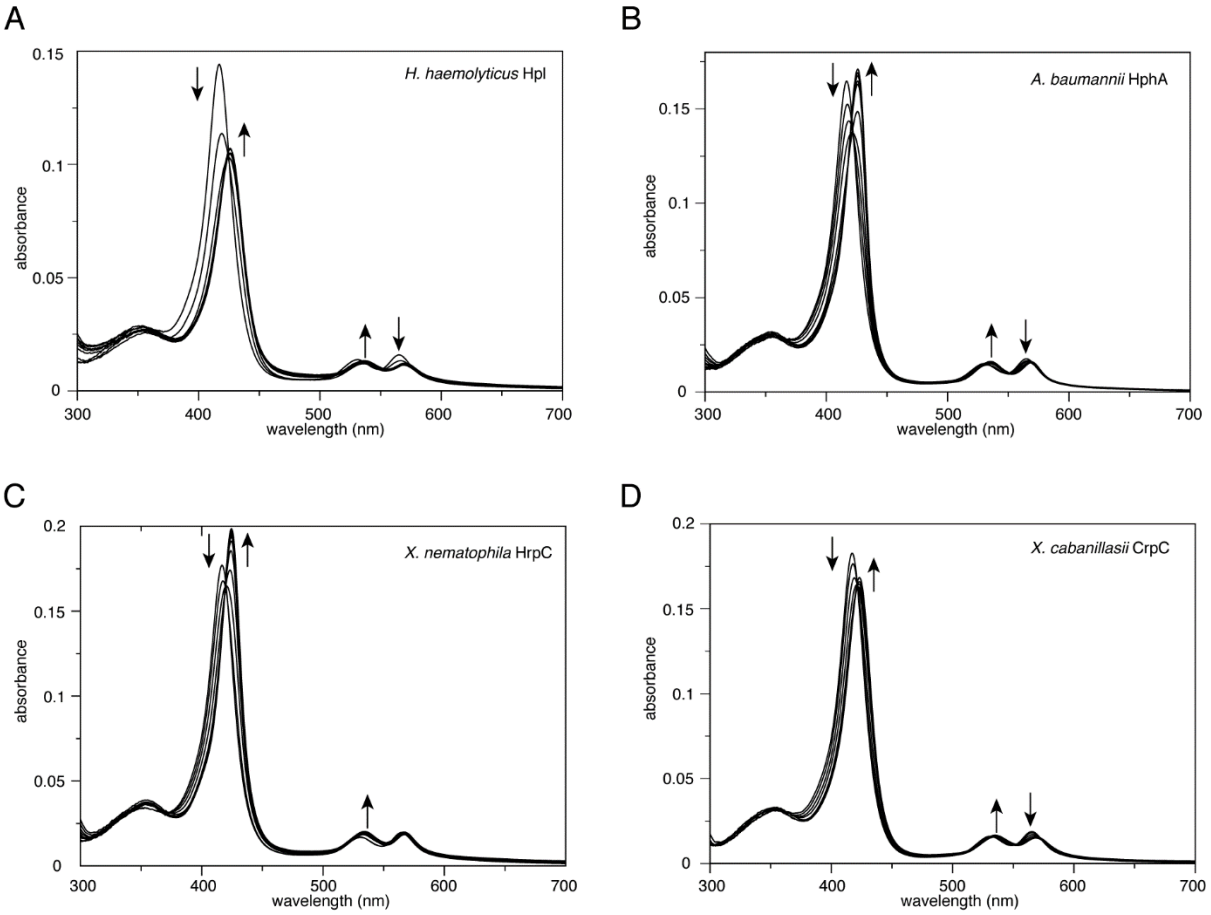
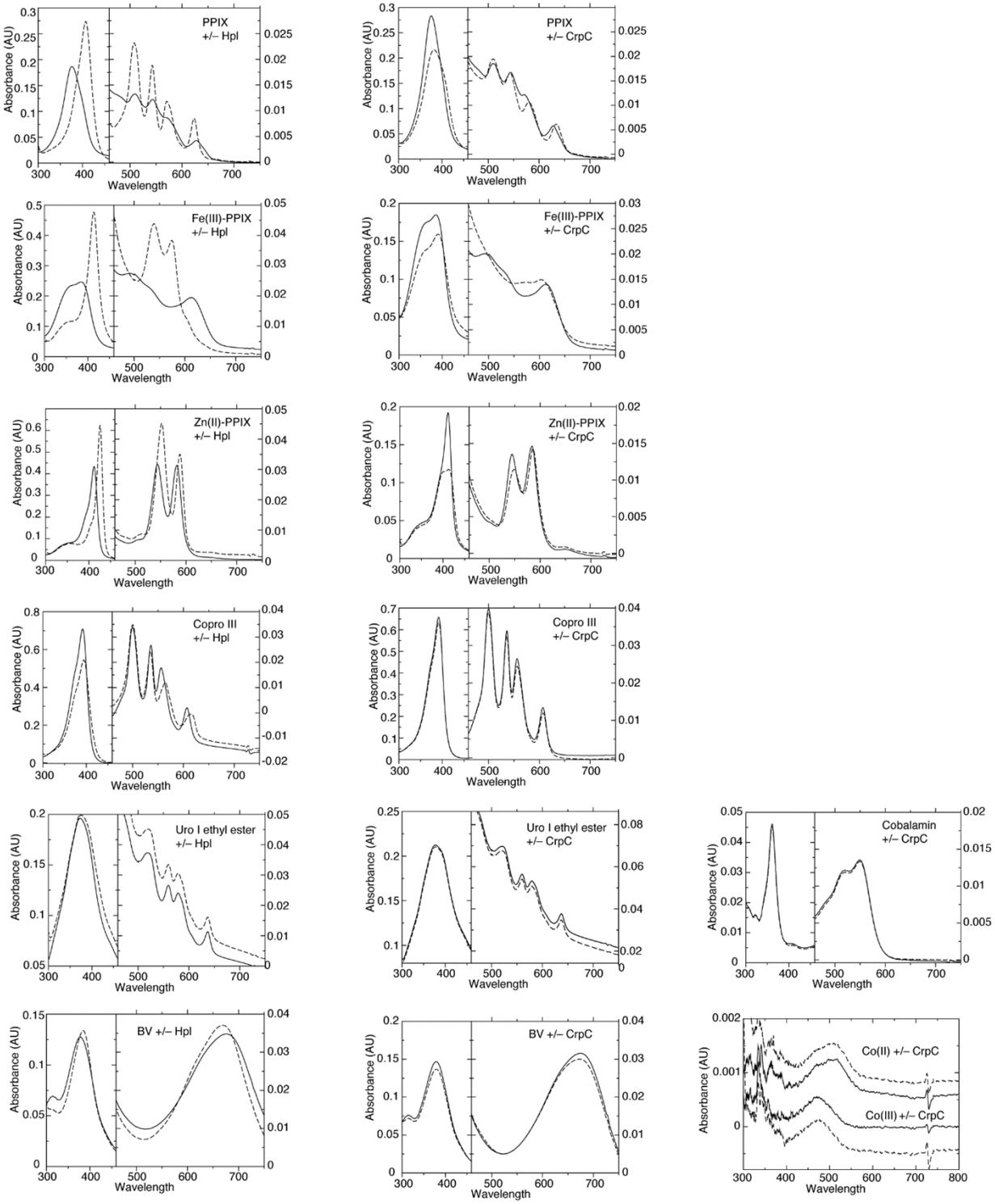


Figure 7. **Titration of hemophilin homologues with Zn(II)PPIX.** Binding of hemophilin homologues was monitored by changes in fluorescence intensity of the Zn(II)PPIX ligand (prepared in 20 mM Tris-HCl, pH 8.0 at 21°C) in the presence of excess BSA (15 μM). Data points (+) were fit (solid lines) with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



Supplementary Figure 4. **Titration of HrpC proteins in Co(III)PPIX monitored by UV-vis spectroscopy. A)** Spectra of Co(III)PPIX (1.5 µM) alone, or with addition of *H. haemolyticus* HpI to a final concentration in the range 0.2–2.2 µM. **B)** Spectra of Co(III)PPIX (1.5 µM) alone, or with addition of *A. baumannii* HphA to a final concentration in the range 0.13–2.6 µM. **C)** Spectra of Co(III)PPIX (1.5 µM) alone, or with addition of *X. nematophila* HrpC to a final concentration in the range 0.3–3.2 µM. **D)** Spectra of Co(III)PPIX (1.5 µM) alone, or with addition of *X. cabanillasii* CrpC to a final concentration in the range 0.15–4.8 µM. Arrows show the direction of spectral changes. All spectra were recorded in 20 mM Tris-HCl, pH 8.0 (at 21°C).



Supplementary Figure 5. **Propensity of Hpl and CrpC to bind to selected linear and cyclic tetrapyrroles.** UV-visible spectra of selected tetrapyrroles (4 µM) in the absence (solid lines) or presence of Hpl (left panels, broken lines) or CrpC (right panels, broken lines) at a concentration of 6 µM. Selected ligands are: PPIX, hemin (Fe(III)PPIX), Zn(II)PPIX, Coproporphyrin III (an intermediate in the coproporphyrin pathway of heme synthesis), uroporphyrin I ethyl ester (from porphyric bovine urine), BV (a linear tetrapyrrole and a natural breakdown product of

hemin), cobalamin, Co(II) chloride, and hexamine Co(III) chloride. The buffer condition was 20 mM Tris-HCl, pH 8.0 (21°C). Addition of Hpl to PPIX, Zn(II)PPIX, Fe(III)PPIX caused large spectral peak shifts consistent with stoichiometric binding. By comparison, minor changes in absorption peak wavelengths accompanied addition of CrpC to PPIX, Zn(II)PPIX and Fe(III)PPIX reflecting the weak binding characterized in titration studies. Copro III contains four propionate groups and is therefore larger and more polar than PPIX. Weak binding of Hpl to copro III was indicated by small spectroscopic shifts, whereas no binding was evident with CrpC. Similarly, small spectral changes indicated some propensity for Hpl, but not CrpC, to bind to the linear tetrapyrrole, BV. Neither Hpl nor CrpC showed any indication of binding to uroporphyrin I ethyl ester (compared to PPIX, this is a larger amphipathic porphyrin with four alkylated propionates). No spectroscopic changes were induced upon mixing CrpC with cobalamin, Co(II) chloride, and hexamine Co(III) chloride.

Table 1. **Relative porphyrin binding affinity of hemophilin homologs**

Protein	Hemin K_d (nM)†	Zn(II)-PPIX IC_{50} (nM)*	Zn(II)-PPIX K_d (nM)†	Zn(II)-PPIX K_d (nM)‡	PPIX K_d (nM)†	Co(III)-PPIX K_d (nM)†
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) × 10^3	N/A	3 (2 – 4) × 10^2	N/A	7 (3–11) × 10^3	24 (21 – 26)
BSA	2 (1–5) × 10^3	N/A	3 (2 – 4) × 10^2	N/A	11 (6–14) × 10^3	70 (50 – 80)

Values in brackets are 95% confidence intervals

* Calculated from fluorescence measurements in competition with BSA

† Calculated from absorbance data

‡ Calculated from IC_{50} values according to Cheng and Prussoff 1973, using measured K_d = 0.3 µM for BSA binding Zn(II)PPIX and [BSA] = 15 µM

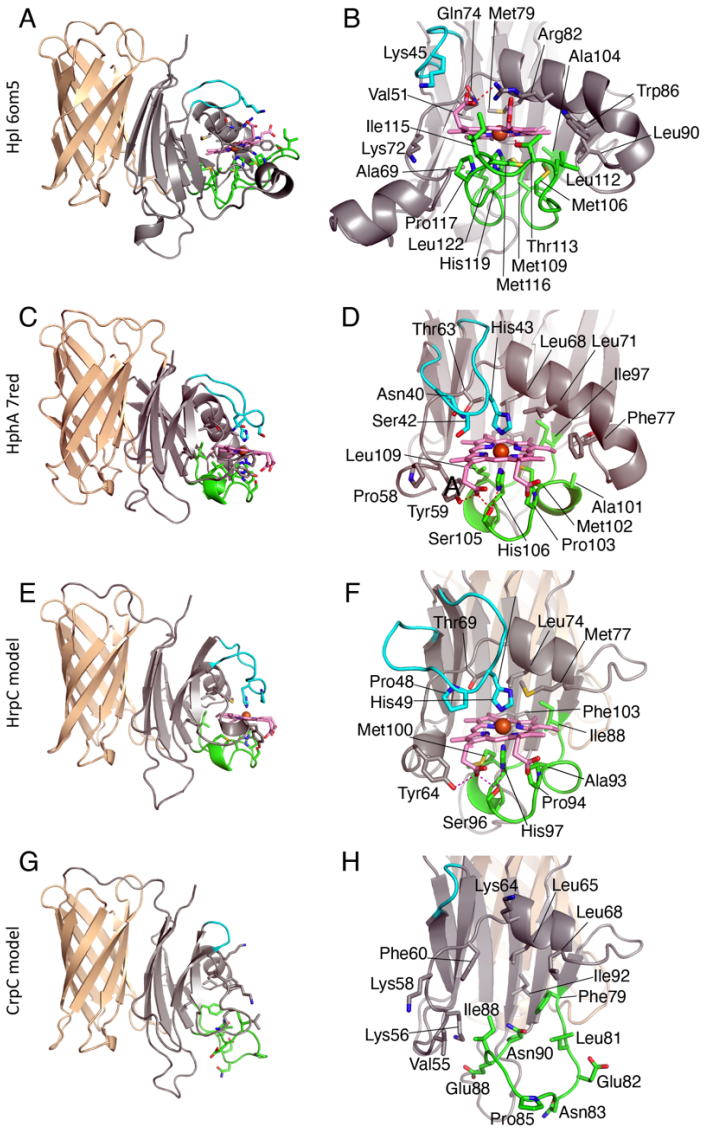


Figure 8. **Structural modelling of HrpC and CrpC**. **A)** Domain structure of Hpl (pdb 6om5) showing the N-terminal hemophilin ligand binding domain (grey, cyan, green) and C-terminal 8-stranded β-barrel domain (BBD; tan color). **B)** Heme binding site of Hpl showing structural elements that contribute to heme binding including the β2-β3 loop (cyan), β3-β4 α-helix that carries Arg82 and Trp86, and the β5-β6 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 β2-β3 loop and His106 in the β4-β5 loop. **E-F)** Structural model of *X. nematophila* HrpC based on the alignment in Figure 10, modelled with bound heme. The heme binding site has features similar to HphA, including bis-histidyl heme coordination. **G-H)** Modelled structure of *X. cabanillasii* CrpC without a heme ligand. The precise conformation of the β5-β6 loop cannot be modelled by threading due to the lack of a suitable template.

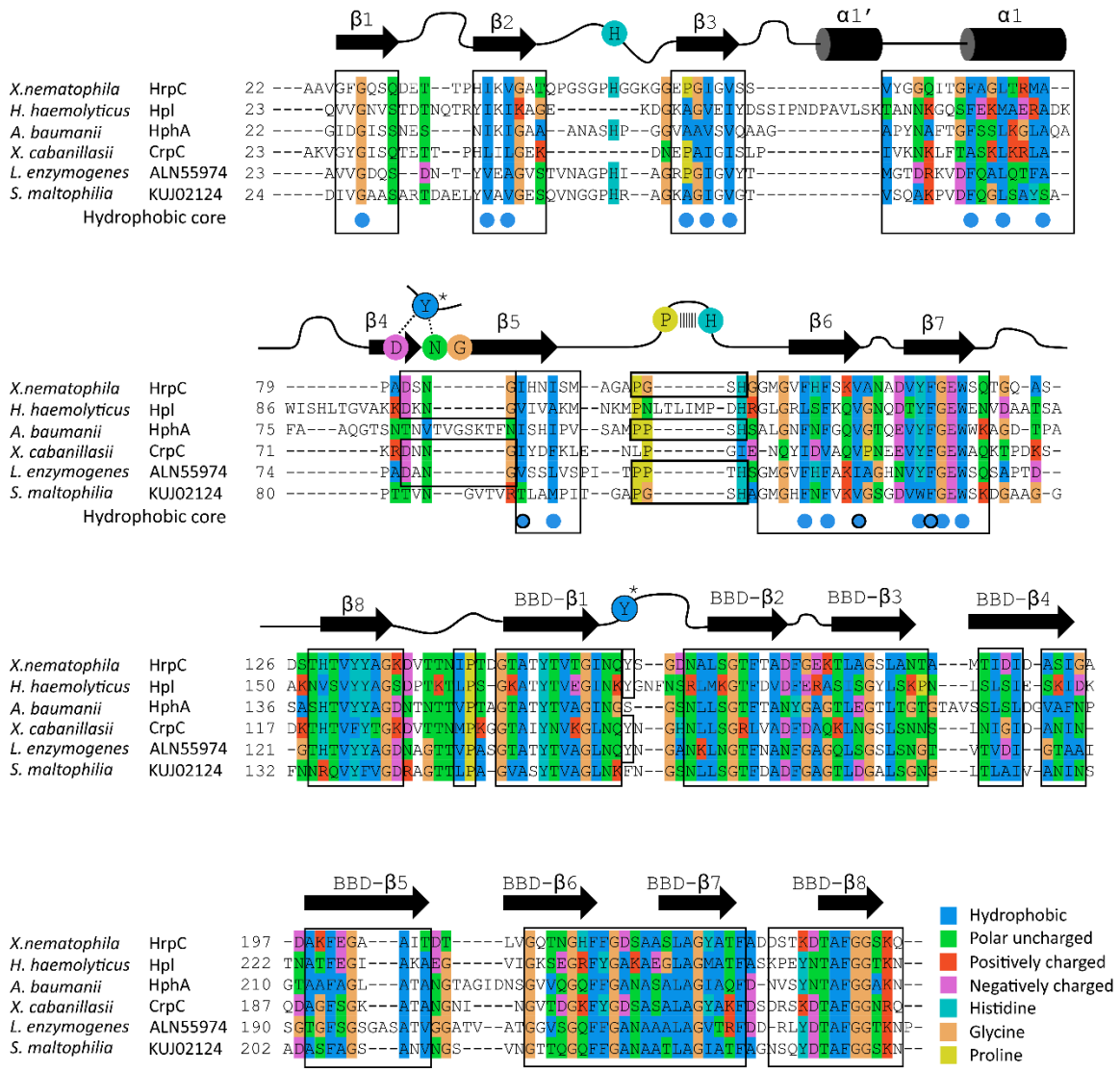


Figure 9. **Aligned sequences of representative members from hemophilin subclusters defined by sequence similarity network analysis.** The alignment is a composite of information including structure-based alignment of *H. haemolyticus* Hpl and *A. baumannii* HphA sequences and sequence-based alignments 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 modelling (see Figure 8 and supplementary Figure 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 Figure 10) shown in colored circles. Boxes indicate regions where query sequences could be modelled 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.

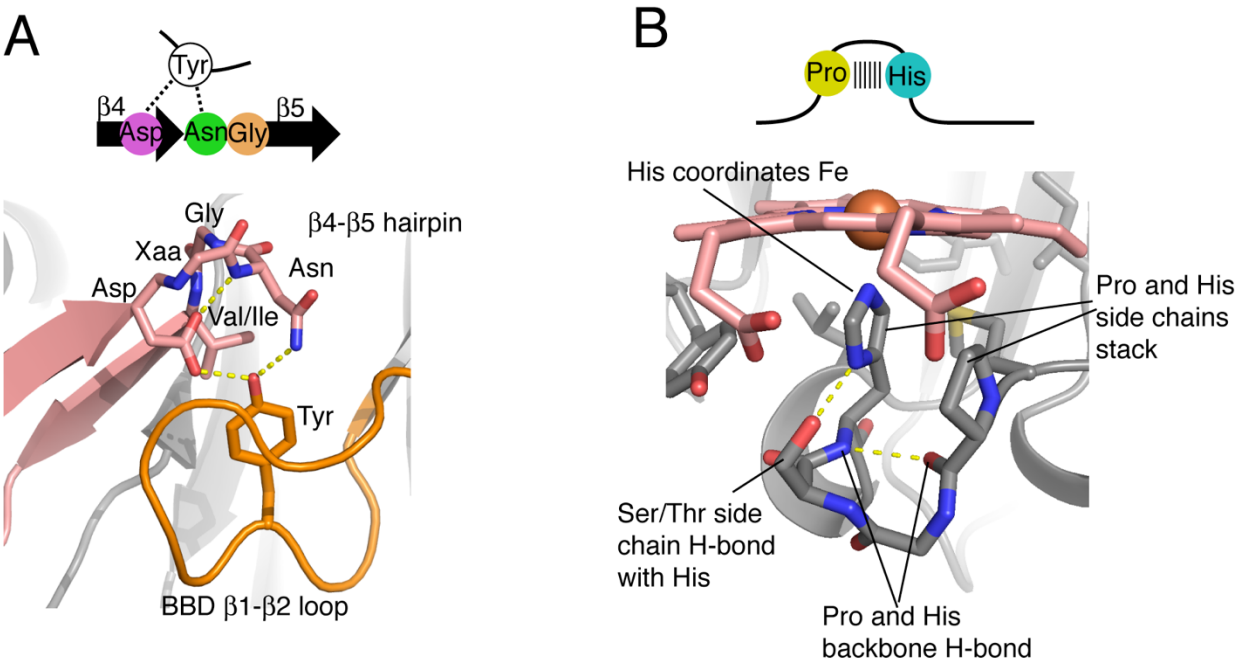
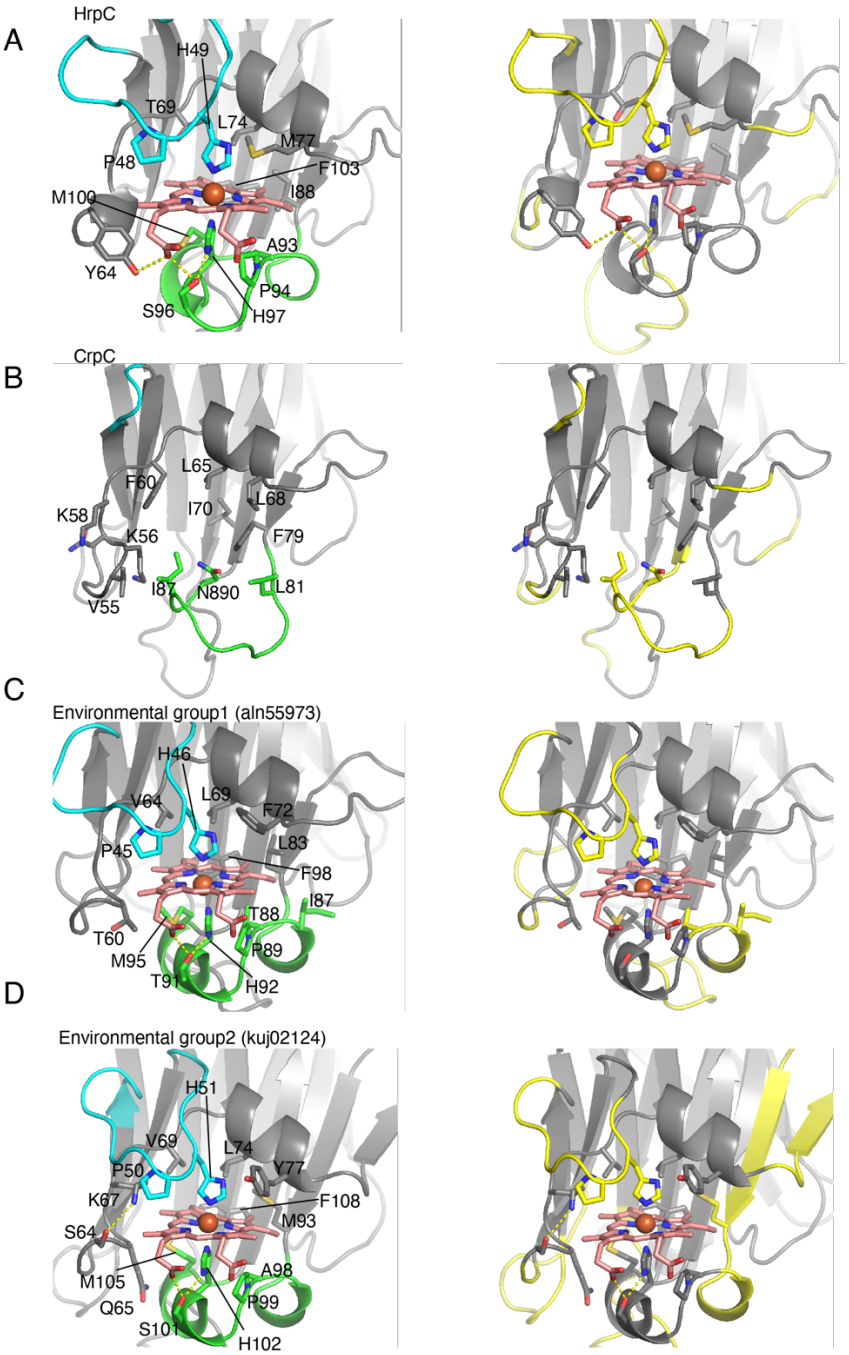


Figure 10. **Conserved sequence-structure motifs. A)** DXNG[V/I] motif structure in the β4-β5 hairpin. **B)** PX[S/T]H motif incorporating the heme-ligating His in the β5-β6 loop.



Supplementary Figure S6. **Structural models of hemophilin homologues.** Structural model of the hemophilin ligand binding domain are shown for sequences in Figure 9. In the left-hand panels, residues of the β2-β3 (cyan) and β4-β5 (green) loop are shown as in Figure 8. In the right-hand panels, regions of the structure/sequence that do not have sequence-similar threading templates from either *H. haemolyticus* Hpl or *A. baumannii* HphA are colored yellow.

Supplemental file 1: **SubclusterRodeoResultsAndNetworkNodeTable.xlsx** Node information from a sequence similarity network of predicted unlipidated T11SS-dependent cargo proteins generated using EFI-EST, alongside the results of a co-occurrence analysis of hemophilin family subclusters generated using RODEO.
Supplemental file 2: **WesternImages.xlsx** Images of immunoblots performed to demonstrate secretion of hemophilin homologs and chimeric hemophilin acquired using a Li-cor odyssey at 700nm.
Supplemental file 3: **StrainAndPrimerTable.xlsx** Collated descriptions of the bacterial strains utilized within this study and the oligonucleotides used for cloning and sequencing.