Comparison of insect and human cytochrome b561 proteins: Insights into candidate ferric reductases in insects

Jessica D. Holst[1], Laura G. Murphy[1], Maureen J. Gorman[2], Emily J. Ragan[1]*

[1]Department of Chemistry and Biochemistry, Metropolitan State University of Denver, Denver, Colorado, United States of America

[2]Department of Biochemistry and Molecular Biophysics, Kansas State University, Manhattan, Kansas, United States of America

*Corresponding author

E-mail: eragan@msudenver.edu (EJR)

Abstract

Cytochrome b561 (cytb561) proteins comprise a family of transmembrane oxidoreductases that transfer single electrons across a membrane. Most eukaryotic species, including insects, possess multiple cytb561 homologs. To learn more about this protein family in insects, we carried out a bioinformatics-based investigation of cytb561 family members from nine species representing eight insect orders. We performed a phylogenetic analysis to classify insect cytb561 orthologous groups. We then conducted sequence analyses and analyzed protein models to predict structural elements that may impact the biological functions and localization of these proteins, with a focus on possible ferric reductase activity. Our study revealed three orthologous groups, designated CG1275, Nemy, and CG8399, and a fourth group of less-conserved genes. We found that CG1275 and Nemy proteins are similar to a human ferric reductase, duodenal cytochrome b561 (Dcytb), and have many conserved amino acid residues that function in substrate binding in Dcytb. Notably, CG1275 and Nemy proteins contain a conserved histidine and other residues that play a role in ferric ion reduction by Dcytb. Nemy proteins were distinguished by a novel cysteine-rich cytoplasmic loop sequence. CG8399 orthologs are similar to a putative ferric reductase in humans, stromal cell-derived receptor 2. Like other members of the CYBDOM class of cytb561 proteins, these proteins contain reeler, DOMON, and cytb561 domains. *Drosophila melanogaster* CG8399 is the only insect cytb561 with known ferric reductase activity. Our investigation of the DOMON domain in CG8399 proteins revealed a probable heme-binding site and a possible site for ferric reduction. The fourth clade includes a group of proteins with a conserved "KXXXXKXH" non-cytoplasmic loop motif that may be a substrate binding site and is present in a potential ferric reductase, human tumor suppressor cytochrome b561. This study provides a foundation for future investigations of the biological functions of cytb561 genes in insects.

Introduction

39 The cytochrome b561 (cytb561) proteins are a family of transmembrane oxidoreductases that
40 function as electron transporters by accepting a single electron from a cytoplasmic reducing agent –
41 generally assumed to be ascorbate – and transferring it across a membrane to reduce a non-cytoplasmic
42 substrate [1,2]. Known non-cytoplasmic substrates (electron acceptors) include ferric (Fe^{3+}) ions, which
43 are reduced to ferrous (Fe^{2+}) ions, and monodehydroascorbate (MDHA), which is reduced to ascorbate
44 [2]. Animal and plant species possess multiple cytb561 homologs, some with ubiquitous expression and
45 others with restricted expression patterns [1–3]. Depending on the particular protein, cytb561s can be
46 located in the plasma membrane, the membranes of organelles in the secretory or endocytic pathways,
47 or multiple membrane locations [4–7].

48 The biological functions of most cytb561s are uncertain, although some have established roles in
49 iron reduction or ascorbate regeneration [1]. The well-characterized human duodenal cytochrome b561
50 (Dcytb) is a ferric reductase that facilitates iron uptake at the duodenal brush border, and it is also
51 suspected to play other physiological roles [8–10]. The structurally well-characterized *Arabidopsis*
52 *thaliana* (plant) cytochrome b561 B-paralog (AtCytb561-B) has in vitro ferric-chelate reductase activity,
53 but its biological functions are unknown [11–13]. Other cytb561 family members with in vitro ferric
54 reductase activity include mammalian stromal cell-derived receptor 2 (SDR2), tumor suppressor 101F6
55 (TScytb), chromaffin granule cytb561 (CGcytb), and lysosomal cytb561 (Lcytb), as well as *Drosophila*
56 *melanogaster* (fly) CG8399, and *A. thaliana* AtCytb561-A [6,11,12,14,15]. The physiological significance
57 of these proteins' ferric reductase activity is not clear; for example, although CGcytb can reduce ferric-
58 chelates in vitro, it is thought to function in chromaffin granules by reducing MDHA to produce
59 ascorbate, which is then used by dopamine β-hydroxylase (DBH) in the synthesis of norepinephrine
60 [16,17]. It seems likely that at least some cytb561s are multifunctional proteins, and that substrate
61 specificity may be influenced by the substrate availability where a particular cytb561 is located.

62 To date, two cytb561 structures have been solved: human Dcytb and plant AtCytb561-B [13,18].
63 These structures confirmed many predicted cyb561 structural elements. Cytb561s have six
64 transmembrane helices, including four consecutive helices that form a cytb561 core domain [2,13,18].
65 Inside the core domain are two b-type heme groups located near each side of the membrane, which are
66 coordinated by four universally conserved histidine residues: the first- and third-conserved histidines
67 (H1, H3) coordinate the non-cytoplasmic heme group, while the second- and fourth-conserved histidines
68 (H2, H4) coordinate the cytoplasmic heme group [2,19].

69 Most cytb561s are single-domain proteins in which the core domain encompasses
70 transmembrane helices two through five (TM2 through TM5) [2,13,18]. An exception is the CYBDOM
71 proteins, which are multidomain cytb561s characterized by having at least one extracellular DOMON
72 domain upstream of the cytb561 domain [1]. The CYBDOM cytb561 core domain encompasses TM1
73 through TM4, rather than TM2 through TM5 [1]. Single-domain cytb561s are oriented in a membrane
74 with their N- and C-termini exposed to the cytoplasm [2]. In contrast, CYBDOMs have an extracellular-
75 facing N-terminus [1], and, given that they have an even number of transmembrane regions, must also
76 have an extracellular-facing C-terminus. Single domain cytb561s and CYBDOMs have three core domain
77 loops that occur in the following order: cytoplasmic, non-cytoplasmic, cytoplasmic [1]. The function of
78 the DOMON domain(s) in CYBDOMs is unknown, but some members of the DOMON domain superfamily
79 bind sugar or heme, and at least one heme-binding DOMON domain is involved in interdomain electron
80 transfer [20]. Some CYBDOMs also have an extracellular, N-terminal adhesion module, such as the reeler
81 domain in SDR2 and *D. melanogaster* CG8399 [1]. The function of the reeler domain in CYBDOMs is not
82 known.

83 Our study aims to better understand the cytb561 family in insects, and we are particularly
84 interested in the possibility that insect cytb561 proteins play a role in cellular iron uptake. Iron
85 homeostasis in insects is poorly understood, and mechanisms of cellular iron uptake are still unknown

86 [21–23]. Most iron in biological systems is in the insoluble ferric form, but the known iron transporters
87 in animals are specific for ferrous ions; thus, for iron to enter an animal cell through a metal transporter,
88 it typically is first reduced to its soluble, ferrous state [24–26]. In mammals, the confirmed ferric
89 reductases are Dcytb and various STEAP proteins [8,27–29]. Insects do not have STEAP homologs [21];
90 therefore, members of the cytb561 family are the most promising ferric reductase candidates in insects.

91 Surprisingly little is known about insect cytb561 proteins. Only one, *D. melanogaster* Nemy, has
92 been studied in depth [4,30]. Nemy participates in a peptide amidation process in neuroendocrine
93 neurons and influences olfactory learning and memory [4,30]. The *D. melanogaster* CYBDOM, CG8399, is
94 reducible by ascorbate and has in vitro ferric reductase activity, but its biological function is unknown
95 [6,31]. Nemy and *D. melanogaster* CG1275, an uncharacterized cytb561, have been described as
96 possible ferric reductases [22,32].

97 The primary aims of this project were to 1) perform a phylogenetic analysis of cytb561 genes
98 from nine diverse insect species to identify orthologous groups, 2) compare their amino acid sequences
99 and predicted protein structures to those of better-studied human cytb561s to reveal sequence motifs
100 and structural elements that may be involved in the localization and biological functions of these
101 proteins, and 3) predict which of the orthologous groups may function as ferric reductases in insects. Of
102 the four cytb561 groups identified, the CG1275 and Nemy groups are similar to human Dcytb; the
103 CG8399 proteins are similar to human SDR2; and a subset of proteins in the fourth group is similar to
104 human TScytb. We found that CG1275 and Nemy proteins have conserved amino acid residues and
105 predicted protein structures that are consistent with ferric reductase activity. CG8399 proteins have
106 highly-conserved amino acid residues in cytoplasmic regions and the DOMON domain that may be
107 involved in substrate binding and ferric reductase activity. TScytb-like proteins have a “KXXXXKXH” motif
108 that may be involved in non-cytoplasmic substrate binding and potential ferric reductase activity.

Materials and methods

Identification of insect cytb561 protein sequences

For this study, we selected *D. melanogaster* as our model insect species, along with eight additional insect species of interest: *Anopheles gambiae* (mosquito), *Acyrthosiphon pisum* (aphid), *Apis mellifera* (bee), *Ctenocephalides felis* (flea), *Papilio xuthus* (butterfly), *Pediculus humanus corporis* (louse), *Tribolium castaneum* (beetle), and *Zootermopsis nevadensis* (termite). These species were chosen for their phylogenetic diversity and their association with high-quality genomic data.

We identified *D. melanogaster* cytb561 protein sequences using the FlyBase protein domain search tool with the query "Cytochrome b561" [33]. Of the nine resulting hits, eight possessed the four conserved heme-coordinating histidine residues characteristic of cytb561 family members: CG1275, Nemy, CG8399, CG10165, CG13077, CG13078, CG10337, and CG3592. The remaining sequence, CG6479, lacked these conserved histidines and was eliminated from our dataset.

To identify cytb561 sequences in the insect species of interest, we performed species-specific BLAST-P searches using the National Center for Biotechnology Information (NCBI) non-redundant database [34]. Eight *D. melanogaster* cytb561 amino acid sequences were used as queries: NP_728727.1 (CG1275), NP_725208.1 (Nemy), NP_611079.2 (CG8399), NP_609982.1 (CG10165), NP_609990.1 (CG13077), NP_609989.1 (CG13078), NP_609986.1 (CG10337), and NP_570039.1 (CG3592). Additional insect cytb561 sequences were identified using the NCBI Conserved Domain Database (CDD) [35]. The NCBI CDD search engine was populated with a combination of "insect species" and either "cyt_b561" or "cytochrome_b_N," and a search was executed.

Isoforms and identical hits were removed from the dataset, resulting in one unique sequence for each gene. Of the 46 non-*D. melanogaster* sequences discovered, 36 resulted from BLAST-P searches

131 and ten resulted from NCBI CDD queries. A sequence from *T. castaneum* (XP_008195477.1) was edited
132 using RNA-seq data to remove an apparently misplaced intron, resulting in the addition of 53 C-terminal
133 amino acids and a complete cytb651 core domain. The NCBI accession number and source of each
134 sequence is listed in S1 Table.

135 Three sequences with unique properties were observed: XP_314065.2 (*A. gambiae*),
136 XP_001950579.2 (*A. pisum*), and XP_001949276.1 (*A. pisum*). These three sequences were used as
137 separate queries to conduct BLAST-P searches against the NCBI non-redundant database for the class
138 Insecta to determine whether other insect species have proteins with similar sequences.

139 Defining homologous regions of protein sequences

140 Prior to performing phylogenetic and sequence analyses, we identified the homologous region
141 of each insect cytb561 protein. We used the HHpred server (Max Planck Institute Bioinformatics Toolkit
142 webserver), which is a tool for protein remote homology detection that generates a multiple sequence
143 alignment, predicts secondary structure, then creates a profile hidden Markov model that is compared
144 with a target database [36]. We evaluated each insect and outgroup sequence using the HHpred
145 PDB_mmCIF70_3_Mar server set to default parameters, including local alignment. This server
146 references the Protein Data Bank (PDB) of known protein structures for possible templates, then
147 generates results based on the probability percentage that the query shares a homologous region with
148 the template. An estimated probability percentage of > 95% indicates that homology is nearly certain
149 [36]. All insect and outgroup sequences were determined to share the highest probability percentage of
150 homology (> 99%) with regions of Dcytb (CYBRD1, NP_079119.3, PDB 5ZLG). The predicted homologous
151 region was not identical for all insect sequences; therefore, we identified the region that all insect
152 sequences have in common and defined this as the common homologous region, which corresponds to

153 Dcytb A44-T179. We used common homologous sequences for all analyses unless otherwise described.
154 The common homologous region of each sequence is listed in S1 Table.

155 To compare insect cytb561s with the six known human cytb561 proteins, we used HHpred and
156 the same methodology described above to define the common homologous regions of the following
157 human sequences: NP_079119.3 (Dcytb; CYBRD1), NP_705839.3 (Lcytb; CYB561A3), NP_001017916.1
158 (CGcytb; CYB561), NP_001278213.1 (TScytb; CYB561D2), NP_872386.1 (CYB561D1), and
159 NP_001347970.1 (SDR2; FRRS1).

160 HHPred analysis of the DOMON domains in insect CG8399 orthologs was performed. All insect
161 DOMONs were predicted to have > 99% predicted homology with regions of the cytochrome domain in
162 the *Phanerodontia chrysosporium* (fungus) cellobiose dehydrogenase (CDH) protein (PDB 1D7B) and the
163 DOMON domain in the human dopamine beta-hydroxylase (DBH) protein (PDB 4ZEL).

164 Multiple sequence alignments, sequence identities, and localization

165 Multiple sequence alignments were performed in Clustal Omega and visualized in JalView,
166 where we applied a custom color scheme and obtained consensus sequence logos and conservation
167 scores [37,38]. Accession numbers for insect sequences used in alignments are in S1 Table. Additional
168 sequences used in alignments include Dcytb and *Arabidopsis thaliana* cytochrome B561-1 (AtCytb561-B,
169 NP_198679.1, PDB 4O7G). To determine sequence identities, we used CLUSTAL Omega to generate
170 multiple sequence alignments for each insect sequence with all *D. melanogaster* and human cytb561
171 sequences, then determined pairwise sequence identities through utilization of the M-View multiple
172 alignment viewer [39].

173 Full-length insect cytb561 protein sequences were submitted to the SignalP-5.0 webserver for
174 detection of predicted signal peptides, and they were analyzed by the DEEPLOC-1.0 webserver for

175 prediction of subcellular protein localization [40,41]. Sequences were also investigated by eye to identify
176 putative late-endosomal/lysosomal targeting signals in the C-terminal portion of the protein [42].

177 Phylogenetic analysis

178 For our phylogenetic analysis, we selected an outgroup species, *Trichoplax adhaerens*, based on
179 two qualifying criteria: *T. adhaerens* belongs to an ancient animal lineage near the base of the metazoan
180 tree, and its cytb561 genes are homologous to insect cytb561 genes [43]. To identify an outgroup
181 sequence, we evaluated the cytb561 genes from *T. adhaerens* and selected the sequence with the
182 lowest identity to the *D. melanogaster* cytb561 sequences. The phylogenetic analysis was conducted
183 using the MEGA version X software [44]. First, all common homologous insect cytb561 sequences and
184 the outgroup cytb561 sequence were aligned in CLUSTAL Omega using the "Pearson/FASTA" output
185 parameter; the resulting alignment was uploaded to MEGA X, and a rooted phylogenetic tree was
186 created using the maximum likelihood (ML) method. To determine the optimal model for ML tree
187 construction, we ran the "Find Best DNA/Protein Models (ML)" tool in MEGA X, which resulted in
188 recommended parameters: the LG with freqs. (+F) substitution model, gamma distributed with invariant
189 sites (G+I) rates among sites, and five discrete gamma [45]. We then used the additional remaining
190 parameters: use all sites in cases of gaps and missing data, the Nearest-Neighbor-Interchange ML
191 heuristic method of tree inference, the default NJ/BioNJ initial tree for ML, no branch swap filters, seven
192 threads, and 1000 bootstrap replicates.

193 AlphaFold structure analysis

194 AlphaFold Monomer v2.0 pipeline structure predictions for *D. melanogaster* CG1275 isoform A
195 (AF-Q9I7U1-F1-model_v4.pdb), *D. melanogaster* Nemy isoform A (AF-Q7JR72-F1-model_v4.pdb), *D.*
196 *melanogaster* CG8399 isoform A (AF-Q8MSU3-F1-model_v4.pdb), *D. melanogster* CG13077 isoform B
197 (AF-M9PG89-F1-model_v4.pdb), and *A. gambiae* XP_3200673.4 (AF-Q7PZD0-F1-model_v4) were

198 downloaded from UniProt and visualized in ChimeraX [46–48]. We removed amino acids 1-30 from the
199 CG8399 structure prediction as they were predicted to be a signal peptide by SignalP 6.0 [49]. AlphaFold
200 structures for CG1275 and Nemy were individually overlayed with Dcytb (PDB 5ZLG) using the
201 MatchMaker feature in ChimeraX with default settings. The number of α-carbon pairs and their root-
202 mean-square deviation (RMSD) was generated by ChimeraX. CG1275 (amino acids Pro100-Ile326) and
203 Nemy (amino acids Cys52-Ser281) were displayed together with Dcytb (amino acids Tyr6-Pro230) to
204 generate figures. Individual AlphaFold models were associated with multiple sequence alignments of
205 same group proteins and colored in ChimeraX using AL2CO and a color gradient from cyan (low
206 conservation score) to magenta (high conservation score) [50]. The heme- and ascorbate-binding
207 residues were selected to calculate possible intermolecular hydrogen bonds using the H-bonds tool and
208 to predict possible direct residue interactions with the Contacts tool in ChimeraX using default settings
209 for overlap cutoff and H-bond allowance parameters. Potential hydrogen bonds were displayed in blue.
210 Coulombic surface coloring was used to visualize surface charges in the potential ion-binding pocket of
211 CG8399 [46].

212 Analysis of DOMON and TScytb sequence motifs

213 Additional sequence analyses were conducted to investigate DOMON domain sequences and a
214 "KXXXXKXH" motif in cytb561 proteins from diverse animal species. Tree of Life was referenced to
215 choose a set of species for these analyses: *Homo sapiens* (human), *Mus musculus* (mouse), *Passer*
216 *montanus* (sparrow), *Xenopus laevis* (frog), *Danio rerio* (zebrafish), *Octopus sinensis* (octopus), *Actinia*
217 *tenebrosa* (sea anemone), *Helobdella robusta* (leech), and *Acanthaster planci* (starfish) [51]. To identify
218 DOMON-containing sequences, the *D. melanogaster* CG8399 sequence (NP_611079.2) was used as a
219 BLAST-P query to search NCBI non-redundant databases; to identify "KXXXXKXH"-containing sequences,
220 the query was *A. gambiae* sequence XP_320673.4. CLUSTAL Omega was then used to generate

221 sequence alignments that allowed us to evaluate amino acid conservation of the DOMON domains and
222 "KXXXXKXH" motifs and to determine the placement of the "KXXXXKXH" motif within the protein [39].

223 Results and Discussion

224 Phylogenetic analysis of insect cytb561 proteins

225 Insects have four distinct cytb561 groups

226 While a limited number of insect cytb561 sequences have been included in broader
227 phylogenetic studies [3,19], we were interested in classifying the cytb561 family across diverse insect
228 species. We identified 54 genes that encode cytb561 family members from nine insect species in eight
229 orders: Blattodea (*Z. nevadensis*), Coleoptera (*T. castaneum*), Diptera (*A. gambiae* and *D. melanogaster*),
230 Hemiptera (*A. pisum*), Hymenoptera (*A. mellifera*), Lepidoptera (*P. xuthus*), Psocodea (*P. humanus*
231 *corporis*), and Siphonaptera (*C. felis*) (S1 Table). All investigated species have between three *(P.*
232 *humanus corporis*) and nine (*P. xuthus*) cytb561 genes (S1 Table).

233 Using the HHpred PDB server for protein remote homology detection, we defined a homologous
234 region (core domain) shared by all of the insect cytb561s [36] (S1 Table). The ~130 amino acid long core
235 domain sequences were then used to generate a multiple sequence alignment (S1 Fig). From this
236 multiple sequence alignment, we generated a phylogenetic tree and found strong support for four
237 cytb561 groups: three orthologous groups and a fourth group that lacks one-to-one orthologs (Fig 1).
238 For simplicity, we will refer to all insect sequences that cluster in the three orthologous groups
239 according to the representative *D. melanogaster* protein name: CG1275, Nemy, or CG8399. The
240 remaining 25 sequences will be referred to as Group 4.

241

242 **Fig 1. Phylogenetic tree of insect cytb561 proteins.** A phylogenetic analysis of cytb561s from nine
243 species of insects (*A. gambiae [Ag]*, *A. mellifera [Am]*, *A. pisum [Ap]*, *C. felis [Cf]*, *D. melanogaster [Dm]*,
244 *T. castaneum [Tc]*, *P. humanus corporis [Ph]*, *P. xuthus [Px]*, and *Z. nevadensis [Zn])* was performed with
245 the maximum likelihood method. An outgroup sequence from *T. adhaerens (Ta)* was used for rooting
246 the tree. Bootstrap values greater than 70% are shown. Three orthologous groups (with one-to-one
247 orthologs in most or all species) were identified: CG1275, Nemy, and CG8399. The remaining genes
248 formed a fourth group that has no one-to-one orthologs but includes many species-specific gene
249 expansions. The bootstrap values associated with the four clades are marked with red circles. A branch
250 length scale bar indicates one substitution per site. Of the Group 4 sequences, seven are more similar to
251 human TScytb than any *D. melanogaster* cytb561s and are referred to as subgroup 4A, whereas the
252 remaining Group 4 proteins belong to subgroup 4B. (Note that the 4A and 4B subgroups are based on
253 sequence comparisons rather than phylogenetic relationship.)

254 CG1275 and Nemy

255 The CG1275 and Nemy groups are closely related, as evidenced by bootstrap support of a
256 CG1275-Nemy cluster and the short branch lengths for the CG1275 and Nemy groups (Fig 1). CG1275
257 proteins were identified for eight of the nine insect species (Fig 1; S1 Table). We did not identify a
258 CG1275 ortholog in the lepidopteran insect *P. xuthus*, and we were unable to find a CG1275 sequence in
259 any other species from the order Lepidoptera (butterflies and moths). Of the species analyzed, *A. pisum*
260 was the only insect with a CG1275 gene expansion (Fig 1; S1 Table). We identified Nemy proteins in
261 eight of the species (Fig 1; S1 Table), but we did not find a Nemy sequence from *A. mellifera* or any other
262 species in the order Hymenoptera (bees, wasps, ants, and similar insects). A previously published
263 phylogenetic analysis of eukaryotic cytb561s suggested that CG1275 and Nemy from *D. melanogaster*
264 and *A. gambiae* may be orthologous to human Dcytb, CGcytb, and Lcytb, but a lack of statistical support

265 for the clade makes the phylogenetic relationship of CG1275 and Nemy with Dcytb, CGcytb, and Lcytb
266 uncertain [19].

267 ### CG8399

268 CG8399 is the only orthologous group with sequences from all nine insect species. This cluster
269 includes ten proteins, nine of which contain multiple domains – reeler, DOMON(s), and cytb561 –
270 consistent with some members of the CYBDOM protein family (Fig 1; S1 Table) [1]. *A. gambiae* was the
271 only species with a second CG8399-like protein (XP_314065.2); this sequence differs from other
272 CG8399s in that it is a single-domain cytb561 lacking both reeler and DOMON domains. A previously
273 published phylogenetic analysis of eukaryotic cytb561s demonstrated that *D. melanogaster* and *A.*
274 *gambiae* CG8399 proteins are orthologous to SDR2 in humans, other animals, and plants [19].

275 ### Group 4

276 The remaining 25 insect genes populate a large clade (Group 4) with a bootstrap value of 92 (Fig
277 1). Group 4 has multiple species-specific clusters with significant bootstrap values, suggesting that
278 Group 4 genes were more likely than other insect cytb561 genes to undergo species-specific expansions
279 (Fig 1). We did not identify any Group 4 genes in *P. humanus corporis* or *A. pisum* (Fig 1).

280 ## Sequence identity between human and insect cytb561s

281 Human cytb561s have been more widely studied than insect cytb561s, and human Dcytb is one
282 of only two cytb561 structures that have been solved [18]. We were interested in determining the
283 sequence similarities between the 54 insect cyb561s and the six human cytb561s (Dcytb, CGcytb, Lcytb,
284 SDR2, TScytb, and CYB561D1) to learn which of the better-characterized human cytb561s is most similar
285 toeach of the insect cytb561s. For this analysis, we identified the common homologous region of each
286 insect and human cytb561 sequence, then determined their pairwise identities. CG1275 and Nemy

proteins had highest sequence identity with Dcytb, CGcytb, or Lcytb (32.4%-49.3%), and CG8399s had highest sequence identity with SDR2 (36.5%-43.4%) (Table 1). These results suggest that CG1275 and Nemy proteins may have physiological similarities with Dcytb, CGcytb, and/or Lcytb, and are consistent with the finding that CG8399s are orthologous to SDR2 [19]. Of the 25 Group 4 cytb561s, 23 had highest sequence identity with TScytb (21; 18.8%-33.1%) or CYB561D1 (2; 22.0%-25.7%) (Table 1). Seven Group 4 proteins are distinct from all other insect cytb561s in that they had higher sequence identity with a human cytb561 (TScytb) than with any of the *D. melanogaster* cytb561s; we will refer to this subset of proteins as Group 4A and the remaining Group 4 proteins as Group 4B (Table 1; Fig 1; S2 Table). A previous phylogenetic study found that the *A. gambiae* gene in Group 4A clustered with human and other TScytbs rather than other insect-specific sequences, but there was a lack of bootstrap support for this cluster; thus, the phylogenetic relationship between Group 4A genes and TScytb is unclear [19].

Table 1. Sequence identity between insect and human cytb561s.

			Human Sequences					
Species	**Group**	**Accession Number**	**CYBRD1 (Dcytb)**	**CYB561A3 (Lcytb)**	**CYB561 (CGcytb)**	**CYB561D2 (TScytb)**	**CYB561D1**	**FRRS1 (SDR2)**
Dm	CG1275	NP_728727.1	**45.7**[1]	40.6	39.3	21.7	20.3	16.7
Ag	CG1275	XP_315519.4	43.2	43.2	**43.4**	23.2	20.4	18.6
Ap	CG1275	NP_001155374.1	32.6	34.1	**40.7**	17.9	23.6	15.3
Ap	CG1275	NP_001280323.1	31.7	29.6	**32.4**	15.2	18.6	15.5
Ap	CG1275	XP_001950854.1	37	37	**41.5**	22	25.5	18.1
Ap	CG1275	XP_003246890.1	30.4	31.2	**33.3**	17.7	20.6	11.7
Am	CG1275	XP_006572086.1	**49.3**	44.9	43	22	20.6	20.1
Cf	CG1275	XP_026462102.1	39.9	38.5	**42.1**	21.2	21.2	18.9
Ph	CG1275	XP_002426701.1	**46.8**	43.9	44.1	22.5	21.1	20.4
Tc	CG1275	XP_008194670.1	44.9	42.8	**45.9**	24.5	20.3	16.1
Zn	CG1275	XP_021935166.1	42	41.3	**44.4**	21.3	20.6	20.8
Dm	Nemy	NP_725208.1	35	**35.7**	32.1	19.7	21.8	16.4

Ag	Nemy	XP_314126.2	38.6	**39.3**	35.8	19	23.2	19.2
Ap	Nemy	XP_001949276.1	33.6	32.9	**36.6**	23.1	27.3	17
Cf	Nemy	XP_026473332.1	37.3	39.4	**39.6**	21.5	29.9	16.9
Px	Nemy	NP_001298968.1	37.1	37.9	**38.7**	23.2	19	16.4
Ph	Nemy	XP_002430226.1	35.2	**36.6**	35.3	20.8	26.4	15.5
Tc	Nemy	XP_008198104.1	36.6	**39.4**	36.7	22.9	29.2	14.9
Zn	Nemy	XP_021939496.1	**39.2**	38.5	38.6	22.8	28.3	14.6
Dm	CG8399	NP_611079.2	18.4	17	17.1	13.9	13.9	**37**
Ag	CG8399	XP_314065.2	18.3	20.4	19.9	16.1	15.4	**36.8**
Ag	CG8399	XP_314066.4	18.3	19	19.1	11	15.2	**41.2**
Ap	CG8399	XP_001950579.2	20	16.6	18.1	10.1	14.2	**36.5**
Am	CG8399	XP_396579.3	15.6	21.3	15.7	9.7	11.1	**41.2**
Cf	CG8399	XP_026481553.1	19	17.6	17	11.7	16.6	**43.4**
Px	CG8399	XP_013164083.1	15.6	19.1	18.6	16	17.4	**39.4**
Ph	CG8399	XP_002423127.1	16.4	19.9	17.2	9.4	14.1	**40.1**
Tc	CG8399	XP_015836986.1	17.6	17.6	15.6	11.7	15.2	**38.2**
Zn	CG8399	XP_021919699.1	15.5	18.3	17	12.4	15.2	**41.9**
Dm	Group 4B	NP_609982.1	17.1	17.1	18.7	**24.8**	22	10.6
Dm	Group 4B	NP_609990.1	14.3	16.4	15.8	**22.7**	20.6	14.2
Dm	Group 4B	NP_609989.1	18.4	14.2	16.4	**19.7**	16.9	14.2
Dm	Group 4B	NP_609986.1	17.3	20.9	**23.2**	21.1	18.3	11.2
Dm	Group 4B	NP_570039.1	18.8	22.2	19.6	**24.5**	21.8	11.4
Ag	Group 4A	XP_320673.4[2]	21.5	20.1	23.1	**33.1**	31.7	19.4
Ag	Group 4B	XP_001238089.2	16.4	14.3	12.9	25	**25.7**	12.6
Ag	Group 4B	XP_552919.3	18.6	19.3	18	**27.3**	21.6	16.1
Am	Group 4A	XP_001122176.1[2]	18.6	17.9	13.7	**27.9**	19.3	17.5
Am	Group 4B	XP_003249671.1	20.4	21.1	19.1	**25.2**	20.9	16.2
Cf	Group 4A	XP_026469480.1[2]	22.3	20.1	19.6	**31.7**	27.3	15.5
Cf	Group 4A	XP_026469242.1[2]	22.3	18.7	15.2	**28.8**	26.6	18.2
Px	Group 4A	XP_013174961.1[2]	20	16.4	23	**33.1**	26.6	14.7
Px	Group 4B	XP_013172799.1	15.8	12.9	16.7	**26.2**	20.6	10
Px	Group 4B	XP_013172813.1	17.3	15.1	16.8	**25.5**	22	12.9
Px	Group 4B	XP_013162691.1	17.1	19.3	17.3	**22**	20.6	13.4
Px	Group 4B	XP_013162686.1	14.3	13.6	12.2	**19.9**	16.3	12.8

Px	Group 4B	XP_013172802.1	17.4	15.2	15.3	**18.8**	18.1	12
Px	Group 4B	XP_013172804.1	20.7	**22.1**	18	19.9	21.3	14.9
Tc	Group 4A	XP_008201603.1[2]	23	20.1	19.6	**33.1**	23.7	18.3
Tc	Group 4B	XP_008195477.1[3]	16.3	19.1	16.4	**26.2**	22.7	16
Tc	Group 4B	XP_015837014.1	20.7	20	18	**26.8**	21.8	11.2
Tc	Group 4B	XP_974632.1	17.9	18.6	16.5	21.3	**22**	14.8
Tc	Group 4B	XP_974652.1	20.9	18.7	15.2	**26.6**	24.5	14.2
Zn	Group 4A	XP_021927739.1[2]	20.9	18	18.8	**32.4**	30.9	14.8

[1]Sequence identity (as a percent) between the homologous region of each insect cytb561 and each human cytb561. The highest identity in each row is shown with a shaded cell and in bold text. Group 4A sequences are shaded in pale orange and all other sequences are shaded in pale blue.
[2]Group 4A insect sequences with higher identity to a human cytb561 than a *D. melanogaster* cytb561.
[3]Edited to correct a gene prediction error.

Analysis of insect cytb561 domains

Locations of conserved amino acid residues

We used amino acid sequence alignments to identify conserved amino acid residues in insect cytb561 domains. To facilitate comparisons of residues from diverse sequences, we denote residues as the number of positions they are upstream (-) or downstream (+) of one of the four conserved heme-coordinating histidine residues; for example, the residue immediately upstream of H1 is referred to as H1-1. To identify amino acid residues that are conserved in all four insect cytb561 groups, we aligned the ~130 amino acid core domain sequences of all 54 insect species and found that only the heme-coordinating histidines, H1-H4, were completely conserved (S1 Fig). Note that for the CG1275, Nemy, and Group 4 proteins, H1-H4 are in TM2-TM5, whereas in the CG8399 proteins, H1-H4 are in TM1-TM4. In addition to the four positions containing conserved histidines, 11 positions have highly-conserved residues (with Jalview conservation scores of 8 or 9), including seven with small residues (Gly, Ala, and Ser), and four with larger hydrophobic residues (S1 Fig). Five of the seven small residue positions were previously noted [3], including two that are present in 106 eukaryotic cytb561s (H1+7, H4+4) [19]. To

319 identify amino acid conservation within specific insect cytb561 groups, we separated the core domain
320 alignment into five group-specific alignments and then used Jalview to generate group-specific
321 consensus logos (Fig 2) [38]. These consensus logos reveal patterns of conservation within groups and
322 allow residues at specific locations to be compared across groups.

323

324 **Fig 2. Group consensus logos for the cytb561 core domain.** Group-specific alignments were used to
325 create group consensus sequences in Jalview. The four conserved histidines are labeled H1-H4 with
326 yellow highlighting for the non-cytoplasmic histidines and blue highlighting for the cytoplasmic
327 histidines. The positions of amino acid residues of interest are labeled relative to the nearest conserved
328 histidine. If the corresponding residue in Dcytb was predicted to contact a heme or ascorbate, that
329 position is indicated by a symbol: non-cytoplasmic heme (square), non-cytoplasmic ascorbate (upside-
330 down triangle), cytoplasmic heme (red diamond), or cytoplasmic ascorbate (red triangle). The Jalview
331 consensus rows show the most common residue at that position listed below the logo; a + is used where
332 there are two or more residues with the highest percentage and a – is used to show gaps.

333 To analyze the conserved amino acids in the context of the protein structure, we used ChimeraX
334 to overlay a protein alignment of the CG1275, Nemy or CG8399 groups with its representative *D.*
335 *melanogaster* AlphaFold protein model (Fig 3). For Group 4B proteins, we selected the *D. melanogaster*
336 CG13077 AlphaFold model (Fig 3). Since *D. melanogaster* does not have a Group 4A protein, we used the
337 *A. gambiae* Group 4A (XP_320673.4) AlphaFold model (Fig 3). The degree of conservation is shown by
338 gradient coloring. By analyzing conservation information from the structural models as well as the
339 group-specific consensus logos, we determined that the insect cytb561 proteins, with the exception of
340 the Group 4B proteins, have relatively high group sequence conservation in the regions near non-
341 cytoplasmic heme-coordinating H1 and H3 (Figs 2 and 3). Additionally, CG8399 proteins have relatively
342 high conservation in the regions preceding the cytoplasmic heme-coordinating H2 and H4 (Figs 2 and 3).

343

344 **Fig 3. Insect cytb561 models with sequence alignment overlay demonstrate residue conservation.**
345 AlphaFold models of mature proteins are shown. Sequence conservation was determined by ChimeraX
346 using AL2CO conservation scores. The model is gradient colored with high conservation (+1.5 score) in
347 magenta and low conservation (-1.5 score) in cyan; sequence gaps are shown in tan. Conserved
348 histidines H1-H4 (not all visible in all panels) are shown in stick form with nitrogen in blue. The models
349 are oriented with the cytoplasmic side of the protein at the bottom. Note that the amino- and carboxyl-
350 termini for all models except the CG8399 model are near the bottom of the panels. Locations of H1-4
351 are indicated. Scale bars depict 10 Å. A) *D. melanogaster* CG1275 (Lys95-Glu340) aligned with the
352 CG1275 sequences. B) *D. melanogaster* Nemy (Lys40-Ile290) aligned with the Nemy sequences. C) *D.*
353 *melanogaster* CG8399 (Ser402-Gly645) aligned with the multi-domain CG8399 sequences. The reeler
354 and DOMON domains are omitted. D) *A. gambiae* XP_320673.4 (all residues) aligned with the Group 4A
355 sequences. E) *D. melanogaster* CG13077 (all residues) aligned with the Group 4B sequences.

356 General comparisons of insect cytb561 proteins with Dcytb and
357 AtCytb561-B

358 This is the first large scale analysis of cytb561 proteins since X-ray crystallography studies
359 identified heme- and ascorbate-binding residues in human Dcytb and *A. thaliana* AtCytb561-B, and a
360 metal ion-coordinating residue in Dcytb [13,18]. We wanted to leverage this structural information to
361 predict whether conserved residues in insect cytb561s might participate in heme, ascorbate, and/or
362 iron-binding. Amino acid sequence alignments and structural models were used to compare insect
363 cytb561 proteins with Dcytb and Atcyb-B. Initial comparisons were restricted to CG1275 and Nemy
364 proteins because of their high similarity to Dcytb and AtCytb561-B; once key residues were identified in
365 CG1275 and Nemy, we used sequence alignments to identify the corresponding residues in the CG8399

366 and Group 4 proteins. References to specific residue numbers are based on Dcytb unless otherwise
367 noted.

368 We first created an alignment of Dcytb, AtCytb561-B, CG1275, and Nemy sequences (Fig 4).
369 From this alignment, which extends beyond the core domain, we identified 23 completely conserved
370 residues and an additional 23 residues that are highly conserved across Dcytb, AtCytb561-B, CG1275,
371 and Nemy sequences (Fig 4). To identify any previously unspecified Dcytb residues that are likely to be
372 important for heme or ascorbate binding, we used ChimeraX to determine predicted protein contacts
373 with each heme and ascorbate group in the Dcytb ascorbate-bound structure (PDB 5ZLG) (Fig 4; S3
374 Table). (Note that Dcytb was selected for this analysis rather than AtCytb561-B for multiple reasons:
375 insects are more closely related to humans than plants, Dcytb is a ferric reductase, and the Dcytb crystal
376 structure contains a metal ion.)

377

378 **Fig 4. Alignment of Dcytb, AtCytb561-B, Nemy, and CG1275 protein sequences.** The two cytb561
379 sequences with known structure, human Dcytb (PDB 5ZLG) and *A. thaliana* AtCytb561-B (PDB 4O7G),
380 were aligned with CG1275 and Nemy sequences from *A. gambiae (Ag)*, *A. mellifera (Am)*, *A. pisum (Ap)*,
381 *C. felis (Cf)*, *D. melanogaster (Dm)*, *T. castaneum (Tc)*, *P. humanus corporis (Ph)*, *P. xuthus (Px)*, and *Z.*
382 *nevadensis (Zn)*. (Note that only the *A. pisum* CG1275 sequence most similar to Dcytb was included.)
383 Alpha helices from crystal structures (Dcytb, AtCytb561-B) or predicted in AlphaFold models (Nemy,
384 CG1275) are shown in boxes: helix 1 carboxyl end (gray), helix 2 (red), helix 3 (yellow), helix 4 (green),
385 and helix 5 (blue). Conserved residues are noted under the alignment (* indicates completely conserved,
386 the four strictly conserved histidines are labeled with their number rather than *, and ^ indicates highly
387 conserved with a Jalview conservation score of 8 or 9). Predicted Dcytb heme-contacting residues are
388 labeled with a square (non-cytoplasmic heme) or diamond (cytoplasmic heme), and ascorbate-

389 contacting residues are labeled with an upside-down triangle (non-cytoplasmic ascorbate) or triangle
390 (cytoplasmic ascorbate). A blue vertical line indicates the position of 51 amino acids found only in *A.*
391 *pisum* XP_001949276.1, which were omitted from the figure. Coloring by amino acid residue: basic
392 residues in blue (Arg) or cyan (Lys); acidic residues (Asp and Glu) in dark pink; amide residues (Asn and
393 Gln) in light pink; hydroxylic residues (Ser and Thr) in pale purple; aromatic residues in light peach (Phe
394 and Tyr) or dark peach (Trp); sulfur-containing residues in dark green (Cys) or light green (Met); His in
395 orange; Pro in yellow; Ile, Leu, and Val in pale yellow; and Ala and Gly in white.

396 Next, we compared structural models of *D. melanogaster* CG1275 and Nemy with the solved
397 structure of Dcytb. To investigate, we overlayed AlphaFold models for *D. melanogaster* CG1275 and
398 Nemy with the ascorbate-bound structure of human Dcytb. The AlphaFold models have very high
399 confidence for the majority of residues in the six transmembrane helices and their connecting loops (S4
400 Table), and the alpha carbons of the two insect structures aligned well over most of the Dcytb sequence
401 (Fig 5) [47,48,52]. Although the AlphaFold cytb561 models do not include the two bound heme groups
402 and the two bound ascorbates, overlaying the AlphaFold structures with Dcytb showed how these
403 molecules may fit within the CG1275 and Nemy structures (Fig 5) [18].

404

405 **Fig 5. *D. melanogaster* CG1275 and Nemy models overlayed on Dcytb heme- and ascorbate-bound**
406 **structure.** AlphaFold structures of *D. melanogaster* CG1275 and Nemy were overlayed with Dcytb (PDB
407 5ZLG) using Chimera Matchmaker. Transmembrane helices are labeled (TM1-TM6), with Dcytb in tan,
408 CG1275 in pale pink, and Nemy in pale blue. The four conserved histidines are shown in stick form; in
409 the Dcytb structure, H1 is in red, H2 in yellow, H3 in green, and H4 in dark blue. Ascorbate (ASC) in black
410 stick, heme in tan stick (with iron shown as orange spheres), and zinc (Zn^{2+}) ion coordination in purple
411 are from the 5ZLG structure. Stick side chains, ascorbate, and heme have oxygen in red and nitrogen in

412 blue. The alpha carbon RMSD for CG1275 with 5ZLG is 0.679 Å for the 208 pruned atom pairs and 1.458
413 Å across all 222 pairs. The alpha carbon RMSD for Nemy with 5ZLG is 0.990 Å for the 188 pruned atom
414 pairs and 2.657 Å across all 222 pairs. A) Overlay of *D. melanogaster* CG1275 and Nemy with Dcytb. B)
415 Non-cytoplasmic ascorbate binding, zinc coordination, and heme coordination with H1 (red) and H3
416 (green) shown in stick form. C) Cytoplasmic ascorbate binding and heme coordination with H2 (yellow)
417 and H4 (dark blue) shown in stick form.
418

419 Conserved insect cytb561 residues likely to be involved in heme 420 positioning

421 We were interested in exploring whether residues involved in heme placement in Dcytb and
422 AtCytb561-B have similar roles in insect cytb561 proteins. Our Dcytb protein contact analysis identified
423 47 residues that make at least one predicted contact with a heme group (S3 Table), including 32
424 residues that are completely or well conserved in all CG1275 and Nemy sequences, as well as in
425 AtCytb561-B (Fig 4). Many of these 32 residues are predicted to make contacts with a heme group in *D.*
426 *melanogaster* CG1275 and Nemy protein models overlayed with Dcytb (S3 Table). In particular, the H1-
427 H4 side chains are positioned very similarly to the heme-coordinating H1-H4 side chains in the Dcytb
428 structure (Figs 5B and C). Amino acid residues that were predicted to interact with a heme group or
429 participate in heme positioning, as well as their degree of conservation in the four insect cytb561
430 groups, are described below.

431 Cytoplasmic heme

432 The cytoplasmic heme group in Dcytb directly interacts with three amino acid residues: H2, H4,
433 and Arg70 (H1+20) [18]. Arg70 forms a salt bridge with the heme A-propionate through its epsilon
434 nitrogen [18]; a similar salt bridge is formed by a lysine at H1+20 in AtCytb561-B (Lys71) [13]. An

435 arginine at H1+20 is completely conserved in all CG1275, Nemy, and CG8399 sequences (Fig 2) and likely
436 interacts with the cytoplasmic heme A-propionate (Fig 6A). The residue at H1+20 tends to be a serine or
437 threonine in Group 4A proteins (Fig 2), and 22 out of 25 Group 4B sequences also have a polar residue at
438 H1+20 (S1 Fig). Serine is a much smaller residue than arginine, but in Dcytb the gamma carbon of Arg70
439 is only ~3.2 Å from the heme A propionate; therefore, a serine or threonine residue could potentially
440 form a hydrogen bond with a heme carboxylate group if properly positioned. Outside of the Dcytb core
441 domain, at the end of helix 6, Lys225 forms two potential hydrogen bonds with the cytoplasmic heme D
442 propionate (Fig 6A; S3 Table). This helix 6 lysine is conserved in CG1275 sequences and is a basic residue
443 in 5 out of 7 Nemy sequences, but it is replaced with a proline in *D. melanogaster* and *A. gambiae* Nemy.
444 (Comparisons of Lys225 with CG8399 and Group 4 sequences were not attempted because helix 6 is
445 outside of their homologous regions.)

446

447 **Fig 6. Heme binding by Dcytb and *D. melanogaster* CG1275 and Nemy.** AlphaFold structures of *D.*
448 *melanogaster* CG1275 and Nemy were overlayed with Dcytb (5ZLG) using Chimera Matchmaker. Dcytb
449 loops are in tan and helices are shown in red (helix 2), yellow (helix 3), green (helix 4), dark blue (helix 5),
450 and purple (helix 6). CG1275 is in pale pink, and Nemy is in pale blue. Ascorbate (ASC) in black stick,
451 heme in tan stick (with iron shown as orange spheres), and zinc (Zn^{2+}) ion coordination in purple are
452 from the 5ZLG structure. Select side chains are shown in stick form and are labeled by position relative
453 to conserved heme-coordinating histidines with Dcytb amino acid numbering in parentheses. Key
454 glycines are shown in black ribbon. Stick side chains, ascorbate, and heme have oxygen in red and
455 nitrogen in blue. Hydrogen bonds shown as cyan dashed lines, including to the A and D heme-
456 propionates (labeled as in [18]). Water molecules shown as red spheres. A) Cytoplasmic heme region. B)
457 Non-cytoplasmic heme region. Hydrogen bonds predicted by ChimeraX or observed in Dcytb structure
458 (Asn115 to D heme-propionate; Glu180 to A heme-propionate, [18]).

459 In the Dcytb and AtCytb561-B structures, we observed interactions between H4 and the
460 glutamine at H3+14 (Gln134). Gln134 forms a predicted hydrogen bond with the H4 backbone carbonyl
461 and contributes to a network of hydrogen bonds linked by a water molecule to the H4 imidazole ring
462 positioned on the opposite side of the heme iron coordination site (Fig 6A). An H3+14 glutamine residue
463 is conserved in all CG1275, Nemy, and CG8399 sequences, suggesting a role in the positioning of H4 in
464 these three protein groups. In Group 4 sequences, a serine (16 out of 25) or threonine (4 out of 25) is
465 often present at H3+14 (S1 Fig) that could both potentially interact with H4, perhaps through hydrogen
466 bonding mediated by water. Thus, within CG1275, Nemy, and CG8399 proteins, a glutamine at H3+14
467 may help position H4 to coordinate the heme, while a salt bridge may form between the cytoplasmic
468 heme A propionate and the arginine at H1+20. In Group 4 proteins, polar residues at H1+20 and H3+14
469 could conceivably have similar heme positioning roles.

470 We found that two of the highly-conserved small residue positions in insect cytb561s are
471 predicted to form cytoplasmic heme contacts in Dcytb: H3+18 (Gly138) and H4+4 (Gly163) (S3 Table). A
472 glycine at H4+4 is conserved across all insect groups, in addition to more diverse cytb561s [3,19], while a
473 glycine at H3+18 is conserved in all insect groups except CG8399, where an alanine substitution occurs
474 (Fig 2). It is likely that small residues at H3+18 and H4+4 are important for proper cytoplasmic heme
475 positioning within the protein's central cavity in diverse cytb561s (Fig 6A).

476 Non-cytoplasmic heme

477 We performed a similar analysis of non-cytoplasmic heme positioning. In Dcytb and AtCytb561-
478 B, two residues form hydrogen bonds with the non-cytoplasmic heme A-propionate group: a serine at
479 H3-2 (Ser118) and a glutamate at H4+21 (Glu180) (Fig 6B)[13,18]. These residues are well conserved in
480 CG1275 and Nemy sequences and form predicted contacts with the heme group in the *D. melanogaster*
481 CG1275 and Nemy models (Fig 4; S3 Table). As mentioned above, it is notable that the most highly-

482 conserved regions among CG1275 and Nemy proteins are the sequences surrounding H1 and H3, which
483 coordinate the non-cytoplasmic heme group (Fig 2). While CG8399 and Group 4 proteins share less
484 conservation in these regions, a serine or threonine at H3-2 is well conserved across Group 4 proteins,
485 and CG8399 sequences most often contain an asparagine (Fig 2; S1 Fig); these polar residues could
486 potentially hydrogen bond with the non-cytoplasmic heme A-propionate (H4+21 is slightly beyond the
487 conserved core cytb561 domain and was therefore omitted from our analysis of CG8399 and Group 4
488 proteins).

489 In Dcytb and AtCytb561-B, an asparagine at H3-5 forms a hydrogen bond with the non-
490 cytoplasmic heme D-propionate (Fig 6B) [13,18]. This position is highly conserved in CG1275 and Nemy
491 proteins (Fig 4); two Nemy sequences have a histidine, which also could form a potential interaction
492 with a heme propionate. All Group 4A and most Group 4B proteins also have a histidine at H3-5 (Fig 2;
493 S1 Fig). In contrast to the other groups, there is no clear conservation at H3-5 in CG8399 proteins, which
494 is part of the core domain non-cytoplasmic loop (Fig 2; S1 Fig).

495 Additional AtCytb561-B residues that form hydrogen bonds with the non-cytoplasmic heme D-
496 propionate are the side chain of H2+22 (His106) and the amide nitrogen of H3-4 (Phe114) [13].
497 Interestingly, in Dcytb the histidine at H2+22 is conserved but was reported to coordinate with a non-
498 cytoplasmic zinc ion (Fig 6B, also discussed in more detail below) [18]. All but two CG1275 and Nemy
499 sequences have a histidine at H2+22, whereas this position is most commonly a lysine in Group 4
500 proteins or a hydrophobic residue in CG8399 proteins (Fig. 2; S1 Fig). Dcytb has a methionine (Met116)
501 at H3-4 whose amide nitrogen is also positioned to hydrogen bond with the non-cytoplasmic heme D-
502 propionate, similarly to Phe114 in AtCytb561-B (Fig 6B). In CG1275 and Nemy sequences, a methionine,
503 phenylalanine, or leucine is found at H3-4 and a phenylalanine is also common in Group 4 proteins (18
504 out of 25) (Figs 2 and 4). In contrast, CG8399s have no notable conservation at this position (S1 Fig). The
505 hydrophobic side chain at H3-4 in Dcytb (Met116) appears to anchor the non-cytoplasmic loop between

506 helices 3 and 4 in the membrane, and its backbone carbonyl is predicted to form a hydrogen bond with
507 the sidechain of the tryptophan (Trp122) at H3+2, providing added stability to the loop region between
508 helices 3 and 4 (Fig 6B). A tryptophan at H3+2 is entirely conserved in CG1275 and Nemy proteins (Fig 4),
509 suggesting a similar positioning of the non-cytoplasmic heme and loop between helices 3 and 4. The
510 tryptophan at H3+2 is not conserved in CG8399 or in most Group 4 proteins (Fig 2).

511 Glycine residues at positions H3+4 (Gly124) and H4+18 (Gly177) are predicted to make non-
512 cytoplasmic heme contacts in the Dcytb, CG1275, and Nemy structures (Figs 4 and 6B; S3 Table) These
513 residues were previously identified as highly conserved across diverse cytb561s [3]. The H3+4 glycine is
514 conserved in all CG1275, Nemy, and CG8399 sequences, and is highly conserved in Group 4 proteins (21
515 out of 25) (Fig 2). A glycine at H4+18 is conserved in all CG1275 and Nemy sequences and most Group 4
516 proteins (19 of 25) also have a glycine at H4+18, whereas CG8399s have an alanine substitution (Fig 2;
517 S1 Fig). Therefore, it is likely that small residues at H3+4 and H4+18 (non-cytoplasmic side), as well as
518 H4+4 and H3+18 (cytoplasmic side, described above) are important for positioning the bulky heme
519 groups in the central cavity of the protein across a range of cytb561s.

520 Heme positioning summary

521 Overall, heme positioning appears to be similar in Dcytb, AtCytb561-B, CG1275, and Nemy
522 proteins, whereas heme positioning in CG8399 and Group 4 proteins is less similar. Non-cytoplasmic
523 heme positioning is most dissimilar in CG8399 proteins, and cytoplasmic heme positioning is most
524 dissimilar in Group 4 proteins.

525 Ascorbate binding

526 Dcytb and AtCytb561-B were found to bind to ascorbate (or the closely related
527 monodehydroascorbate) in the cytoplasm through the interactions of basic residues with some of the
528 ascorbate oxygens [13,18]. On the non-cytoplasmic side, phenylalanine was implicated in stacking with

529 the ascorbate ring [13,18]. We located these Dcytb and AtCytb561-B residues in relation to the
530 conserved heme-coordinating histidines and examined our insect sequences for conserved residues,
531 especially basic and aromatic residues, at these locations. We also performed ChimeraX contact analysis
532 with the Dcytb structure and overlayed CG1275 and Nemy structures to identify other amino acids
533 proximate to ascorbate to identify similarities and differences in how ascorbate might bind.

534 Potential cytoplasmic ascorbate binding residues

535 The Dcytb and AtCytb561-B crystal structures clearly identified three basic residues that bind
536 cytoplasmic ascorbate: H2-7 (Lys79), H2-3 (Lys83), and H4-7 (Arg152) [13,18]. Additionally, ChimeraX
537 analysis predicted that Dcytb has ascorbate contacts at H3+22 (Phe142) and H4-3 (Met156) (Fig 7A; S3
538 Table). These five residues are also predicted to have cytoplasmic ascorbate contacts in *D. melanogaster*
539 CG1275 (S3 Table), and most CG1275 proteins have these conserved residues except for some variation
540 at H4-3 (Fig 4). In Nemy sequences, a lysine at H2-3, an arginine at H4-7, and a phenylalanine at H3+22
541 are entirely conserved, whereas either a valine or threonine is present at H4-3 (Fig 4); *D. melanogaster*
542 Nemy is predicted to have ascorbate contacts at these four positions (S3 Table). More variation is seen
543 at Nemy H2-7, which is most often an arginine (4 out of 8) and in one sequence is a lysine (Fig 4).
544 However, H2-7 is not a basic residue in the *D. melanogaster*, *A. gambiae*, and *A. pisum* Nemy sequences
545 (Fig 4). The *D. melanogaster* Nemy model suggests a shift in the placement of the lysine at H2-3 (Fig 7A),
546 which may compensate for the lack of a basic residue at H2-7. Additionally, the Nemy model predicts
547 that a serine (Ser205) at H4-6 hydrogen bonds with ascorbate, suggesting a unique functional role for
548 this serine in *D. melanogaster* Nemy and in *A. gambiae* Nemy, which also has a serine at this position
549 (Fig 7A).

550

551 **Fig 7. Ascorbate binding in Dcytb, CG1275, and Nemy.** AlphaFold structures of *D. melanogaster* CG1275
552 and Nemy were overlayed with Dcytb (5ZLG) using Chimera Matchmaker. Dcytb loops are in tan and
553 helices are shown in yellow (helix 3), green (helix 4), and dark blue (helix 5). CG1275 is in pale pink, and
554 Nemy is in pale blue. Ascorbate (ASC) in black stick, heme in tan stick (with iron shown as orange
555 spheres), and metal ion coordination in purple are from the 5ZLG structure. Select side chains are
556 shown in stick form and are labeled by position relative to conserved heme-coordinating histidines with
557 Dcytb amino acid numbering in parentheses (unless otherwise noted as Nemy numbering). Stick side
558 chains, ascorbate, and heme have oxygen in red and nitrogen in blue. Hydrogen bonds shown as cyan
559 dashed lines. A) Cytoplasmic ascorbate region. Ascorbate oxygens involved in bonding are labeled 1-3
560 and 6. B) Non-cytoplasmic ascorbate region with predicted hydrogen bonds to Dcytb, CG1275, and
561 Nemy shown in cyan; an additional predicted hydrogen bond between Nemy Lys161 and Glu236 is also
562 shown. Zinc coordination bonds are shown as orange dashed lines.

563 Of the five residues proximate to cytoplasmic ascorbate in Dcytb and AtCytb561-B (basic
564 residues at H2-7, H2-3, and H4-7, aromatic at H3+22, bulky hydrophobic at H4-3), only an arginine at H4-
565 7 (Arg509) is completely conserved across CG8399 sequences (amino acid numbering in this section
566 corresponds with *D. melanogaster* CG8399) (Fig 2). Most CG8399 sequences also have a conserved
567 lysine at H2-7, although *D. melanogaster* CG8399 has a threonine (Thr443) (S1 Fig). To visualize
568 conservation on the cytoplasmic side of CG8399 proteins, we overlayed the CG8399 consensus
569 sequence alignment onto the *D. melanogaster* CG8399 AlphaFold model (Fig 8). Of the remaining
570 positions discussed in the interaction of cytoplasmic ascorbate with Dcytb, AtCytb561-B, CG1275, and
571 Nemy proteins (H2-3, H3+22, and H4-3), CG8399 sequences have conserved residues with differing
572 properties. At H2-3, CG8399s have a phenylalanine (Phe447) rather than the lysine found in Dcytb,
573 AtCytb561-B, CG1275, and Nemy proteins, and a conserved tryptophan (Trp446) precedes H2-3 at H2-4
574 (Fig 8). Instead of a bulky hydrophobic residue at H3+22, a proline (Pro502) is present in CG8399s; that

575 proline is preceded by a completely conserved CG8399 arginine (Arg501) at H3+21 (Fig 8). Instead of a
576 bulky hydrophobic residue at H4-3, CG8399 proteins have an asparagine (Asn513). We noted additional
577 highly-conserved positions in this region among CG8399 proteins, including a second proline at H3+24
578 (Pro504) and a basic residue at H1+23 (Lys432), which is either a lysine or arginine in all the CG8399
579 sequences. H1+23 is an arginine in most CG1275 and Nemy sequences although not in Dcytb (Trp73) or
580 AtCytb561-B (Gln74) (Fig 4). While side-chain placement cannot be determined with high confidence, it
581 is notable how highly conserved this region is across multi-domain CG8399 proteins, including conserved
582 residues suitable for binding ascorbate: basic residues at H1+23, H3+21, and H4-7, and aromatics at H2-
583 4 and H2-3 (Fig 8; S5 Fig).

584

585 **Fig 8. Cytoplasmic side of *D. melanogaster* CG8399 model.** AlphaFold structure of *D. melanogaster*
586 CG8399 overlayed with multi-domain CG8399 consensus sequence alignment. The model is gradient
587 colored with high conservation (1.68 score) in magenta to low conservation (-1.89 score) in cyan. Select
588 side chains are shown in stick form with oxygen in red and nitrogen in blue and are labeled by position
589 relative to conserved heme-coordinating histidines with *D. melanogaster* CG8399 amino acid numbering
590 in parentheses. Ascorbate binding on the cytoplasmic surface of CG8399 may be facilitated by a cluster
591 of aromatic and basic residues: Lys432, Trp446, Phe447, Arg501, and Arg509. Also shown are conserved
592 H2 (His450) and H4 (His516), which coordinate the cytoplasmic heme iron (not present in model). Two
593 well-conserved prolines at positions H3+22 and H3+24 stabilize the loop between helices 3 and 4. Other
594 cytoplasmic side positions with well conserved residues include H2-7, which typically has a lysine but is
595 Thr443 in *D. melanogaster*, and H4-3 (Asp513). Scale bar depicts 10 Å.

596 Some Group 4 proteins also have a basic residue at two of the three positions that hydrogen
597 bond with cytoplasmic ascorbate in Dcytb and AtCytb561-B: H2-7 (11 out of 25) and H4-7 (11 out of 25)

598 (S1 Fig). While only six Group 4 proteins have a basic residue at H2-3, nearly all have basic residues at
599 H2-4 (23 out of 25) and H4-3 (23 out of 25) (S1 Fig). This is notable since H4-3 is one of the two positions
600 we identified as having a potential hydrophobic contact with cytoplasmic ascorbate in Dcytb, CG1275,
601 and Nemy (Fig 7A; S3 Table). It is possible that the basic residues at these highly-conserved positions are
602 involved in Group 4 cytoplasmic ascorbate binding. At the other position (H3+22) with predicted
603 hydrophobic ascorbate contact, most Group 4 proteins (19 out of 25) also have a large, hydrophobic
604 residue that could potentially interact with ascorbate or other possible cytoplasmic reducing agents,
605 such as dihydrolipoic acid (S1 Fig)[53]. The adjacent position, H3+23, is also typically an aromatic residue
606 (17 out of 25) or occasionally a lysine (3 out of 25) and may be involved in cytoplasmic ascorbate binding
607 (S1 Fig).

608 Non-cytoplasmic interactions with ascorbate and zinc: implications for ferric 609 reductase activity

610 Dcytb and AtCytb561-B also bind ascorbate at their non-cytoplasmic surfaces [13,18]. In Dcytb,
611 this appears to be part of its ferric reduction mechanism [18]. To identify an iron-binding site in Dcytb,
612 the protein was crystallized (PDB: 5ZLG) in the presence of redox inactive zinc, which presumably binds
613 Dcytb in a manner similar to iron [18]. Details of Dcytb ascorbate-complexed metal binding are
614 described below, and this information was used to predict non-cytoplasmic ascorbate and metal binding
615 by insect cytb561 proteins.

616 In Dcytb the only residue found to directly interact with non-cytoplasmic ascorbate was Phe184
617 (H4+25) via a van der Waals interaction, although several residues were discussed as contributing to the
618 environment of the ascorbate-binding pocket [18]; in AtCytb561-B the same H4+25 (Phe182) interaction
619 was reported, and several residues were also suggested to form possible hydrogen bonds, including
620 H2+22 (His106) and H3-3 (Tyr115) [13]. Our analysis revealed that three Dcytb residues have predicted

621 contacts with non-cytoplasmic ascorbate: H1-3 (Phe47), H2+22 (His108), and H4+25 (Phe184) (Fig 7B; S3
622 Table). A phenylalanine at H1-3 is also predicted to make ascorbate contacts in *D. melanogaster* CG1275
623 and Nemy (S3 Table); a phenylalanine at this position is conserved across CG1275 and Nemy sequences,
624 and is centrally located – near the zinc ion, the non-cytoplasmic ascorbate, and the non-cytoplasmic
625 heme – in Dcytb, CG1275, and Nemy structures (Fig 7B). In Dcytb, Phe47 was the only residue that had
626 an observed change in electron density upon zinc ion and ascorbate binding, indicating a change in
627 position [18]. Most Group 4A proteins have a phenylalanine at H1-3; however, there is less conservation
628 at this position in CG8399 and Group 4B proteins (S1 Fig).

629 As mentioned above, the phenylalanine (Phe184) at H4+25 was determined to form a van der
630 Waals interaction with the non-cytoplasmic ascorbate in Dcytb and AtCytb561-B [13,18]. An aromatic
631 amino acid, most often phenylalanine, is typical at H4+25 across CG1275 and Nemy sequences with a
632 small number of exceptions; the most dramatically different is a glutamate (Glu236) in *D. melanogaster*
633 Nemy (Fig 4). In the absence of ascorbate, our analysis revealed that Glu236 and a lysine at H3-7
634 (Lys161) are predicted to form a hydrogen bond in the *D. melanogaster* Nemy model that would run
635 straight through the region occupied by the non-cytoplasmic ascorbate ring (Fig 7B). The *D.*
636 *melanogaster* Nemy model also has extensive contacts predicted between the non-cytoplasmic
637 ascorbate and Lys161 (23, including one hydrogen bond) and Glu236 (29) (S3 Table). These results
638 suggest that *D. melanogaster* Nemy could bind non-cytoplasmic ascorbate though a combination of
639 compensatory substitutions in lieu of an aromatic residue at H4+25. H4+25 is slightly outside the
640 conserved core domain and therefore was not included in our analyses of CG8399 and Group 4
641 sequences.

642 In Dcytb, the histidine (His108) at H2+22 was determined to be part of an ascorbate-metal-
643 binding complex; His108 was found to directly coordinate with a redox inactive zinc ion, which directly
644 interacted with the non-cytoplasmic ascorbate [18]. The histidine at H2+22 is conserved across CG1275

645 and Nemy sequences and is predicted to be similarly positioned in *D. melanogaster* CG1275 and Nemy
646 models (Fig 4; Fig 7B). In contrast, a histidine at H2+22 is not conserved in other insect cytb561s (Fig 2).
647 The residue at H2+22 is nonpolar in CG8399 sequences, whereas many Group 4 proteins have a lysine at
648 this position, which could contribute to ascorbate binding but would not be expected to coordinate a
649 metal cation (Fig 2; S1 Fig).

650 His108 was also found to be critical for Dcytb ferric reductase activity, as were residues at H2+21
651 (Asn107), H3-3 (Tyr117), and H4+25 (Phe184, discussed above) [18]. Like Dcytb, CG1275 and Nemy
652 proteins have a tyrosine at H3-3 in addition to the histidine at H2+22 and phenylalanine at H4+25, and
653 they tend to have a conservatively-substituted serine at the H2+21 position (Fig 2). These conserved
654 residues support the hypothesis that CG1275 and Nemy proteins have ferric reductase activity. In
655 contrast, conservation of these residues in the other insect cytb561 proteins is limited to a glutamate
656 that is entirely conserved at H2+21 in multi-domain CG8399 proteins (discussed further below) (Fig 2).

657 Some Group 4A proteins may have a different type of non-cytoplasmic ascorbate-binding site. A lysine-
658 containing "KXXXXKXH" motif was observed in the non-cytoplasmic loop of the core domain in six of the
659 seven Group 4A proteins (excluding one of the two *C. felis* sequences) (Fig 2; S1 Fig). Group 4A includes
660 the proteins identified as more like human TScytb than any *D. melanogaster* cytb561 (Table 1; S2 Table).
661 We aligned these six Group 4A proteins with human TScytb (which can reduce ferric ions in nanodiscs
662 [14,54]) and TScytb-like proteins of diverse animal species and found that all sequences have this
663 "KXXXXKXH" motif (Fig 9A), which contributes to the high degree of conservation across TScytb proteins
664 on the non-cytoplasmic side (Fig 9B). The location of the "KXXXXKXH" motif in the non-cytoplasmic loop
665 between helices 3 and 4 (where substrate binding occurs in Dyctb), as well as its strict conservation in
666 TScytb-like proteins of diverse species, suggests an important functional role, possibly related to
667 ascorbate binding (Fig 9C). This motif was not identified in other human or insect cytb561s. A completely
668 conserved tryptophan at H4+23 could also potentially contribute to a non-cytoplasmic binding site, and

669 the histidine at H3-5 (within the "KXXXXKXH" motif) could potentially coordinate a ferric ion complexed
670 with ascorbate, similar to the substrate binding site of Dytcb, providing a site for ferric reduction in
671 TScytb-like proteins (Fig 9C).

672

673 **Fig 9. A conserved "KXXXXKXH" motif in Group 4A and related proteins.** A) Truncated alignment of
674 animal TScytb and Group 4A sequences shows a conserved "KXXXXKXH" motif upstream of the
675 conserved non-cytoplasmic heme-coordinating H3. Coloring by amino acid residue as in Fig 4.
676 Abbreviations: Cf, *C. felis*; Aten, *A. tenebrosa*; Hrob, *H. robusta*; Drer, *D. rerio*; Xlae, *X. laevis*; Pmon, *P.*
677 *montanus*; Mmus, *M. musculus*; Hsap, *H. sapiens*; Apla, *A. planci*; Osin, *O. sinensis*; Am, *A. mellifera*; Tc,
678 *T. castaneum*; Zn, *Z. nevadensis*; Px, *P. xuthus*; Ag, *A. gambiae*. B) Full alignment of animal Tscytb and
679 Group 4A sequences overlayed on AlphaFold model of *A. gambiae* XP_3200673.4 (AF-Q7PZD0-F1-
680 model_v4). H1-H4 shown in stick form with nitrogen in blue. C) Non-cytoplasmic side of the protein
681 contains completely conserved lysines (H2+22, H3-7), a histidine (H3-5), a tryptophan (H4+23), and a
682 phenylalanine (H1-3), in addition to H1 and H3; these residues are shown in stick with nitrogen atoms in
683 blue. B-C) Amino acid conservation is shown by a gradient of color as in Fig 3. Scale bars show 10 Å.

684 Other amino acids of interest

685 Aside from H1-H4, the most highly-conserved residues in the core domain of insect cytb561s are
686 small residues and hydrophobic residues (S1 Fig). As discussed above, four well-conserved, small
687 residues (H3+4, H3+18, H4+4, and H4+18) appear to be important for heme placement. The remaining
688 well-conserved residues may play a role in helical packing [55]. The small amino acids at H1+7 and
689 H2+14 are positioned where TM2 and TM3 face each other, while the small residue at H1+13 is on the
690 side of TM2 that faces TM5.

691 Of the investigated insect cytb561s, Nemy proteins possess unique cysteine-rich sequences in
692 their cytoplasmic loops, with double cysteines common in both cytoplasmic loops of the core domain
693 (S1 Fig). This observation is notable, as cysteines are among the least common amino acids in proteins,
694 yet they are often at positions of functional importance [56]. Cytoplasmic cysteine residues can
695 participate in redox-active disulfide formation and metal ligation, and they can also serve as sites of
696 post-translational modifications [56]. While the pH and reducing environment of the cytoplasm is not
697 conducive to disulfide bond formation, solvent exposed cysteines can be regulated by redox state [57].
698 The cytoplasm-exposed Nemy cysteines could form redox active disulfides, potentially functioning as an
699 on/off switch for electron transfer activity. Another possibility involves cysteine S-palmitoylation, given
700 that *D. melanogaster* Nemy was identified as a palmitoylated protein [58]. Cysteine S-palmitoylation is a
701 reversible post-translational modification used to regulate protein function [59]. Given its dynamic
702 nature, cysteine S-palmitoylation can direct varied cellular processes, including subcellular trafficking,
703 protein localization, and enzyme activity [59]. We did not detect multiple cysteine-rich cytoplasmic loop
704 sequences in the cytb561 domain of other insect cytb561s, nor in human cytb561 proteins, suggesting a
705 function unique to Nemy proteins. Future research could indicate whether these cysteines are
706 palmitoylated and whether they influence Nemy localization, cytoplasmic substrate binding, or protein
707 function.

708 Analysis of the CG8399 DOMON domain

709 Lack of conservation in the extracellular sections of the CG8399 cytb561 domain
710 suggests DOMON domain involvement in ferric reductase activity

711 As discussed above, many of the Dyctb non-cytoplasmic surface residues that interact with
712 ascorbate or function in ferric reduction are missing in CG8399 proteins (Fig 2) [13,18]. Nevertheless, *D.*
713 *melanogaster* CG8399 and its ortholog in mice, SDR2, possess in vitro ferric reductase activity [6,31].

714 These observations indicate that ferric reductase activity in CG8399 and SDR2 occurs through a
715 mechanism different from that of Dcytb and suggests the possibility that iron-binding in CG8399 and
716 SDR2 occurs in the extracellular DOMON domain. It has been previously proposed that CYBDOMs may
717 act as a "redox bridge" between cytosolic ascorbate and the extracellular environment [1]. Using the
718 redox bridge concept as a model, we suggest that the CG8399 cytb561 domain binds the reducing agent
719 (ascorbate) on the cytoplasmic surface, then transfers the electron to a DOMON domain heme, which
720 then transfers the electron to an electron-accepting substrate (possibly a metal ion).

721 CG8399 DOMONs may bind a b-type heme

722 To further investigate the DOMON domain in CG8399 proteins, we used the HHpred PDB server
723 to identify proteins with known structure that have the highest probability percentage of homology to
724 CG8399 DOMON sequences. We found that all insect CG8399 DOMONs have > 99% probability of
725 homology to the DOMON domain of human DBH and the cytochrome domain of *P. chrysosporium* CDH.
726 (Note that despite the use of "cytochrome" in the domain name of *P. chrysosporium* CDH, this domain is
727 not homologous with the cytochrome b561 domain of CG8399 proteins; rather, the *P. chrysosporium*
728 CDH cytochrome domain is homologous with the DOMON domain in CG8399 proteins.) Intriguingly, DBH
729 and CDH have only 10% sequence identity and bind different ligands, yet their structural alignment
730 suggests that both domains have very similar folds [60].

731 The crystal structure of the CDH cytochrome domain (DOMON homolog) revealed a b-type
732 heme group coordinated by an unusual methionine and histidine ligation of the heme iron [61]. HHpred
733 alignments of CDH with each individual insect CG8399 DOMON showed conservation of these
734 methionine and histidine residues in all sequences (Met286 and His379 in *D. melanogaster* CG8399 [S2
735 Fig]). Notably, the methionine and histidine are not conserved in the DOMONs of proteins that bind
736 sugar rather than heme, nor in DBH, which binds neither sugar nor heme [20,60]. We then generated an

737 alignment of CG8399 DOMON domains with the DOMON domains of diverse animal CYBDOM proteins,
738 including insect CG8399s, and found universal conservation of the methionine and histidine across all
739 species investigated, including both DOMON domains in the *A. pisum* CG8399 (S3 Fig). These results
740 agree with previous observations that CYBDOM proteins likely have a b-type heme group in their
741 DOMON domain [1]. Given the DOMON domain's suspected role in interdomain electron transfer and
742 the likelihood that CG8399s and their CYBDOM orthologs coordinate a heme group within their DOMON
743 domains, it seems likely that their putative ferric reductase activity occurs outside of the cytb561
744 domain.

745 Spatial relationship between the DOMON domain and cytb561 domain in 746 CG8399

747 To study the possibility of electron transfer to the DOMON domain, we analyzed the multi-
748 domain CG8399 consensus sequence overlay of the AlphaFold model for D. *melanogaster* CG8399,
749 which includes the extracellular reeler and DOMON domains in addition to the transmembrane cytb561
750 domain (Fig 10). In the following discussion, residue numbers relate to the *D. melanogaster* CG8399
751 protein sequence. The model predicts a DOMON domain largely composed of two five-strand beta
752 sheets, similar to the DOMON domain in DBH [60]. The predicted aligned error plots for the DOMON
753 and cytb561 domains were relatively low, suggesting their models could be informative for positional
754 insights [47,48]. (In contrast, the predicted position of the reeler domain was reported with high error.)
755 Notably, the *D. melanogaster* DOMON domain is oriented adjacent to the cytb561 domain, with
756 predicted hydrogen bonds between the side chains of Lys285 and Arg383 in the DOMON domain with
757 the backbone carbonyl oxygens of Leu472 and Pro538 in the cytb561 domain, respectively (Fig 10B). Salt
758 bridges are predicted to form in the non-cytoplasmic loops of the cytb561 domain between Arg403 and
759 Glu541, as well as Glu471 and Lys539 (Fig 10B).

760 **Fig 10. Model of *D. melanogaster* CG8399 demonstrating interactions between domains, and a**
761 **putative iron binding site.** A-C) Models were colored by ChimeraX based on amino acid conservation in
762 the nine multidomain CG8399 sequences. Color gradient as in Fig 3. Residues in stick form show oxygen
763 in red, nitrogen in blue and sulfur in yellow. A-D) Scale bars depict 10 Å. A) CG8399 has three distinct
764 domains: reeler (top left), DOMON (middle right), and cytb561 (bottom). Residues expected to be heme-
765 coordinating are shown in stick (some not visible). B) Selected sidechains at the interface of the DOMON
766 (top) and cytb561 (bottom) domains. Predicted heme-coordinating residues (Met286 and His379) in the
767 DOMON domain are positioned at the bottom of the domain where it is predicted to interact with the
768 cytb561 domain. Interdomain hydrogen bonds were predicted between the sidechain of Lys285
769 (DOMON) and the backbone of Leu472 (cytb561), as well as the sidechain of Arg383 (DOMON) and the
770 backbone of Pro538 (cytb561). Potential salt bridge interactions may form in the cytb561 domain loop
771 regions between Glu471 and Lys539, as well as Arg403 and Glu541. C) A putative metal binding site in a
772 highly conserved loop with two aspartates and a serine residue. D) Space filling model of the same part
773 of the structure depicted in C, with Coulombic coloring calculated by ChimeraX, red for negative charge,
774 white for neutral, and dark blue for positive charge.

775 We then located the predicted positions of the putative heme-coordinating residues (Met286
776 and His379) in the DOMON domain, with particular interest in their relative positions to the non-
777 cytoplasmic heme-coordinating histidines (His409 and His480) in the cytb561 domain. The location of
778 the putative DOMON heme-binding pocket is predicted to be directly adjacent to the non-cytoplasmic
779 heme-binding region of the cytb561 domain (Fig 10B) [47,48]. These predictions further support the
780 hypothesis that the CG8399 cytb561 domain functions in interdomain electron transport rather than
781 being the site for oxidoreductase activity.

782 ## The CG8399 DOMON domain may contain a metal-binding pocket

783 If CG8399 ferric reductase activity occurs within the DOMON domain, we would expect to locate
784 an iron-binding pocket within this domain. Our DOMON alignment revealed multiple well-conserved
785 aspartates, including strict conservation of Asp283 and Asp289 (S3 Fig). Asp283 and Asp289 are in a
786 highly-conserved region (sequence "SXDXXMGXD") that includes the putative heme-coordinating
787 methionine (Met286) (S3 Fig). This region also contains additional aspartates, especially in insect
788 CG8399s, of which two are particularly well conserved (S3 Fig). We located this region in the *D.*
789 *melanogaster* CG8399 model and found that a negatively-charged pocket is predicted to be directly
790 behind the putative heme-binding pocket, similar to the position of the putative metal ion-binding
791 pocket in DBH relative to its ligand-binding pocket (Figs 10C and D) [60]. These results are consistent
792 with the DOMON domain being the site of ferric reductase activity, wherein strictly-conserved Ser281,
793 Asp283, and Asp289 could form an iron-binding pocket in insect CG8399s and CYBDOMs from other
794 species (S3 Fig). Metal-binding sites formed by aspartate residues have been suggested for other
795 proteins. For example, the DBH DOMON domain has a putative metal ion-binding site composed of
796 aspartate sidechains and the backbone carbonyl oxygens of two other amino acids [60], and three
797 aspartate residues form a binding site for a variety of cations, including Fe^{2+} and Fe^{3+}, in LarE [62]. This
798 model also provides a possible explanation for the ferric reductase activity of *D. melanogaster* CG8399
799 and mouse SDR2, despite their lack of conserved ascorbate- and metal-binding residues on the non-
800 cytoplasmic surface of the cytb561 domain.

801 ## Three insect cytb561s with distinctive features outside of the core
802 ## domain

803 Of the 54 identified insect cytb561s, we noted that three sequences possess unique features
804 within the cytb561 domain but outside the domain core (TM2-TM5 in Nemy proteins, and TM1-TM4 in

805 CG8399s). To determine whether other insect cytb561 proteins have any of these unique features, we
806 performed additional BLAST searches against the entire class Insecta.

807 Compared to other Nemys, the *A. pisum* sequence has approximately 150 additional N-terminal
808 residues preceding helix 1 and approximately 50 additional residues in the non-cytoplasmic loop
809 connecting helix 1 and helix 2, which immediately precedes the core cytb561 domain (S4 Fig). We found
810 proteins with similar extended regions in eight additional species of the Aphididae family, but not in
811 other insects, suggesting that these regions may serve a unique function within aphids (S4 Fig; S5 Table).
812 The extended N-terminal region preceding helix 1 is acidic and includes relatively high percentages of
813 aspartate, glycine, serine, and threonine (S4 Fig).

814 Of the investigated CG8399s, only the *A. pisum* sequence possesses two DOMON domains. We
815 identified 888 CYBDOM hits in the class Insecta, of which only 19 were found to have two DOMON
816 domains: nine from other species of aphids, eight from other insects from the order Hemiptera, and two
817 from insects in the order Coleoptera (S6 Table). It appears that multiple DOMONs are uncommon in
818 insect cytb561s; however, they have a relatively high occurrence in Hemipterans. Both *A. pisum* DOMON
819 domains contain the conserved methionine and histidine that are assumed to coordinate a heme group
820 (S3 Fig); however, it is unknown how these two DOMON domains are spatially arranged.

821 While all nine investigated insect species possess a multidomain CG8399, *A. gambiae* (mosquito)
822 is the only species with a second, single domain CG8399-like protein (XP_314065.2); this sequence lacks
823 both reeler and DOMON domains. Further investigation revealed similar single-domain (cytb561 only)
824 CG8399-like proteins in ten additional mosquito species, as well as two species of fungus gnats, which
825 are also in the order Diptera (S7 Table). The single-domain CG8399-like protein is more similar to the
826 cytb561 domain in multidomain CG8399 proteins than to the cytb561 domain in other groups of
827 cytb561s (Table 1; S2 Table), including its core domain spanning TM1-TM4, rather than TM2-TM5.

828 ## Cyb561 subcellular localization

829 Subcellular localization is important for protein function in an organism. We submitted each of
830 the 54 full-length insect cytb561s to the SignalP-5.0 webserver for detection of predicted signal
831 peptides, and to the DEEPLOC-1.0 webserver for prediction of subcellular protein localization [40]. Only
832 the multidomain CG8399 proteins had a predicted signal peptide, and they are the only sequences with
833 predicted plasma membrane localization (S8 Table). This result is consistent with the presence of *D.*
834 *melanogaster* CG8399 in a plasma membrane fraction of fly head proteins [63]. The *A. gambiae* single-
835 domain CG8399-like protein is predicted to be in the lysosome, as are all the CG1275 and Nemy proteins
836 (S8 Table). The subcellular localization predictions for Group 4 proteins were not particularly robust, and
837 the predicted location varied depending on the specific Group 4 protein (S8 Table).

838 An alignment of the Dcytb, Nemy, and CG1275 carboyxl-termini revealed that only the CG1275
839 proteins have a "EDXXLL" motif near TM6 that is consistent with a dileucine-based late-
840 endosomal/lysosomal targeting signal "(D/E)XXXL(L/I)" (Fig 11) [42]. *A. mellifera* and *T. castaneum*
841 CG1275 proteins lack one conserved leucine, and one of the four *A. pisum* sequences (XP_003246890.1)
842 lacks both leucines, yet these three sequences are still predicted to localize to the lysosome (Fig 11; S8
843 Table).

844

845 **Fig 11. Lysosomal targeting signal in CG1275.** An alignment of the C-terminal portion of CG1275
846 proteins. Helix predictions from the AlphaFold model of *D. melanogaster* CG1275 shown in gray (helix 6)
847 and an additional helical region is in black. The EDXXLL motif is boxed in blue in the consensus logo.

848 ## Summary of cytochrome b561s as potential ferric reductases in insects

849 A major goal of this study was to predict which of the insect cytb561 orthologous groups may
850 function as ferric reductases and, thus, have a possible role in iron uptake by insect cells. Our analyses
851 suggest that none of the four groups of insect cytb561 proteins can be excluded as ferric reductase
852 candidates.

853 Based on sequence similarity, predicted structural similarity, and conservation of key amino acid
854 residues, we conclude that CG1275 and Nemy proteins are strikingly similar to the well-established
855 ferric reductase Dcytb, as well as the closely-related human proteins Lcytb and CGcytb. These similarities
856 extend to the conservation of a metal-binding histidine and other amino acid residues that are required
857 for ferric reductase activity by Dcytb. To date, no functions of CG1275 proteins have been discovered;
858 however, if CG1275 proteins have ferric reductase activity, the constitutive, nearly ubiquitous
859 expression of *D. melanogaster* CG1275 [64,65] would be consistent with a role in iron uptake by many
860 cell types. We found that CG1275 proteins have a C-terminal dileucine motif that directs proteins to a
861 late-endosomal/lysosomal location, suggesting that CG1275 proteins could function similarly to the non-
862 homologous STEAP3 protein, which reduces Fe^{3+} to Fe^{2+} prior to uptake across endosomal and lysosomal
863 membranes [27,28,66]. Among mammalian cytb561s, a C-terminal dileucine motif is present only in
864 Lcytb proteins [7]. Interestingly, Lcytb appears to function as a lysosomal ferric reductase in Burkitt
865 lymphoma cells, filling the role that STEAP3 plays in some other cell types [67]. Although Nemy proteins
866 have many similarities with Dcytb and Lyctb, their one known function is unrelated to iron homeostasis
867 [4,30]; however, Nemy proteins could be multifunctional enzymes that also play a role in iron uptake.

868 *D. melanogaster* CG8399 and the mammalian ortholog SDR2 have in vitro ferric reductase activity
869 [6,15,31], but the in vivo functions of CG8399 and SDR2 proteins are not known. Based on a previous
870 biochemical study [63] and our subcellular localization predictions, CG8399 proteins localize to the

871 plasma membrane, with their reeler and DOMON domains in the extracellular compartment. Their
872 presence in the plasma membrane may enable reduction of extracellular ferric ions prior to iron uptake.
873 Our analyses of the predicted structures of the CG8399 DOMON and cytb561 domains suggest that the
874 two domains may function together in ferric reduction, with the interdomain transfer of an electron
875 from cytoplasmic ascorbate to a heme in the extracellular DOMON domain and then to a ferric ion that
876 is perhaps coordinated nearby by multiple aspartate residues. Because *D. melanogaster* CG8399
877 expression is limited to a subset of tissues [64,65], its hypothetical role in iron uptake would be limited
878 to specific cell types.

879 We have the least amount of evidence to support the hypothesis that the diverse Group 4
880 proteins function as ferric reductases, and a lack of one-to-one orthologs indicates that these proteins
881 may have non-conserved physiological functions. The expression profiles of all but one of the *D.*
882 *melanogaster* Group 4 genes is moderately to highly restricted [64,65], excluding a role in iron uptake by
883 most cell types. The Group 4 proteins have low similarity to the well-studied mammalian cytb561s,
884 lacking even conserved residues that bind ascorbate in the cytoplasm. However, multiple basic residues
885 and a large hydrophobic residue at the cytoplasmic surface of most Group 4 proteins may enable the
886 Group 4 proteins to use ascorbate as an electron donor. On the non-cytoplasmic side, a subset of Group
887 4 proteins contain a "KXXXXKXH" motif that is also present in TScytb, a protein with an unknown
888 biochemical function in vivo, but ferric reductase activity in vitro [14]. We propose that the "KXXXXKXH"
889 motif may bind ascorbate and a ferric ion in a hypothetical ferric reduction site.

890

891

Acknowledgements

We wish to acknowledge the undergraduate research students who participated in aspects of this project: Clare Burnett, Chrissy Chan, Caleb Garris, Keyata Lewis, Gregory Kane, Karen Nunez Sifuentes, Connor Sullivan, and Nataliya Yusopov. Thanks also to the MSU Denver CHE 4320 students from Spring 2018, Fall 2018, Spring 2019, Fall 2019 and Spring 2020 who participated in iterations of the Iron Uptake in Insects CURE. We thank Dr. Andrew McMillan for thoughtful comments on the manuscript. Molecular graphics and analyses were performed with UCSF ChimeraX, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco. FlyBase was used for gene identification and as a source of expression data.

References

1. Asard H, Barbaro R, Trost P, Bérczi A. Cytochromes *b* 561: Ascorbate-mediated trans-membrane electron transport. Antioxid Redox Signal. 2013;19(9):1026–35. doi: 10.1089/ars.2012.5065
2. Berczi A, Zimanyi L. The trans-membrane cytochrome b561 proteins: Structural information and biological function. Curr Protein Pept Sci. 2014;15(8):745–60. doi: 10.2174/1389203715666140828100351
3. Verelst W, Asard H. A phylogenetic study of cytochrome b561 proteins. Genome Biol. 2003;4(6) R38. doi: 10.1186/gb-2003-4-6-r38
4. Iliadi KG, Avivi A, Iliadi NN, Knight D, Korol AB, Nevo E, et al. nemy encodes a cytochrome b561 that is required for Drosophila learning and memory. Proc Natl Acad Sci. 2008;105(50):19986–91. doi: 10.1073/pnas.0810698105
5. Latunde-Dada GO, Van der Westhuizen J, Vulpe CD, Anderson GJ, Simpson RJ, McKie AT. Molecular and functional roles of duodenal cytochrome B (Dcytb) in iron metabolism. Blood Cells Mol Dis. 2002;29(3):356–60. doi: 10.1006/bcmd.2002.0574
6. Vargas JD, Herpers B, McKie AT, Gledhill S, McDonnell J, van den Heuvel M et al. Stromal cell-derived receptor 2 and cytochrome b561 are functional ferric reductases. Biochim Biophys Acta BBA - Proteins Proteomics. 2003;1651(1–2):116–23. doi: 10.1016/S1570-9639(03)00242-5
7. Zhang D liang, Su D, Bérczi A, Vargas A, Asard H. An ascorbate-reducible cytochrome b561 is localized in macrophage lysosomes. Biochim Biophys Acta BBA - Gen Subj. 2006;1760(12):1903–13. doi: 10.1016/j.bbagen.2006.07.019

8. Lane D, Bae DH, Merlot A, Sahni S, Richardson D. Duodenal cytochrome b (DCYTB) in iron metabolism: An update on function and regulation. Nutrients. 2015;7(4):2274–96. doi: 10.3390/nu7042274
9. Lemler DJ, Lynch ML, Tesfay L, Deng Z, Paul BT, Wang X, et al. DCYTB is a predictor of outcome in breast cancer that functions via iron-independent mechanisms. Breast Cancer Res. 2017;19(1):25. doi: 10.1186/s13058-017-0814-9
10. Su D, May JM, Koury MJ, Asard H. Human erythrocyte membranes contain a cytochrome b561 that may be involved in extracellular ascorbate recycling. J Biol Chem. 2006;281(52):39852–9. doi: 10.1074/jbc.M606543200
11. Cenacchi L, Busch M, Schleidt PG, Müller FG, Stumpp TVM, Mäntele W, et al. Heterologous production and characterisation of two distinct dihaem-containing membrane integral cytochrome b(561) enzymes from *Arabidopsis thaliana* in *Pichia pastoris* and *Escherichia coli* cells. Biochim Biophys Acta. 2012;1818(3):679–88. doi: 10.1016/j.bbamem.2011.10.030
12. Klein M, Deniz E, Heit S, Wille G, Mäntele W, Lancaster CRD. Proton-coupled electron transport in two distinct CYBASC paralogs of *Arabidopsis thaliana*: A comparative characterization of highly conserved tyrosine and lysine residues. Biochemistry. 2020;59(25):2328–39. doi: 10.1021/acs.biochem.0c00155
13. Lu P, Ma D, Yan C, Gong X, Du M, Shi Y. Structure and mechanism of a eukaryotic transmembrane ascorbate-dependent oxidoreductase. Proc Natl Acad Sci. 2014;111(5):1813–8. doi: 10.1073/pnas.1323931111
14. El Behery M, Fujimura M, Kimura T, Tsubaki M. Direct measurements of ferric reductase activity of human 101F6 and its enhancement upon reconstitution into phospholipid bilayer nanodisc. Biochem Biophys Rep. 2020;21:100730. doi: 10.1016/j.bbrep.2020.100730
15. Su D, Asard H. Three mammalian cytochromes b561 are ascorbate-dependent ferrireductases. FEBS J. 2006;273(16):3722–34. doi: 10.1111/j.1742-4658.2006.05381.x
16. Kelley PM, Njus D. Cytochrome b561 spectral changes associated with electron transfer in chromaffin-vesicle ghosts. J Biol Chem. 1986;261(14):6429–32. doi: 10.1016/S0021-9258(19)84579-9
17. Njus D, Knoth J, Cook C, Kelly PM. Electron transfer across the chromaffin granule membrane. J Biol Chem. 1983;258(1):27–30. doi: 10.1016/S0021-9258(18)33212-5
18. Ganasen M, Togashi H, Takeda H, Asakura H, Tosha T, Yamashita K, et al. Structural basis for promotion of duodenal iron absorption by enteric ferric reductase with ascorbate. Commun Biol. 2018;1(1):120. doi: 10.1038/s42003-018-0121-8
19. Tsubaki M, Takeuchi F, Nakanishi N. Cytochrome b561 protein family: Expanding roles and versatile transmembrane electron transfer abilities as predicted by a new classification system and protein sequence motif analyses. Biochim Biophys Acta BBA - Proteins Proteomics. 2005;1753(2):174–90. doi: 10.1016/j.bbapap.2005.08.015
20. Iyer LM, Anantharaman V, Aravind L. The DOMON domains are involved in heme and sugar recognition. Bioinformatics. 2007;23(20):2660–4. doi:10.1093/bioinformatics/btm411
21. Gorman MJ. Iron Homeostasis in Insects. Annu Rev Entomol. 2023;68(1):51-67. doi: 10.1146/annurev-ento-040622-092836
22. Tang X, Zhou B. Iron homeostasis in insects: Insights from *Drosophila* studies: Iron Homeostasis in Insects. IUBMB Life. 2013;65(10):863–72. doi: 10.1002/iub.1211
23. Wu S, Yin S, Zhou B. Molecular physiology of iron trafficking in *Drosophila melanogaste*r. Curr Opin Insect Sci. 2022;50:100888. doi: 10.1016/j.cois.2022.100888
24. Andrews NC, Fleming MD, Gunshin H. Iron transport across biologic membranes. Nutr Rev. 1999;57(4):114–23. doi: 10.1111/j.1753-4887.1999.tb06934.x

25. Kosman DJ. A holistic view of mammalian (vertebrate) cellular iron uptake. Metallomics. 2020;12(9):1323–34. doi: 10.1039/D0MT00065E
26. Montalbetti N, Simonin A, Kovacs G, Hediger MA. Mammalian iron transporters: Families SLC11 and SLC40. Mol Aspects Med. 2013;34(2):270–87. doi: 10.1016/j.mam.2013.01.002
27. Gomes IM, Maia CJ, Santos CR. STEAP proteins: From structure to applications in cancer therapy. Mol Cancer Res. 2012;10(5):573–87. doi: 10.1158/1541-7786.MCR-11-0281
28. Knutson MD. Steap Proteins: Implications for Iron and Copper Metabolism. Nutr Rev. 2007;65(7):335–40. doi: 10.1111/j.1753-4887.2007.tb00311.x
29. Schlottmann F, Vera-Aviles M, Latunde-Dada GO. Duodenal cytochrome b (Cybrd1) ferric reductase functional studies in cells. Metallomics. 2017;9(10):1389–93. doi: 10.1039/C7MT00254H
30. Knight D, Iliadi KG, Iliadi N, Wilk R, Hu J, Krause HM, et al. Distinct regulation of transmitter release at the Drosophila NMJ by different isoforms of nemy. PLOS ONE. 2015;10(8):e0132548. doi: 10.1371/journal.pone.0132548
31. Picco C, Scholz-Starke J, Naso A, Preger V, Sparla F, Trost P, et al. How are cytochrome b561 electron currents controlled by membrane voltage and substrate availability? Antioxid Redox Signal. 2014;21(3):384–91. doi: 10.1089/ars.2013.5809
32. Mandilaras K, Pathmanathan T, Missirlis F. Iron absorption in *Drosophila melanogaster*. Nutrients. 2013;5(5):1622–47. doi: 10.3390/nu5051622
33. Larkin A, Marygold SJ, Antonazzo G, Attrill H, dos Santos G, Garapati PV, et al. FlyBase: updates to the *Drosophila melanogaster* knowledge base. Nucleic Acids Res. 2021;49(D1):D899–907. doi: 10.1093/nar/gkaa1026
34. NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2018;46(D1):D8–13. doi: 10.1093/nar/gkx1095
35. Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, et al. CDD/SPARCLE: the conserved domain database in 2020. Nucleic Acids Res. 2020;48(D1):D265–8. doi: 10.1093/nar/gkz991
36. Gabler F, Nam SZ, Till S, Mirdita M, Steinegger M, Söding J, et al. Protein sequence analysis using the MPI bioinformatics toolkit. Curr Protoc Bioinforma. 2020;72(1):e108. doi: 10.1002/cpbi.108
37. Sievers F, Higgins DG. Clustal omega. Curr Protoc Bioinforma. 2014;48(1):3.13.1-3.13.16. doi: 10.1002/0471250953.bi0313s48
38. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics. 2009;25(9):1189–91. doi: 10.1093/bioinformatics/btp033
39. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011;7:539. doi: 10.1038/msb.2011.75
40. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat Biotechnol. 2019;37(4):420–3. doi: 10.1038/s41587-019-0036-z
41. Almagro Armenteros JJ, Sønderby CK, Sønderby SK, Nielsen H, Winther O. DeepLoc: prediction of protein subcellular localization using deep learning. Hancock J, editor. Bioinformatics. 2017;33(21):3387–95. doi: 10.1093/bioinformatics/btx431
42. Bonifacino JS, Traub LM. Signals for sorting of transmembrane proteins to endosomes and lysosomes. Annu Rev Biochem. 2003;72(1):395–447. doi: 10.1146/annurev.biochem.72.121801.161800

43. Srivastava M, Begovic E, Chapman J, Putnam NH, Hellsten U, Kawashima T, et al. The Trichoplax genome and the nature of placozoans. Nature. 2008;454(7207):955–60. doi: 10.1038/nature07191
44. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol. 2018;35(6):1547–9. doi: 10.1093/molbev/msy096
45. Hall BG. Phylogenetic Trees Made Easy: A How-To Manual. Fifth Edition. Oxford, New York: Oxford University Press; 2017. 368 p.
46. Goddard TD, Huang CC, Meng EC, Pettersen EF, Couch GS, Morris JH, et al. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. Protein Sci Publ Protein Soc. 2018;27(1):14–25. doi: 10.1002/pro.3235
47. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583–9. doi: 10.1038/s41586-021-03819-2
48. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, et al. AlphaFold protein structure database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res. 2022;50(D1):D439–44. doi: 10.1093/nar/gkab1061
49. Teufel F, Almagro Armenteros JJ, Johansen AR, Gíslason MH, Pihl SI, Tsirigos KD et al. SignalP 6.0 predicts all five types of signal peptides using protein language models. Nat Biotechnol. 2022;40(7):1023–5. doi: 10.1038/s41587-021-01156-3
50. Pei J, Grishin NV. AL2CO: calculation of positional conservation in a protein sequence alignment. Bioinforma Oxf Engl. 2001;17(8):700–12. doi:10.1093/bioinformatics/17.8.700
51. Maddison DR, Schulz KS, Maddison WP. The tree of life web project. Zootaxa. 2007;1668(1):19–40.
52. Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, et al. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Sci. 2021;30(1):70–82. doi: 10.1002/pro.3943
53. Bérczi A, Zimányi L, Asard H. Dihydrolipoic acid reduces cytochrome b561 proteins. Eur Biophys J EBJ Heidelb. 2013 Mar;42(2–3):159–68. doi: 10.1007/s00249-012-0812-x
54. Recuenco MC, Rahman MdM, Sakamoto Y, Takeuchi F, Hori H, Tsubaki M. Functional characterization of the recombinant human tumour suppressor 101F6 protein, a cytochrome b561 homologue. J Biochem (Tokyo). 2013;153(2):233–42. doi: 10.1093/jb/mvs139
55. Walters RFS, DeGrado WF. Helix-packing motifs in membrane proteins. Proc Natl Acad Sci. 2006;103(37):13658–63. doi: 10.1073/pnas.0605878103
56. Bak DW, Bechtel TJ, Falco JA, Weerapana E. Cysteine reactivity across the subcellular universe. Curr Opin Chem Biol. 2019;48:96–105. doi: 10.1016/j.cbpa.2018.11.002
57. Held JM. Redox systems biology: Harnessing the sentinels of the cysteine redoxome. Antioxid Redox Signal. 2020;32(10):659–76. doi: 10.1089/ars.2019.7725
58. Strassburger K, Kang E, Teleman AA. Drosophila ZDHHC8 palmitoylates scribble and Ras64B and controls growth and viability. PLOS ONE. 2019;14(2):e0198149. doi: 10.1371/journal.pone.0198149
59. Bannan BA, Van Etten J, Kohler JA, Tsoi Y, Hansen NM, Sigmon S, et al. The Drosophila protein palmitoylome: Characterizing palmitoyl-thioesterases and DHHC palmitoyl-transferases. Fly (Austin). 2008;2(4):198–214. doi: 10.4161/fly.6621
60. Vendelboe TV, Harris P, Zhao Y, Walter TS, Harlos K, El Omari K, et al. The crystal structure of human dopamine β-hydroxylase at 2.9 Å resolution. Sci Adv. 2016;2(4):e1500980. doi: 10.1126/sciadv.1500980

61. Hallberg BM, Bergfors T, Bäckbro K, Pettersson G, Henriksson G, Divne C. A new scaffold for binding haem in the cytochrome domain of the extracellular flavocytochrome cellobiose dehydrogenase. Structure. 2000;8(1):79–88. doi: 10.1016/S0969-2126(00)00082-4
62. Fellner M, Huizenga KG, Hausinger RP, Hu J. Crystallographic characterization of a tri-Asp metal-binding site at the three-fold symmetry axis of LarE. Sci Rep. 2020;10(1):5830. doi: 10.1038/s41598-020-62847-6
63. Khanna MR, Stanley BA, Thomas GH. Towards a membrane proteome in *Drosophila*: a method for the isolation of plasma membrane. BMC Genomics. 2010;11:302. doi: 10.1186/1471-2164-11-302
64. Chintapalli VR, Wang J, Dow JAT. Using FlyAtlas to identify better *Drosophila melanogaster* models of human disease. Nat Genet. 2007;39(6):715–20. doi: 10.1038/ng2049
65. Leader DP, Krause SA, Pandit A, Davies SA, Dow JAT. FlyAtlas 2: a new version of the *Drosophila melanogaster* expression atlas with RNA-Seq, miRNA-Seq and sex-specific data. Nucleic Acids Res. 2018;46(D1):D809–15. doi: 10.1093/nar/gkx976
66. Meng F, Fleming BA, Jia X, Rousek AA, Mulvey MA, Ward DM. Lysosomal iron recycling in mouse macrophages is dependent upon both LcytB and Steap3 reductases. Blood Adv. 2022;6(6):1692–707. doi: 10.1182/bloodadvances.2021005609
67. Wang Z, Guo R, Trudeau SJ, Wolinsky E, Ast T, Liang JH, et al. CYB561A3 is the key lysosomal iron reductase required for Burkitt B-cell growth and survival. Blood. 2021;138(22):2216–30. doi: 10.1182/blood.2021011079

Supporting Information

S1 Fig. Alignment of the homologous cytb561 core domain in all analyzed insect sequences. Cytb561 protein sequences from *A. gambiae (Ag)*, *A. mellifera (Am)*, *A. pisum (Ap)*, *C. felis (Cf)*, *D. melanogaster (Dm)*, *T. castaneum (Tc)*, *P. humanus corporis (Ph)*, *P. xuthus (Px)*, and *Z. nevadensis (Zn)* were aligned with an outgroup sequence from *T. adhaerens (Ta)*. AlphaFold models of the *D. melanogaster* cytb561s were used to predict transmembrane helices and loop regions: helix 2 is boxed in red, helix 3 in yellow, helix 4 in green, and helix 5 in blue (helices 1-4 for CG8399); short helices predicted between the transmembrane helices are shown in gray. The approximate location of transmembrane helices and loops are labeled at the top of the alignment. The four strictly-conserved histidines are numbered in order of occurrence; 11 highly-conserved residues (Jalview conservation score of 8 or 9) are indicated with a caret (^). The Jalview consensus row shows the consensus logos with the most common residue at that position listed below the logo; a + is used where equal top residues were calculated by Jalview.

1099 Coloring by amino acid residue: basic residues in blue (Arg) or cyan (Lys); acidic residues (Glu and Asp) in
1100 dark pink; amide residues (Asn and Gln) in light pink; hydroxylic residues (Ser and Thr) in pale purple;
1101 aromatic residues in light peach (Phe and Tyr) or dark peach (Trp); sulfur-containing residues in dark
1102 green (Cys) or light green (Met); His in orange; Pro in yellow; Ile, Leu, and Val in pale yellow; and Ala and
1103 Gly in white.

1104 **S2 Fig. Alignment between CG8399 DOMON domain and cellobiose dehydrogenase (CDH) cytochrome**
1105 **domain.** An HHPred search using the sequence of the *D. melanogaster* CG8399 DOMON domain
1106 matched with the *P. chrysosporium* cellobiose dehydrogenase (CDH) cytochrome domain (PDB 1D7B).
1107 The CDH methionine and histidine residues that bind to bind a b-type heme group in CDH are conserved
1108 in the CG8399 DOMON domain and are marked with an asterisk (*).

1109 **S3 Fig. Alignment of DOMON domains from insect CG8399 proteins and CYBDOMs from other animal**
1110 **species.** DOMON domains from insect species (as in S1 Fig., including both *A. piusm* DOMON domains)
1111 and non-insect animals *D. rerio,* (Drer, zebrafish); *X. laevis* (Xlae, frog); *P. montanus* (Pmon, sparrow); *M.*
1112 *musculus* (Mmus, mouse); *H. sapiens* (Hsap, human) were aligned. Conserved residues include putative
1113 heme-binding methionine and histidine residues (*), residues surrounding the conserved methionine
1114 (+), cysteines predicted to form disulfide bonds (blue brackets), and Lys233 and Glu264, which are
1115 positioned to form a salt bridge (black bracket). Alignment colored as in S1 Figure. The Jalview
1116 consensus rows show the most common residue at that position listed below the logo; a + is used where
1117 there are equal top residues.

1118 **S4 Fig. Alignment highlighting extended regions in the aphid Nemy sequences.** The *A. pisum* Nemy
1119 sequence was aligned with insect sequences exhibiting similarly extended regions in the N-terminus and
1120 the non-cytoplasmic loop connecting helix 1 and helix 2. BLAST-identified insect sequences are from:
1121 *Aphis craccivora* (Ac), *Aphis glycines* (Agl), *Aphis gossypii* (Ago), *Cinara cedri* (Cc), *Diuraphis noxia* (Dn),
1122 *Melanaphis sacchari* (Ms), *Myzus persicae* (Mp), *Rhopalosiphum maidis* (Rm), and *Sipha flava* (Sf). The

1123 longest *D. melanogaster* Nemy isoform is included for comparison. Colored as in S1 Fig., including

1124 aspartate and glutamate in magenta, glycine and alanine in white, and serine and threonine in purple.

1125 H1-H4 are indicated by asterisks. Lines above the alignment show predicted helices based on the

1126 ChimeraX structure of *D. melanogaster* Nemy.

1127 **S1 Table. Insect and outgroup cytb561 sequences used for this study.**

1128 **S2 Table. Sequence identity between insect and D. melanogaster cytb561s.**

1129 **S3 Table. ChimeraX-predicted contacts for Dcytb, CG1275, and Nemy with heme and ascorbate**

1130 **molecules.**

1131 **S4 Table. AlphaFold model confidence for D. melanogaster CG1275, Nemy, and CG8399 structures**

1132 **based on per-residue confidence score (pLDDT) between 0 and 100.**

1133 **S5 Table. Insect sequences with similar extended regions to A. pisum Nemy.**

1134 **S6 Table. Insect sequences with two DOMON domains.**

1135 **S7 Table. Single-domain insect sequences with highest similarity to CG8399.**

1136 **S8 Table. Predicted subcellular localization of insect cytb561 sequences based on DEEPLOC-1.0**

1137 **webserver.**

Fig 1. Phylogenetic analysis of insect cytb561 proteins.

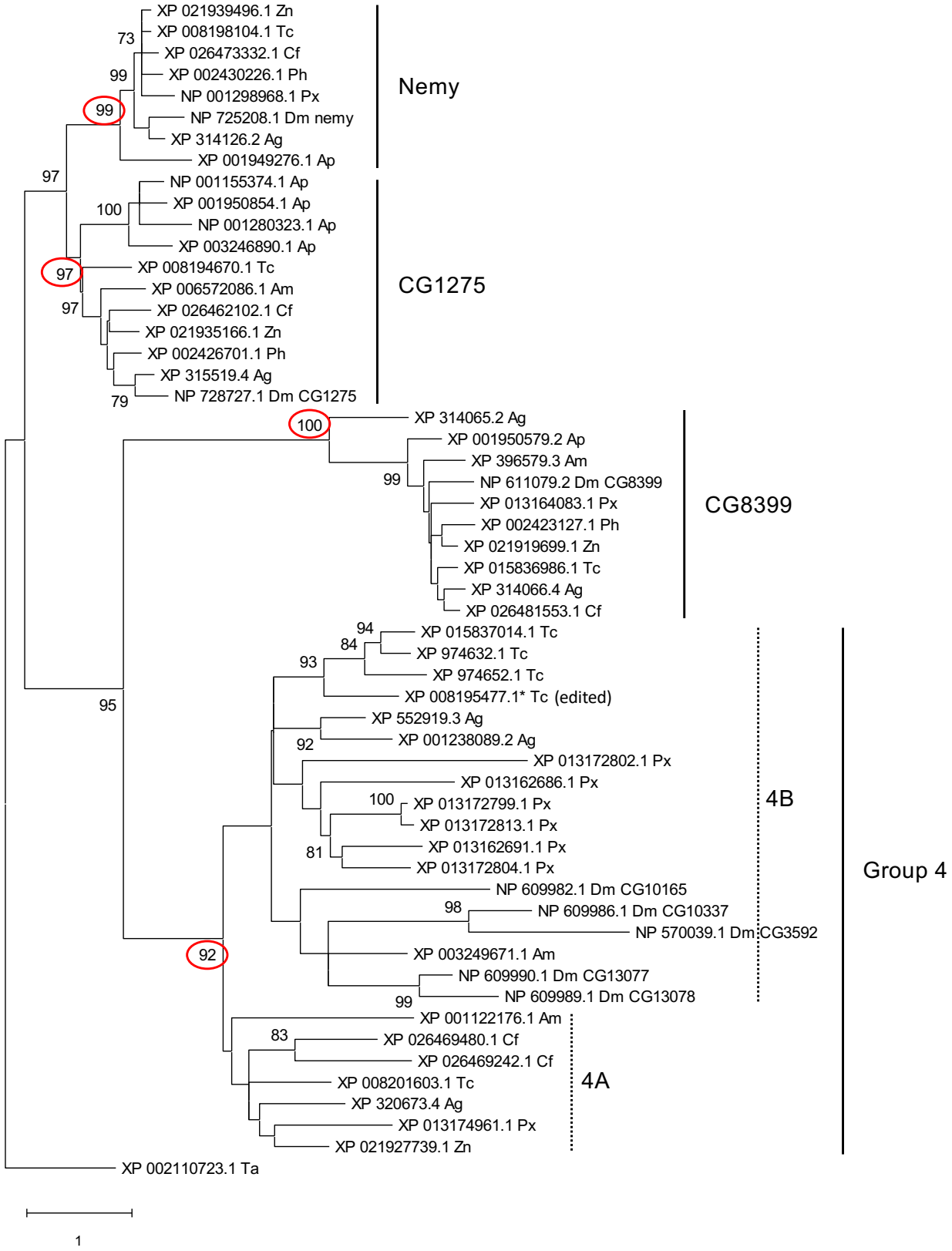


Fig 2. Group consensus logos for the cytb561 core domain.

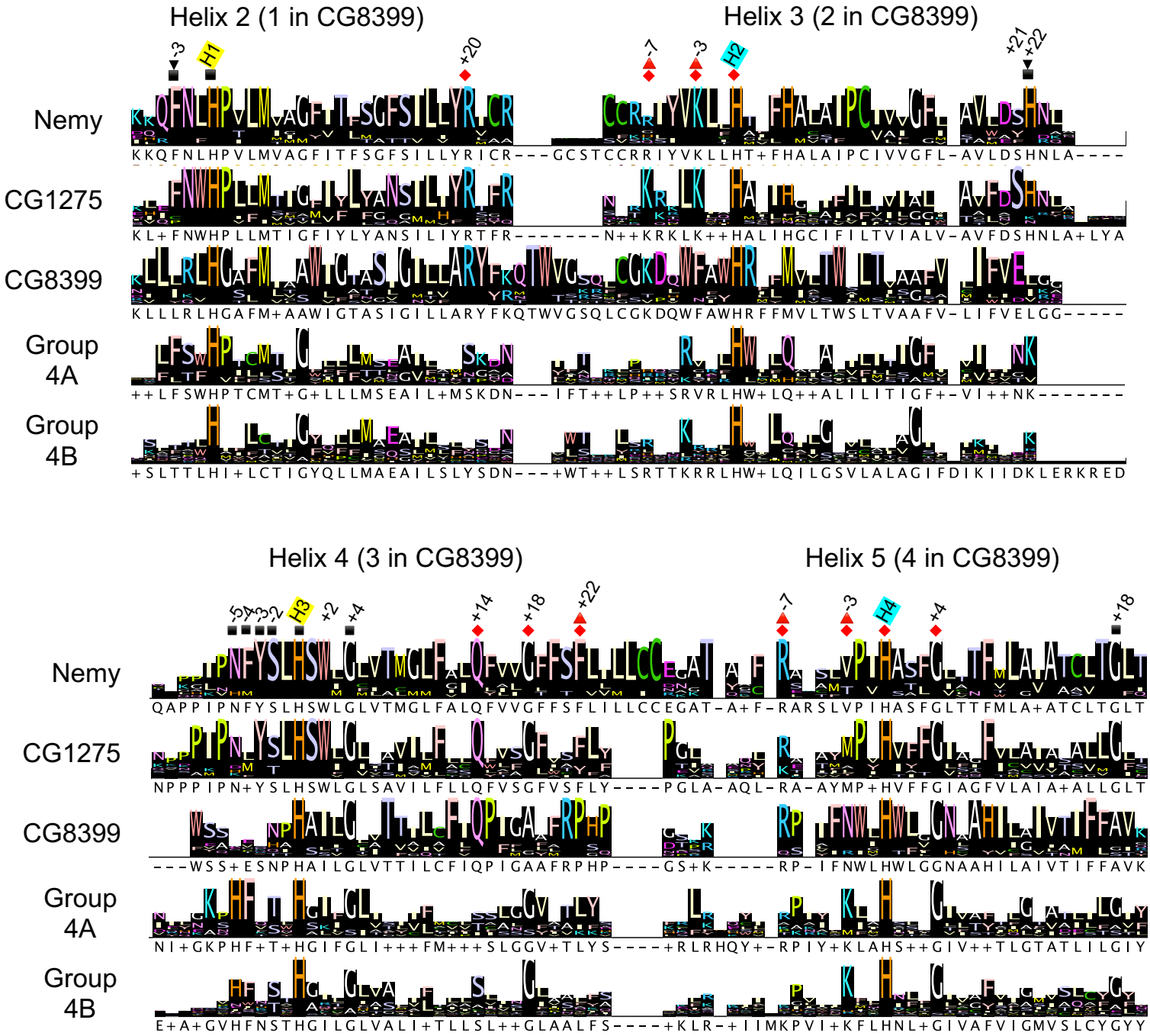
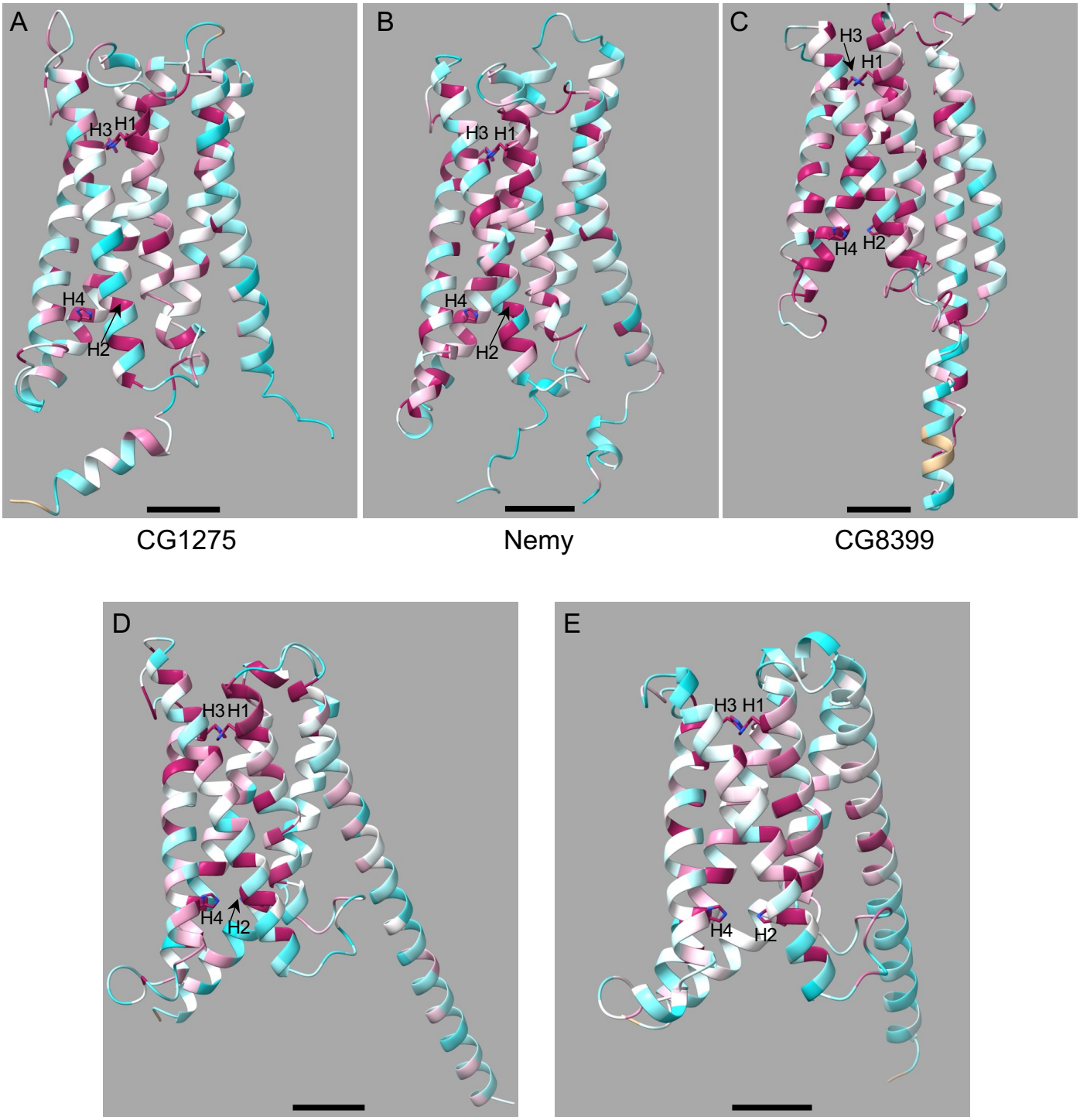


Fig 3. Cytb561 models with coloring based on residue conservation



CG1275 Nemy CG8399

Group 4A (*A. gambiae*) Group 4B (CG13077)

Fig 4. Alignment of Dcytb, Atcytb-B, Nemy, and CG1275 protein sequences.

```
NP_079119.3_Hs          30 WLHYREGLG-----WD-G-SALEFNWHPVLMVTGFVFIQGIAIIVYRLPWTW--KCSKLLM  82
NP_198679.1_At          29 WTVHYRGGLA-----LSSDNKDHIFNVHPVMMVIGLILFNGEAMLAYKSVQ-----GTKNLK  80
XP_001949276.1_Ap      237 WTMAYRGGYDPNWFPWPPSIEDQRFNLHPTLMTIGFVTLTGFSILVYRMAAGCSTSCRGSYV 349
Nemy_A_Dm               79 WMFYRDGFA-----WSES-PKQQFNLHPILMIAGFVTLSGFSILIYRLCR----CVKQIYV 130
XP_314126.2_Ag          49 WTIYYRKGFN-----M-DD-PKLQFNLHPVLMIGGYITLSGFSILLYRICR----CCSHLIV  99
NP_001298968.1_Px       50 WVFFYREGYA-----WADDIPEKQFNLHPTLMVAGFITFSGFSVLLYRICR----CFRRIYV 102
XP_002430226.1_Ph       73 WALYYRGGFA-----WQEN-PKKQFNLHPVLMMMM----MATTILLYRICR----CCRKIYV 120
XP_021939496.1_Zn      103 WVLYYRGGFT-----WTED-PKHQFNLHPVLMVAGFITFSGFSILLYRVCR----CCRRIYV 154
XP_026473332.1_Cf       49 WVLYYRDGFA-----WTED-PKKQFNLHPVLMVAGFITFSGFSILLYRVCR----CCRRIYV 100
XP_008198104.1_Tc       91 WVIKHRGGFA-----WRDD-PIKQFNFHPVLMVAGFITFSGFSILLYRICR----CCRRIYV 142
XP_008194670.1_Tc       46 WLVKYRDGFS-----WS-D-PQLQFNWHPLLMVTGLVFLYGQSILIYRTGR----NATKKRL  96
XP_006572086.1_Am       36 WCGNYRGGFA-----WNSN-PELEFNWHPLLMVIGFVFLYANGMLIYRTQR----NIRKQKL  87
XP_026462102.1_Cf       32 WTGGYLGGFA-----WRST-PELQFNWHPLLMSIGMIFLYSNSMLIYRSYR----SCCKRTL  83
XP_021935166.1_Zn       32 WAGHYRGGFT-----WRSN-PDHEFNWHPVLMTFGMIYLYANSIMIYRAFR----NNQKRRL  83
XP_002426701.1_Ph       30 WLGSYRGGFT-----WTSN-PKLEFNWHPLLMMMGMIFLYANGILVYRSFR----NSRKRKL  81
CG1275_A_Dm            126 WIGQHFGGLA-----GTSN-PGVEFNWHPLFMTIGFIYLYGNSILIYRGFR----TTRKKTL 177
XP_315519.4_Ag          35 WIGIHLNGLG-----WSAR-PSVQFNWHPLLMSVGMIFLYGNSILIYRGFR----YARKKPL  86
XP_001950854.1_Ap       39 WIVQYRGGFS-----IT-E-PKIIFNWHPLLMTIAFIYLLANSILHYRTFQ----NNTKRKL  89
                           *   ^  *                 ** 1*  ^*   ^      ^   ^^ *^

NP_079119.3_Hs          83 KSIHAGLNAVAAILAIISVVAVFENHNVN-------NIANMYSLHSWVGLIAVICYLLQLLS 137
NP_198679.1_At          81 KLVHLTLQLTAFILSLIGVWAALKFHIDK-------GIENFYSLHSWLGLACLFLFAFQWAA 135
XP_001949276.1_Ap      350 KLTHGLLHMGTIPCVLFGGVAAMEYHRLK----G---LPHMYSLHSWMGLLTLMLFAIQLIL 404
Nemy_A_Dm              131 KLIHMFFHAVAIPCIALGFISVFASHDAL----H---KVNFYSLHSWLGFVTMGMFVLQFVI 185
XP_314126.2_Ag         100 KLCHTFFHACSIPCIVIGFMAVWDSHNQQ----Q---IPNFYSLHSWLGMITMGLFALQFVL 154
NP_001298968.1_Px      103 KLLHAIFHALAFPCIVIGFLAVLDYHNKK----G---INNFYSLHSWIGLVAMGLFGLQYAV 157
XP_002430226.1_Ph      121 KLFHALFHALAIPCIVVGFLAVWDFHNLG----Q-PPIPNFYSLHSWLGLVTCGLFILQFVV 177
XP_021939496.1_Zn      155 KLFHTIFHALAIPCIVVGFIAVLDSHNLN----PAGAIPNFYSLHSWMGLVTMGLFALQFVV 212
XP_026473332.1_Cf      101 KLLHTVFHALAIPCVVVGFLAVLDSHNLA----T-PPLPNFYSLHSWLGCVTMGLFGIQFLV 157
XP_008198104.1_Tc      143 KLLHTVFHALAIPCVAVGFLTVLDSHNLA----Q-PPIPNFYSLHSWIGLVTMGLFAIQFVV 199
XP_008194670.1_Tc       97 KVTHALLHLSAFILAVIGLKAAFDSHNYA-----KPPKPNLYTLHSWFGLVAVIIFTGQFVF 153
XP_006572086.1_Am       88 KLIHASLMLLTIALVVISLVAVFDSHNLN-----VKPIPNMYSLHSWIGLTSIILFCCQWLA 144
XP_026462102.1_Cf       84 KIAHSAIHGLIFLLTIFALIAVFDSHNLHMKDGQSDPIPNMYSLHSWIGLTAVILFACQLLS 145
XP_021935166.1_Zn       84 KLAHMLIHFTTFILVIIGLVAVFDSHNLK----NP-PMPNLYTLHSWIGLSAVILFACQWVS 140
XP_002426701.1_Ph       82 KIAHAFINAMAFIFTIIGLQAVFDSHNLA----SPSPIPNMYSLHSWLGLASVILFALQWVF 139
CG1275_A_Dm            178 KLTHAGIHMGAFILTVIALKTVFDSHNLA----NP-PIPNMYSLHSWLGLSAVIVFSLQYVA 234
XP_315519.4_Ag          87 KITHATIHGAAFIFTVVALVAVFDSHNLA----KPNPIPNMYTLHSWVGLSAVILFSLQYVF 144
XP_001950854.1_Ap       90 KNQHALIHGCILILIVIAGFAAFVSH----QYAKP-PIPHLYSLHSWLGVLTIVMFLSQLFS 146
                           *   2   ^            ^    ^    *                ^*^*3**  *    ^  ^^   *

NP_079119.3_Hs         138 GFSVFLL----PWAPLSLRA-FLMPIHVYSGIVIFGTVIATALMGLTEKLIFSLRDPAYS   192
NP_198679.1_At         136 GFVTYWY----PGGSRNSRA-SLMPWHVFLGISIYALALVTATTGILEKVTFLQVNQVIT   190
XP_001949276.1_Ap      405 GLFTFVVLLCCRGATAVCRLRCFTPIHATLGLCTFTLAIATCLTGLQQRADFSIFSNNNG   464
Nemy_A_Dm              186 GFFSFLVMLCCENKTYSCRS-AMVPIHASLGLANFWLAIATSVTGLIEKERETVNE-AGV   243
XP_314126.2_Ag         155 GFFSFLILLCCENATYKFRS-TMVPIHASFGVATFMLAIATAVTGLTQKAHFELGE-NYS   212
NP_001298968.1_Px      158 GFFSFLLLLICNKGTAGFRA-SLVPVHAAFGVLTFVLGVAACLTGLTEKAIFTLGAERYS   216
XP_002430226.1_Ph      178 GTFSFLILLCCEGATAAFRA-SLTPVHSSFGITTFMLAVATCLTGFTQKTLITFGE-TYQ   235
XP_021939496.1_Zn      213 GFFSFLILLCCEGATAAFRA-SLVPIHATFGITTFMLAVATCLTGLTEKAFFELGN-KYS   270
XP_026473332.1_Cf      158 GFFSFLLLLCCDAATQSFRE-AMVPIHASFGLVTFVLAVATCLTGLTEKAIYTLGS-EYS   215
XP_008198104.1_Tc      200 GFFSFLILLCCEGATAGFRA-SLVPIHASFGLTTFMLAIAACLTGLTEKVFFSLSD-DYV   257
XP_008194670.1_Tc      154 GFVSFLF----PGLSRTLRA-AFLPVHVATGIATFTLCVVTALVGITEKTIWTLGS-KYQ   207
XP_006572086.1_Am      145 GFLSFLY----PKLQISLRT-AYMPIHVYFGIAGFIGVITSCLLGLNEKAIFALKE-NYS   198
XP_026462102.1_Cf      146 GFVTYLY----PGLTSNIKT-TYMPVHTFFGIAIFICVVVAALLGLSEKGFFVMKS-NYA   199
XP_021935166.1_Zn      141 GFITFFF----PGLQSPLRA-AYMPVHVFFGLAGFIAAVIAALLGLTEKALWVLQD-EYS   194
XP_002426701.1_Ph      140 GLLSFLY----PTLSAPLRA-TYYPIHVFFGINAFVCASASALLGLTEKAIFNV-S-DYS   192
CG1275_A_Dm            235 GFVAFLA----PGLRENYRI-AMMPLHIYFGLFGFVLAIASALMGITEKAIFAIKTPAYS   289
XP_315519.4_Ag         145 GFAAYLF----PGVREQLRA-SYMPVHVFFGVAGFVMAIAAALLGLLEKAIFSVK--DYS   197
XP_001950854.1_Ap      147 GLWCFLY----PGAAKQHRE-TLAPYHVFFGISIFILAVATAVLGFCEKIIFSLDK-QYK   200
                           *   ^                ^      *  4    *^   ^   ^        *^  ^^
```

Fig 5. *D. melanogaster* CG1275 and Nemy models overlayed on Dcytb heme- and ascorbate-bound structure.

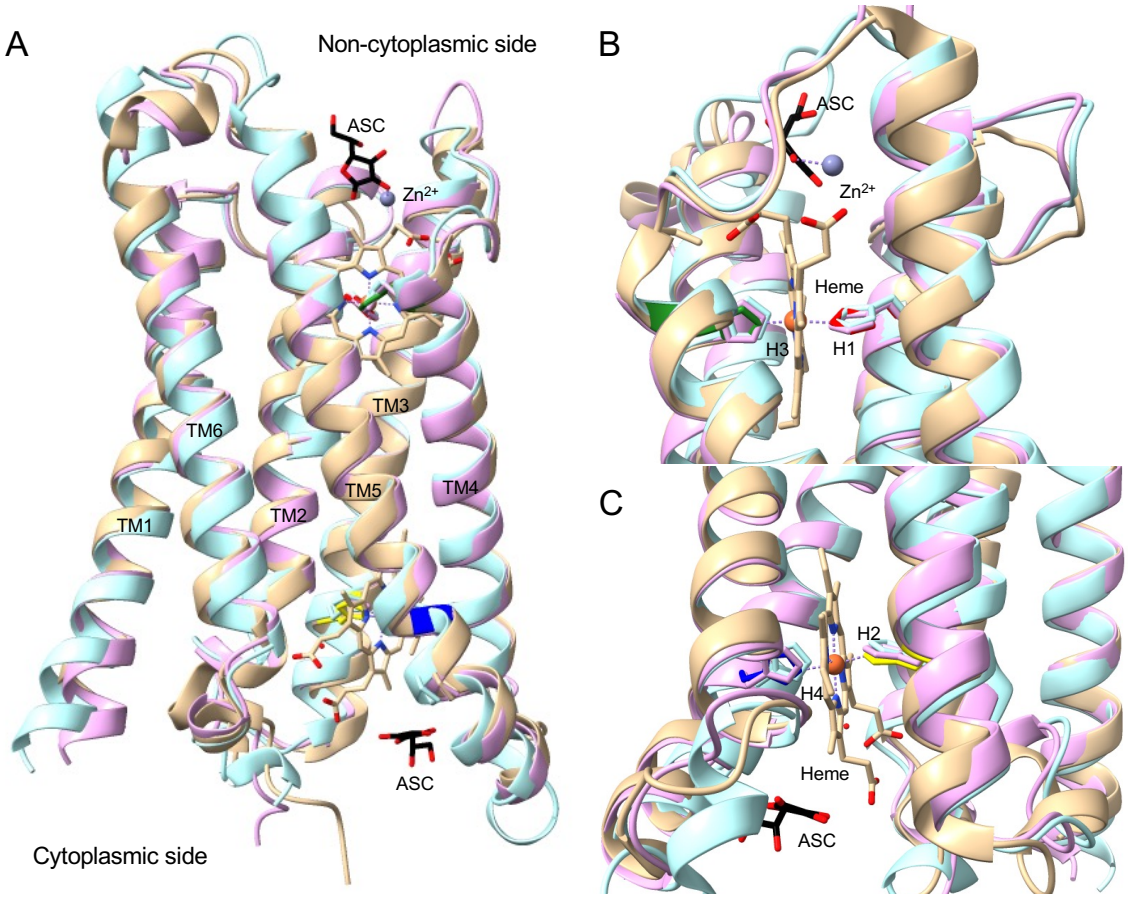
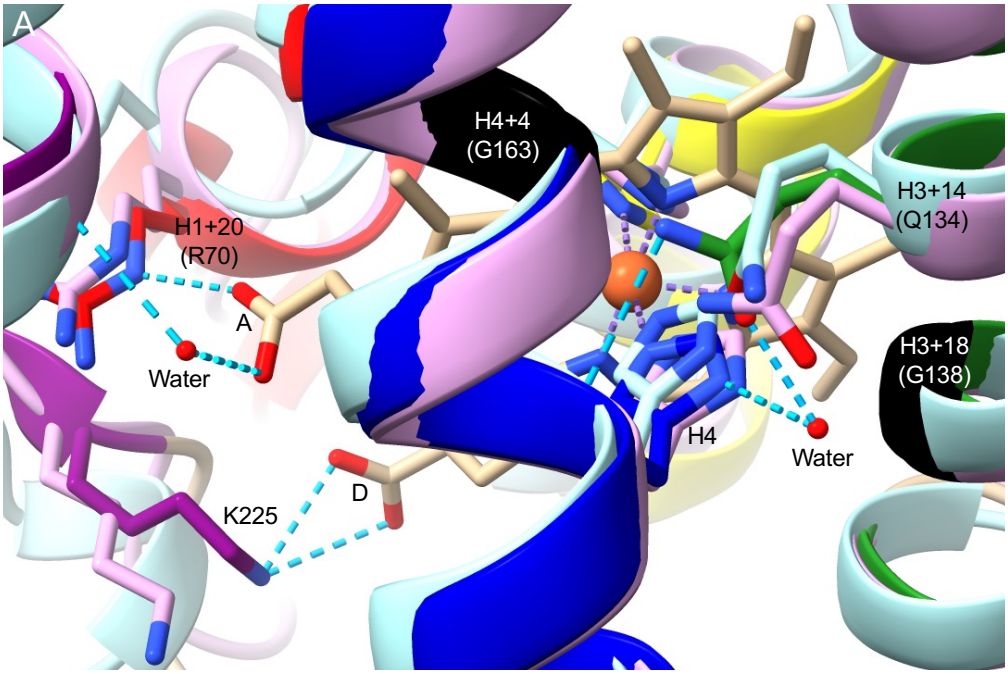
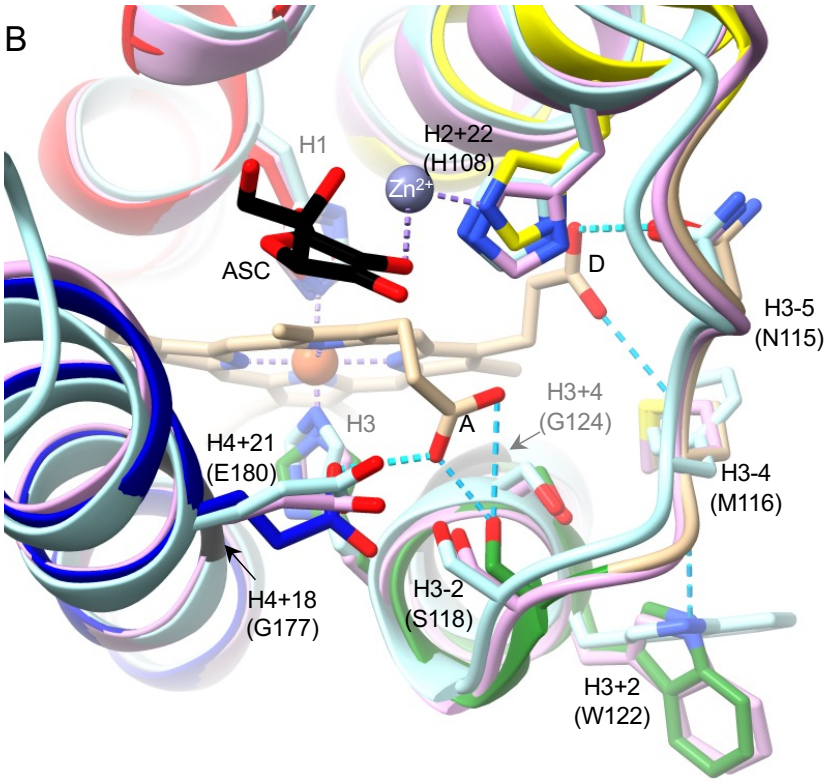


Fig 6. Dcytb and *D. melanogaster* CG1275 and Nemy heme coordination.

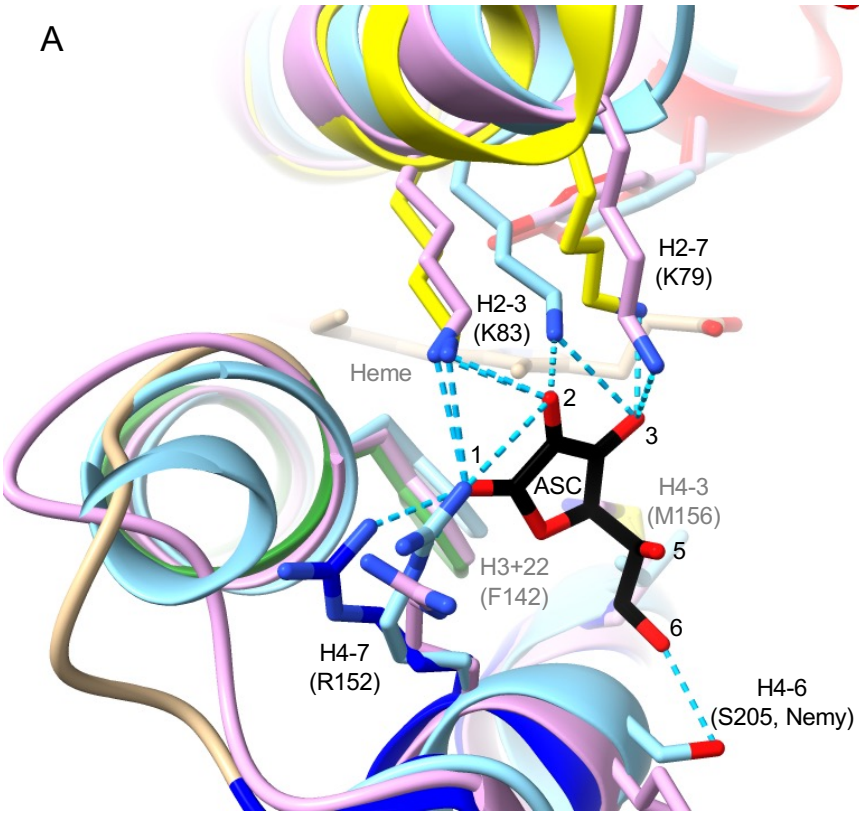


Cytoplasmic Heme

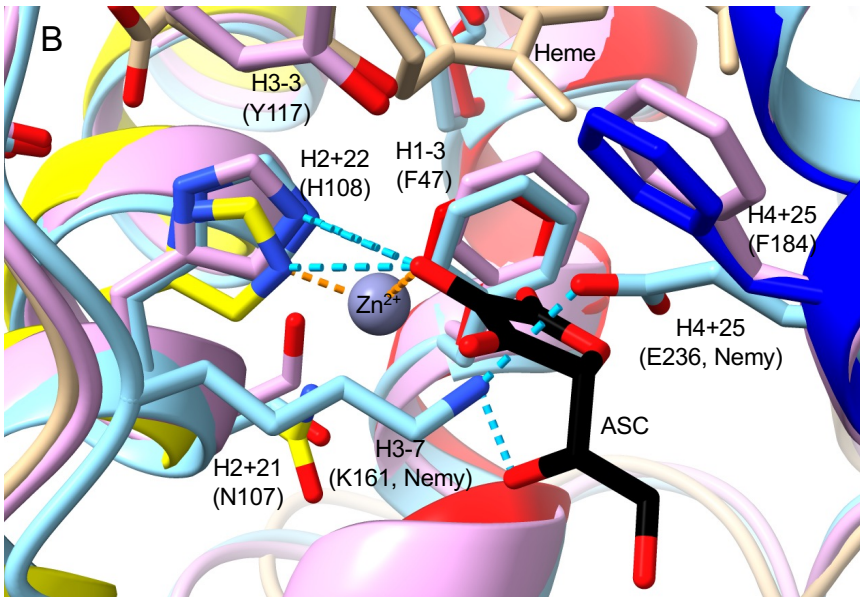


Non-cytoplasmic Heme

Fig 7. Ascorbate binding in Dcytb, CG1275, and Nemy.



Cytoplasmic Ascorbate



Non-cytoplasmic Ascorbate

Fig 8. Cytoplasmic side of *D. melanogaster* CG8399 model with consensus sequence overlay.

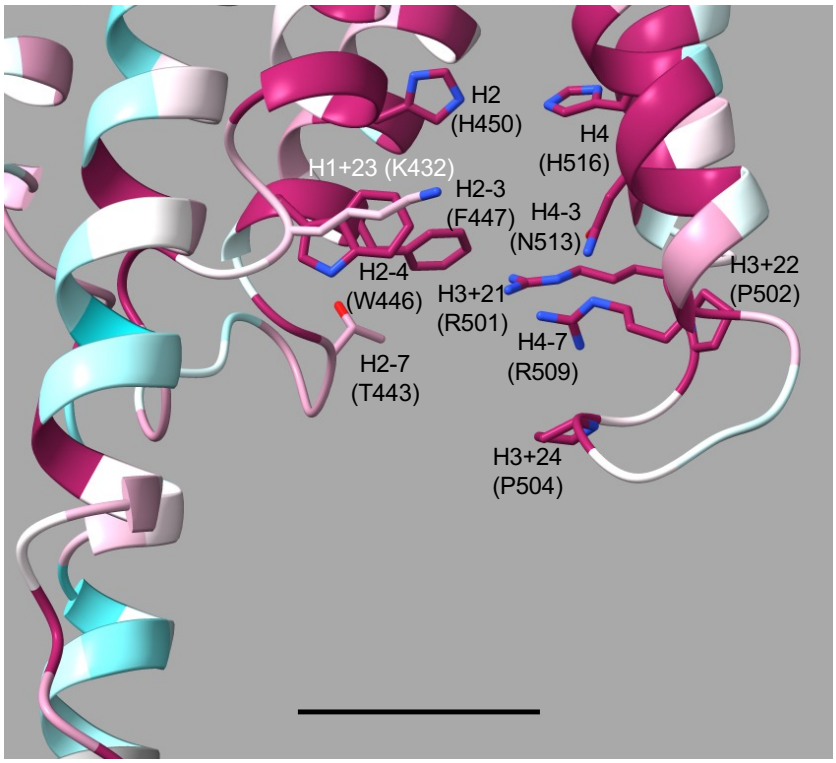


Fig 9. A conserved "KXXXXKXH" motif in Group 4A and related proteins.

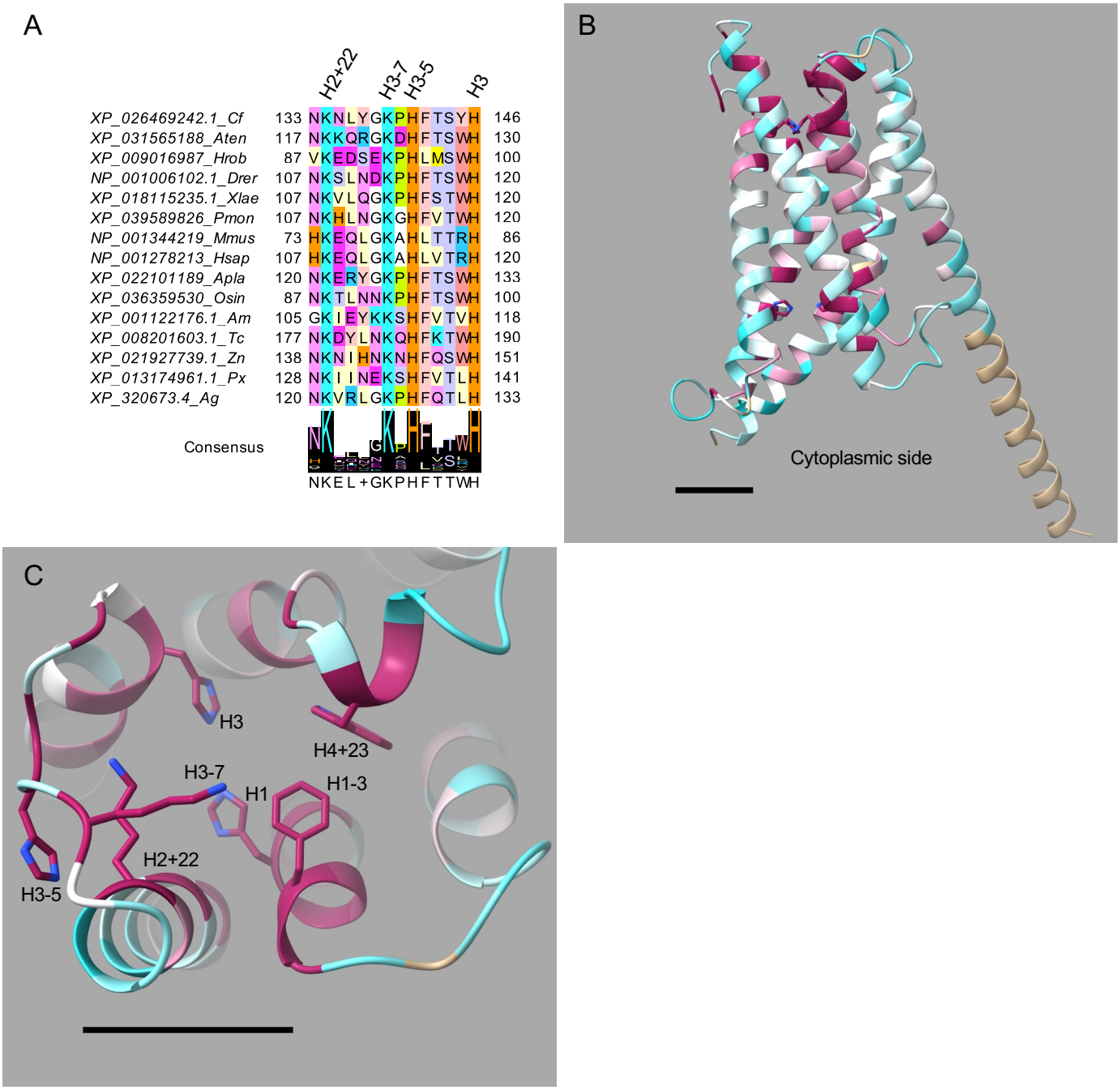


Fig 10. AlphaFold model of *D. melanogaster* CG8399.

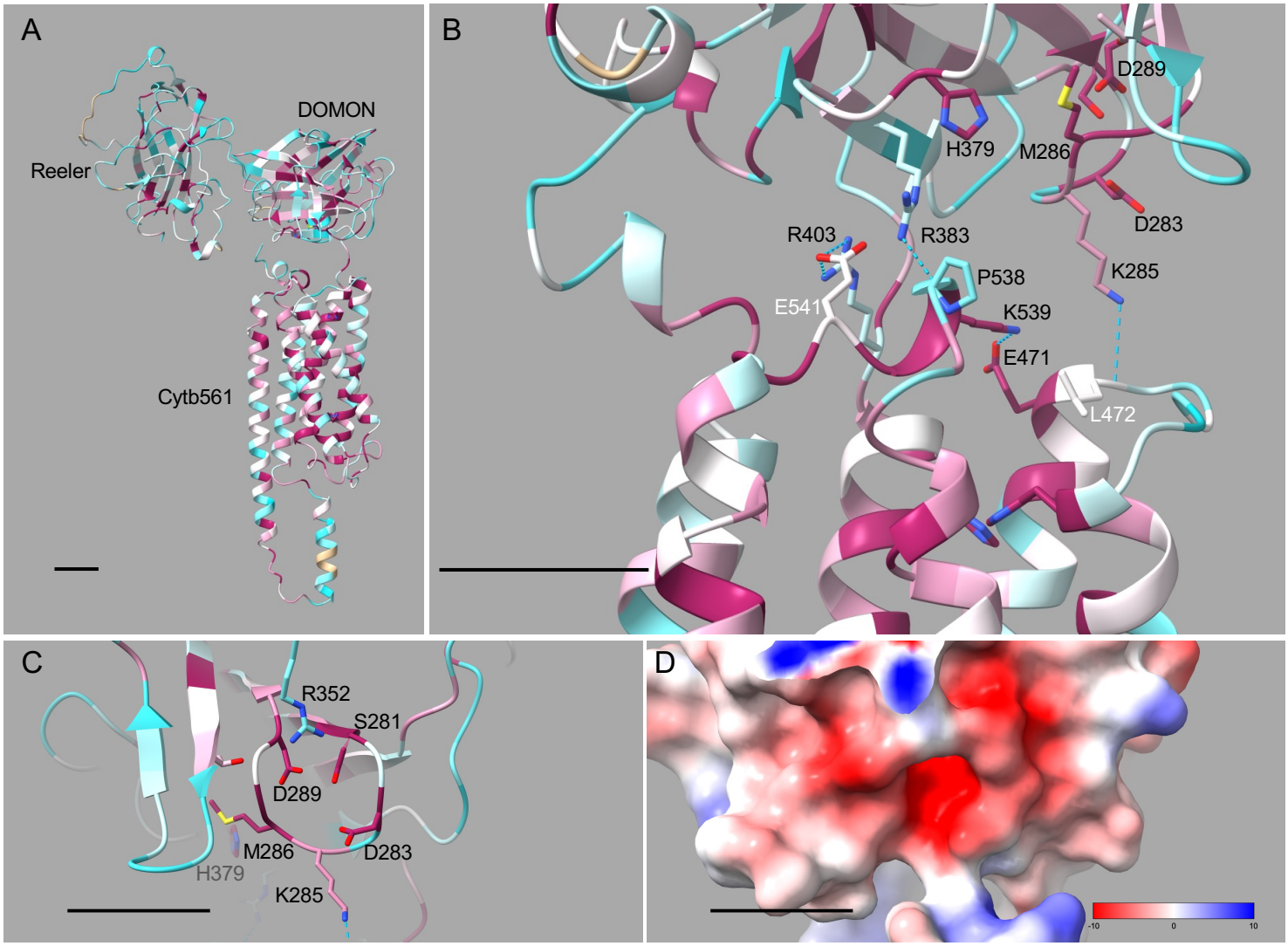
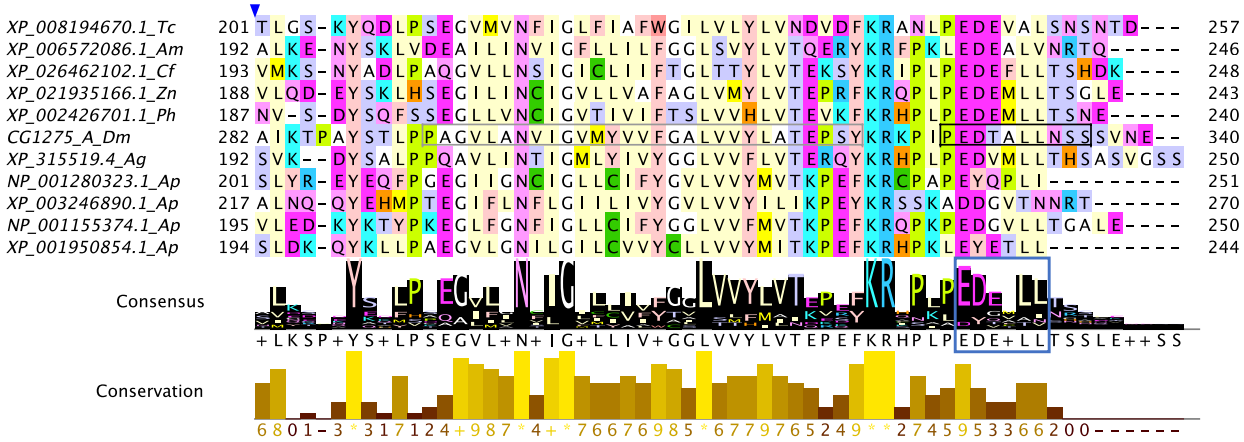



Fig 11. Helix 6 and C-terminus of CG1275 proteins