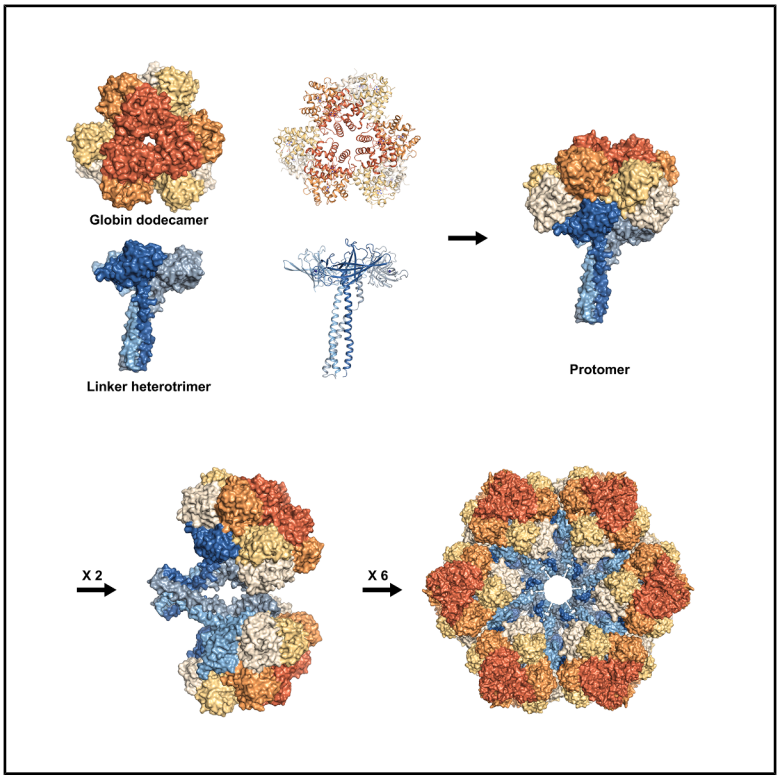
Article

Structure

Structure of *Perinereis linea* erythrocruorin reveals a compact extracellular globin megacomplex

Graphical abstract



Authors

Jia-Xin Deng, Wen-Bin Cheng, Kang Xu, Pu Hou, Yuxing Chen, Yong-Liang Jiang, Cong-Zhao Zhou

Correspondence

jyl@ustc.edu.cn (Y..-L.J.), zcz@ustc.edu.cn (C..-Z.Z.)

In brief

Deng et al. determined the intact structure of the giant extracellular hemoglobin erythrocruorin from *Perinereis linea*, revealing how 180 protein subunits assemble into a compact megacomplex. This study advances our understanding of invertebrate erythrocruorins and provides a structural framework for developing next-generation artificial oxygen carriers.

Highlights

- *Pl*Ec assembles into a compact 3.3-MDa hexagonal bilayer
- Linker proteins shape erythrocruorin architecture
- *Pl*Ec is thermally stable and of interest as a scaffold for oxygen carrier design

Deng et al., 2026, Structure *34*, 1–11
August 6, 2026
https://doi.org/10.1016/j.str.2026.06.004

CellPress

Structure

CellPress

Article

Structure of *Perinereis linea* erythrocruorin reveals a compact extracellular globin megacomplex

Jia-Xin Deng,[1] Wen-Bin Cheng,[1] Kang Xu,[1] Pu Hou,[1] Yuxing Chen,[1] Yong-Liang Jiang,[1,*] and Cong-Zhao Zhou[1,2,*]
[1]Department of Radiology, the First Affiliated Hospital of USTC, and School of Life Sciences, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei 230026, China
[2]Lead contact
*Correspondence: jyl@ustc.edu.cn (Y.-L.J.), zcz@ustc.edu.cn (C.-Z.Z.)
https://doi.org/10.1016/j.str.2026.06.004

SUMMARY

Many invertebrates lack erythrocytes and instead rely on extracellular hemoglobin assemblies, termed erythrocruorins, for oxygen transport. Here we report a 2.84 Å cryo-electron microscopy (cryo-EM) structure of *Perinereis linea* erythrocruorin (*Pl*Ec). *Pl*Ec is a ~3.3 MDa megacomplex composed of 180 polypeptide chains organized into 12 protomers, forming a hexagonal bilayer with D6 symmetry. Each protomer consists of 12 globin subunits and three linker subunits, adopting a mushroom-like architecture. The cap of the mushroom is formed by a globin dodecamer associated with a heterotrimeric linker head, and the stem consists of a triple-stranded coiled coil derived from the N-terminal helices of three linker subunits. Biochemical assays show that *Pl*Ec has thermal stability and auto-oxidation rate comparable to those of other erythrocruorins, but displays relatively lower oxygen-binding affinity. These findings provide mechanistic insights into the quaternary assembly of invertebrate erythrocruorins and lay the groundwork for the potential biomedical applications.

INTRODUCTION

Globins play a fundamental role in oxygen transport and storage in diverse organisms, enabling the efficient respiration and energy metabolism.[1] In human, hemoglobin (hHb) serves as the primary oxygen carrier in red blood cells and exists as a tetrameric protein complex with a molecular mass of approximately 64.5 kDa and a diameter of about 5.5 nm.[2] In contrast, many invertebrates, such as annelids, lack erythrocytes and instead rely on extracellular respiratory proteins known as erythrocruorins (Ecs).[3] These giant globin complexes are freely diffusible in the hemolymph and exhibit remarkable structural stability, forming assemblies with a molecular mass of ~3.6 MDa and diameter of roughly 30 nm.[4] This striking difference in size and organization between hHb and Ecs reflects distinct physiological adaptations and highlights the potential biomedical applications of Ecs.[5]

As the best characterized invertebrate Ecs, the crystal structure of *Lumbricus terrestris* erythrocruorin (*Lt*Ec)[4] revealed that this 3.6-MDa megacomplex is composed of 12 protomers, with each protomer consisting of 12 globin and 3 linker subunits. The globin subunit shares a high structural similarity with hHb and adopts the canonical myoglobin fold of the M-family (PF00042). In contrast, the linker protein belongs to the PF16915 family and contains three domains: an N-terminal extended α-helix, a highly conserved cysteine-rich region (~40 residues) homologous to the ligand-binding repeats of the human low-density lipoprotein receptor (LDLR),[6] and a C-terminal β-barrel domain.[7]

The development of safe and effective hemoglobin-based oxygen carriers (HBOCs) is of great biomedical importance, offering potential alternatives to donor blood transfusion and solutions for oxygen delivery.[8] However, conventional cell-free hHb has a couple of physiological drawbacks due to its small molecular size, which leads to extravasation and nitric oxide (NO) scavenging, causing vasoconstriction and hypertension.[9] Ecs provide a promising natural alternative, which possess an enhanced molecular size and intrinsic stability without the need for artificial polymerization or PEGylation. Their massive quaternary structure at ~30 nm in diameter inherently restricts vascular permeability and exhibits greatly reduced vasoactivity *in vivo*, providing an ideal platform for designing the next-generation HBOCs.[5] In addition, Ecs and other hemoglobins show considerable potential in wound healing applications.[10,11]

Despite their physiological and therapeutic significance, high-resolution structural information on Ecs remains limited, which impedes our understanding of their assembly and the regulation of oxygen binding. Here we report the 2.84 Å cryo-electron microscopy (cryo-EM) structure of Ec from the polychaete *Perinereis linea* (*Perinereis* for short). The structure reveals the molecular architecture and fine assembly pattern of *Pl*Ec megacomplex, which is composed of 180 polypeptide chains forming 12 protomers. Biochemical assays reveal that *Pl*Ec possesses a high thermal stability, making it an ideal candidate for designing oxygen carriers in biomedical applications. Our findings not only expand the structural framework for understanding the giant globins but also provide a foundation for the rational design of oxygen carriers.

Please cite this article in press as: Deng et al., Structure of *Perinereis linea* erythrocruorin reveals a compact extracellular globin megacomplex, Structure (2026), https://doi.org/10.1016/j.str.2026.06.004

RESULTS

Identification, purification, and cryo-EM analysis of *Pl*Ec

Genomic analyses indicate that *Perinereis* relies exclusively on Ec for respiratory function, making it an ideal system for investigating the structural and functional properties of this class of oxygen carriers. Whole-genome sequencing and *de novo* assembly yielded approximately 1.09 Gb of high-quality chromosome-level data (Figure S1). Screening the complete genome for genes encoding the globin (PF00042) and linker (PF16915) domains enabled us to identify 14 putative globin and two linker protein sequences (Table S1).

We isolated the native *Pl*Ec complex from the hemolymph of *Perinereis* by ultracentrifugation, following the procedures described in previous studies.[12–14] The resulting pellet was resuspended and further purified using size-exclusion chromatography (Figure S2). SDS-PAGE analysis revealed at least one protein band around 30 kDa, corresponding to the linker subunits, and two protein bands around 15 kDa, corresponding to globin subunits.

Using cryo-EM, we determined the structure of *Pl*Ec at 2.84 Å resolution (Figure S3). The AlphaFold3-predicted models of all candidate subunits were individually docked into the cryo-EM map. Assignment of each chain was validated by comparing the bulky side-chain densities with the model, which were further confirmed by mass spectrometry (Table S1).

The previous studies of extracellular hemoglobins from invertebrates showed that Ec typically contains four classes of globins (each class may have two subtypes) and four classes of linker proteins.[3,4,15–17] The globin A, usually featuring four cysteine residues, serves as a bridge connecting B and C. Two cysteine residues of globin A form an intramolecular disulfide bond, whereas the two additional cysteines form intermolecular disulfide bonds with subunits B and C, respectively. Based on this feature, we assigned that the *FUN_080850* gene encodes the subunit A (Figures S4A and S4B). Subsequently, *FUN_080846* and *FUN_080849* were assigned to the genes encoding subunits B and C, respectively, which have only one additional cysteine residue at the C-terminus and N-terminus, respectively, that forms an intermolecular disulfide bond with subunit A (Figures S4C and S4D). Meanwhile, the *FUN_050619* gene encoding subunit D was assigned based on the best matching degree of bulky residues in the cryo-EM map. Similarly, we modeled the structures of linker proteins encoded by *FUN_021997* and *FUN_079198* genes into the cryo-EM map. These linker proteins assemble into a heterotrimer in a 2:1 ratio, forming a unit of the scaffold of *Pl*Ec structure. Notably, all these globin and linker protein subunits were detected by mass spectrometry analysis (Table S2).

Overall architecture of *Pl*Ec

*Pl*Ec contains 180 polypeptide chains, which assemble into a massive hexagonal bilayer structure (Figure 1A), with a total molecular mass of approximately 3.3 MDa. The two hexagonal layers are partially staggered, with one layer rotated by approximately 10° relative to the other around the 6-fold symmetry axis, forming an eclipsed configuration. The entire complex exhibits a dihedral D6 symmetry and comprises 12 protomers, each consisting of 12 globin subunits and three linker subunits (Figures 1B and 1C).

Each protomer displays a mushroom-like morphology. The cap of the protomer consists of a dodecameric globin assembly associated with the head of a heterotrimeric linker, the stem of which comprises a triple-stranded coiled coil formed by the N-terminal α-helices. The central core of the megacomplex consists of 36 linker subunits that anchor the 12 globin dodecamers, ensuring the structural integrity of the entire bilayer assembly.

Assembly pattern of the globin components in *Pl*Ec

The 144 oxygen-binding globin subunits of *Pl*Ec are organized into 12 identical protomers, with each assembling into a dodecamer with a local 3-fold symmetry (Figure 2A). Each dodecamer is composed of three copies of four distinct globin chains, designated A, B, C, and D, respectively. All four globins also contain internal disulfide bonds that stabilize the protomer structure. Subunits A, B, and C are covalently linked through inter-subunit disulfide bonds to form a heterotrimer (Figure S4B), whereas subunit D sticks on via non-covalent bonds.

Within each globin dodecamer, three D subunits cluster near the local 3-fold axis, forming a concave apical surface that interacts with the linker subunits, while three disulfide-linked ABC trimers radiate outward at the periphery, forming the cap of the "mushroom," which resembles that of the previously reported structure of *Lumbricus* erythrocruorin (*Lt*Ec) dodecamer.[4,17,18]

Three different interfaces were observed to stabilize the globin dodecamer: two EF-dimer interfaces, one intra-tetramer interface, and one inter-tetramer interface (Figure 2). Each of these occurs in three copies per dodecamer. The two EF-dimer interfaces, formed between A–D subunits (Figure 2C) and between B–C subunits (Figure 2D), are the most extensive interfaces of ~2,000 Å^2 and 2,300 Å^2, respectively. Each globin subunit participates in the EF-dimer interface, a hallmark of invertebrate hemoglobins.[17,19] These interfaces are characterized by extensive contacts between the E and F helices and their associated heme groups. All heme groups directly contribute to the EF-dimer interface through their propionate groups, which form hydrogen bonds with His93(A)-F3 and Gln97(A)-F7 of the adjacent subunit. The topological notation (e.g., F3, F7, F8, B10) refers to the residue position within a specific helix of the globin fold, where "F3" denotes the third residue of the F-helix. These contacts likely respond to the ligation state of the heme, consistent with the observations in *Scapharca* and *Lumbricus* EFs. A conserved cation-π interaction between Arg69(A)-E10 and His88(B)-F3′ further stabilizes the interface and may confer pH-dependent allosteric modulation, as proposed previously.[17] Superposition of *Perinereis* globin dodecamer with that of *Lumbricus* dodecamer yields a root-mean-square deviation (RMSD) of 1.43 Å over 2,012 Cα atoms, indicating a conserved mode of inter-subunit packing.[20]

The intra-tetramer interface, which associates one A–D dimer with one B–C dimer, buries a total surface area of ~2000 Å^2. In the *Pl*Ec structure, subunits A, B, C, and D show an asymmetric contribution to the interface, where subunits A and B contribute more substantially through extensive interactions involving their AB corners (Figure 2E), highlighting the non-tetrahedral arrangement of subunits within the tetramer.

The third interface involves the inter-tetramer interface that mediates interactions between adjacent tetramers. The subunits A and D of one tetramer contact the subunits C and D of a

Please cite this article in press as: Deng et al., Structure of *Perinereis linea* erythrocruorin reveals a compact extracellular globin megacomplex, Structure (2026), https://doi.org/10.1016/j.str.2026.06.004

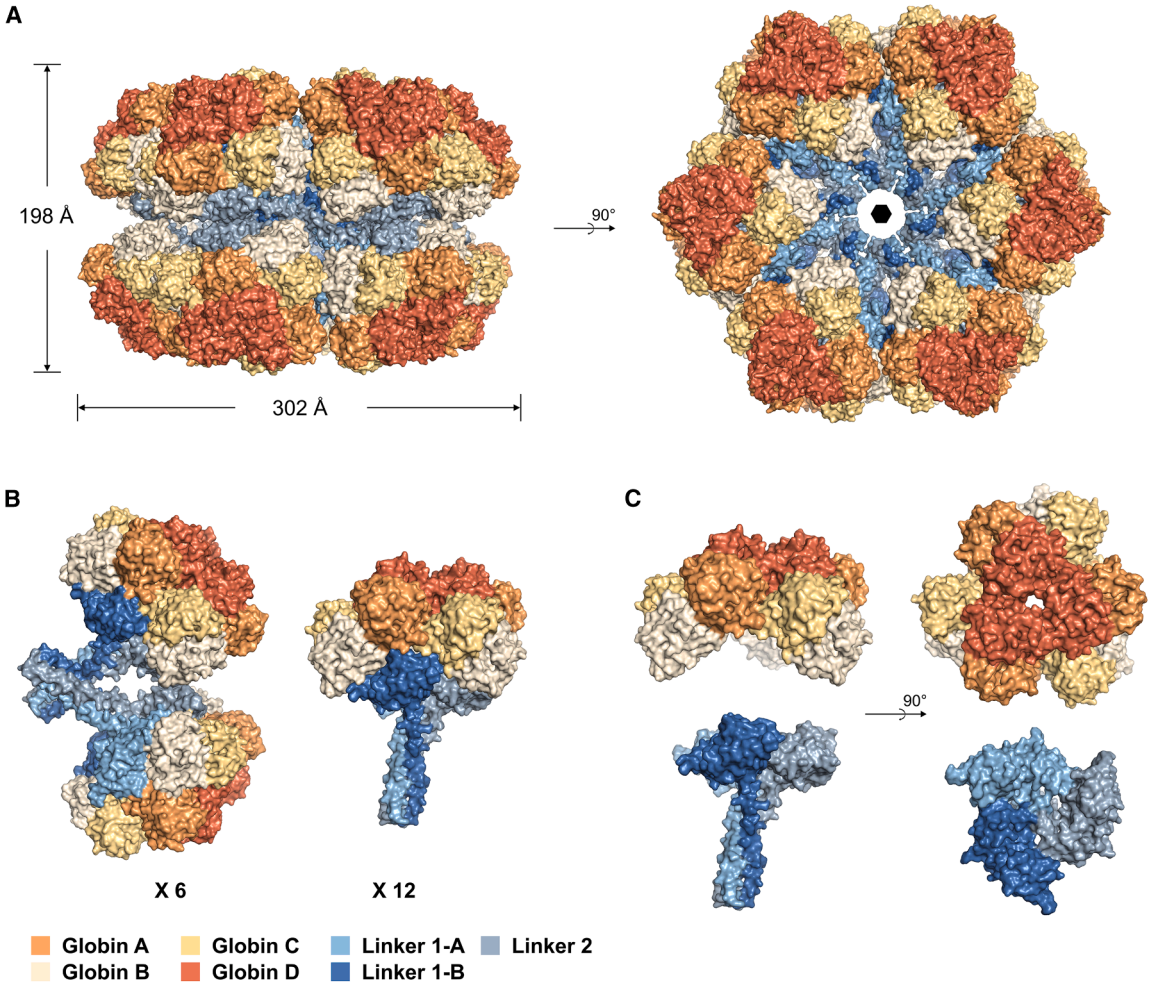


Figure 1. Overall structure of *Pl*Ec megacomplex

(A) Side view (left) along a dyad axis and top view (right) along the 6-fold axis of *Pl*Ec hexagonal bilayer assembly. All *Pl*Ec subunits are shown as a surface with the diameters of the megacomplex labeled.

(B) *Pl*Ec comprises 12 mushroom-like protomers, in which two protomers from different layers are associated by the stem of the mushroom.

(C) A *Pl*Ec protomer includes a hemoglobin dodecamer and a scaffold heterotrimer. The globin A/B/C/D, linker 1A, linker 1B, and linker 2 are colored light orange, pale yellow, yellow orange, orange, blue, light blue and gray, respectively.

neighboring tetramer (Figure 2F), burying an interface area of ~1500 Å^2. This interface is further stabilized by a disulfide bond that links subunits A and C from different tetramers.

The heme-binding pocket

The heme is bound at the heme-binding pocket through extensive noncovalent interactions, similar to those observed in other heme-binding proteins.[21] The heme-binding pocket has two conserved histidine residues, the proximal (F8) and distal (E7) (Figure 3), and the imidazole Nε of F8 is coordinated to the iron directly, whereas E7 plays an important role in the stabilization of heme-bound oxygen via hydrogen bonding and π interaction. The heme is further fixed through a polar contact between the heme-propionate side chains and residues at the heme-binding pocket (e.g., R52 of globin A in Figure 2C).

Residue B10 on the distal side also plays a key role in tuning ligand affinity. All *Pl*Ec globins harbor bulky aromatic residues at this position that modulate ligand access and stabilize the bound state. In the subunits B and D, Trp B10 extends into the pocket to form van der Waals and possible electrostatic contacts with the ligand, whereas in the subunits A or C, Phe B10 performs a similar role. In contrast, subunit C of LtEc contains a leucine at this position instead of an aromatic residue and possesses a rather weak interaction with the ligand. The distal histidine E7 and valine E11 residues in all subunits are positioned to form additional contacts with the bound oxygen or CO ligands, further contributing to ligand stabilization.[4]

Structures of the linker proteins

Each *Pl*Ec protomer contains two distinct linker subunits, designated linker 1 (L1) and linker 2 (L2), which share a similar overall structure (Figure 4). The linker proteins form a heterotrimer that consists of two L1 subunits (L1A and L1B) and one L2 subunit. Each linker protein contains an N-terminal unstructured region

Please cite this article in press as: Deng et al., Structure of *Perinereis linea* erythrocruorin reveals a compact extracellular globin megacomplex, Structure (2026), https://doi.org/10.1016/j.str.2026.06.004

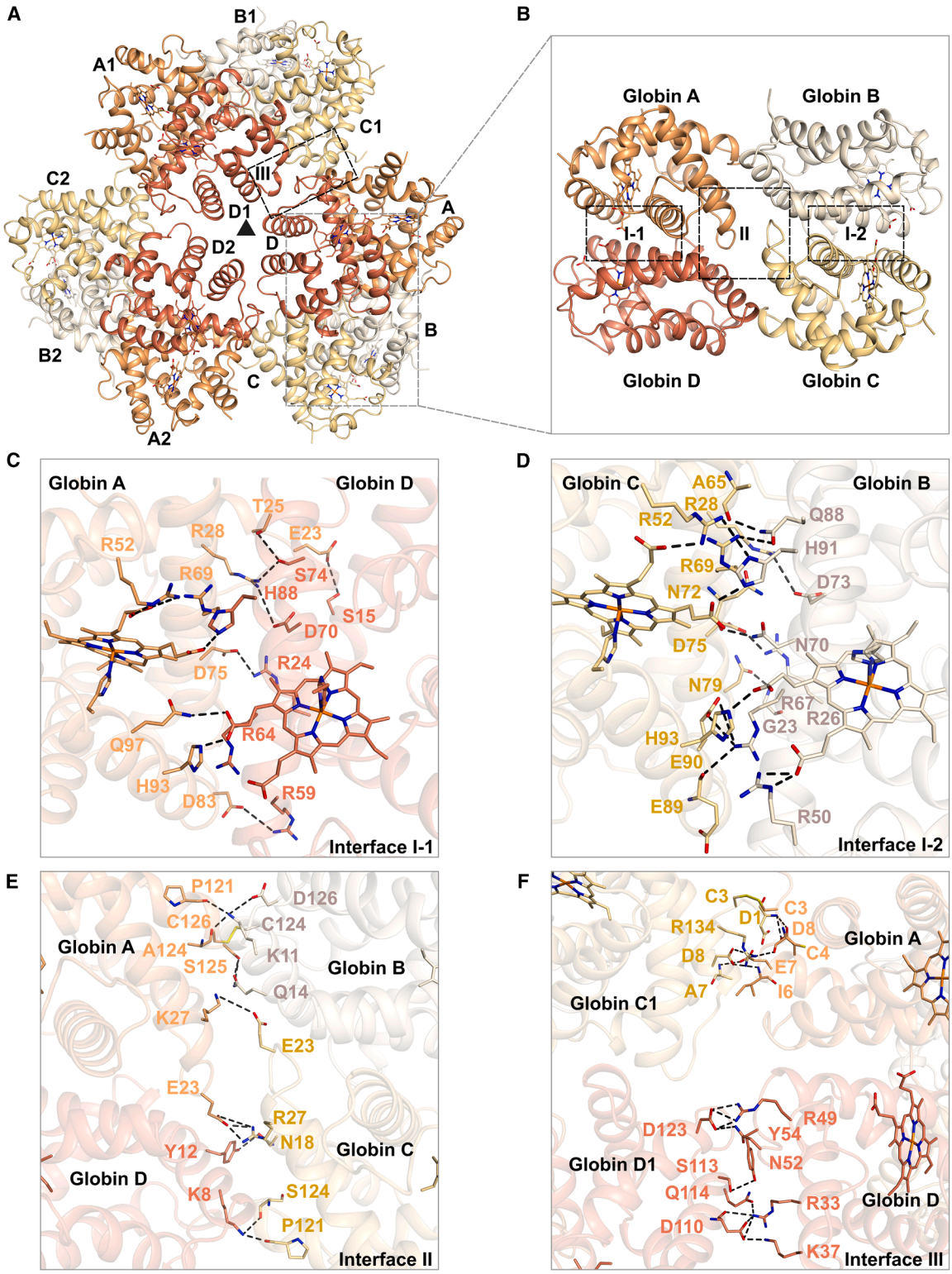


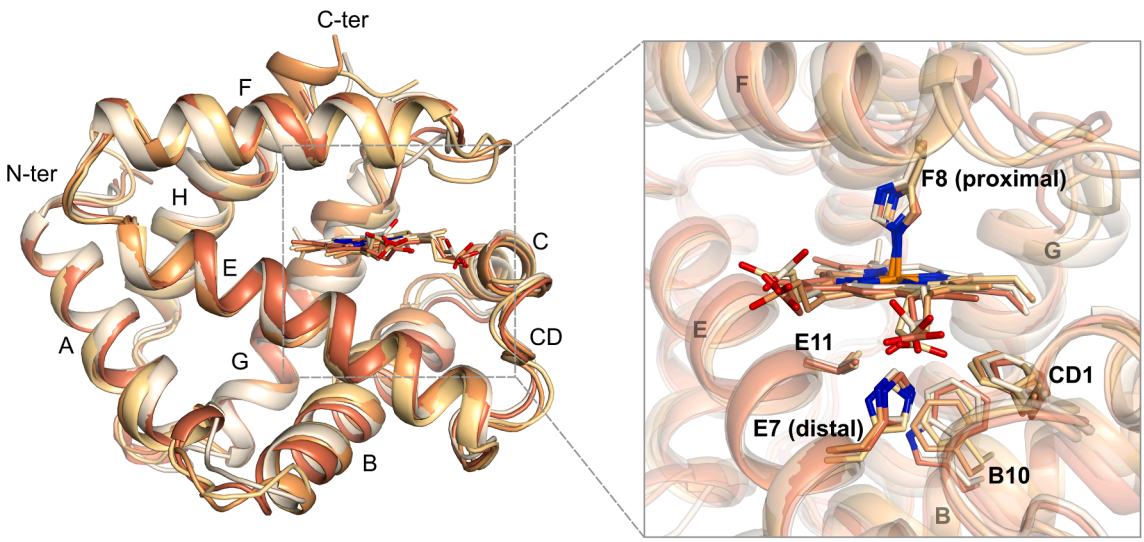
Figure 2. Three distinct interfaces between globins

(A and B) Overview of the *Perinereis* globin dodecamer viewed along the 3-fold axis (A) and the zoomed-in view of one globin tetramer (B). Black rectangles indicate the three different interfaces in globin dodecamer: the interface of dimeric globins (Interface I-1 for subunits A and D, Interface I-2 for subunits B and C), the intra-tetramer interface (Interface II), and the inter-tetramer interface (Interface III).

(C–F) The detailed interactions of three interfaces in the globin dodecamer. Hydrogen bonds and salt bridges are indicated with black dashed lines. The residues involved in hydrogen bonds and salt bridges are labeled and shown as sticks. The O, N, Fe, and S atoms are colored red, blue, orange, and yellow, respectively.

Please cite this article in press as: Deng et al., Structure of *Perinereis linea* erythrocruorin reveals a compact extracellular globin megacomplex, Structure (2026), https://doi.org/10.1016/j.str.2026.06.004

A



B

Globin	Conserved sites				
	F8	E7	E11	B10	CD1
A	H98	H66	V70	F36	F50
B	H96	H64	V68	W36	F48
C	H98	H66	V70	F36	F50
D	H93	H61	V65	W31	F45

Figure 3. Heme binding pockets in four globin subunits

(A) Superposition of the overall structures of four globin subunits. The helices in the globin subunits are named as those in myoglobin. The heme-binding pockets are zoomed in on the right with proximal (F8) and distal (E7) histidines displayed as sticks.

(B) The conserved residues in *Perinereis* globins. In addition to the highly conserved histidines involved in heme binding, the residues at E11, B10, and CD1 in the proximal pocket are also conserved in globin subunits of *Pl*Ec.

and three conserved structural motifs (Figure S5): a long α-helix in L1, which is segmented into two short helices in L2; a cysteine-rich LDLR-like module (~40 residues) homologous to ligand-binding repeats of the human LDLR; and a C-terminal eight-stranded antiparallel β-barrel (~120 residues). Excluding the N-terminal helix, superposition of L1 and L2 yields an RMSD of 1.29 Å over 142 Cα atoms.

As previously reported,[22] the LDLR domain of each linker protein has a calcium-binding site consisting of one glutamate and three aspartate residues. A calcium ion binds to this conserved site of both L1 and L2 of *Pl*Ec. In contrast, *Lumbricus* L2 contains more metal-binding motifs with glycine-rich sequences and polar DDH repeats at the termini,[7] which is absent in *Pl*Ec linkers.

In the assembled protomer, a globin dodecamer associates with the head of the heterotrimeric linker, with their local 3-fold axes aligned (Figure 1C). The external surface of the linker head packs against the globin assembly, forming a hollow space at the interface. The LDLR and β-barrel domains of each linker make extensive contacts with one ABC disulfide-linked trimer in the globin dodecamer (Figure 4), burying an interface area of approximately 3,900 Å^2, which is the most extensive interface in the complex.

Assembly of the *Pl*Ec megacomplex

Previous studies on Ecs have identified two different assembly patterns.[4,23] In *Lt*Ec, the vertices of the two hexagonal layers are partially staggered, with one hexagonal layer rotated by approximately 16° relative to the other layer. In contrast, in *Arenicola marina* Ec, the vertices of the two hexagonal layers are essentially eclipsed. Structural analysis shows that *Pl*Ec

Please cite this article in press as: Deng et al., Structure of *Perinereis linea* erythrocruorin reveals a compact extracellular globin megacomplex, Structure (2026), https://doi.org/10.1016/j.str.2026.06.004

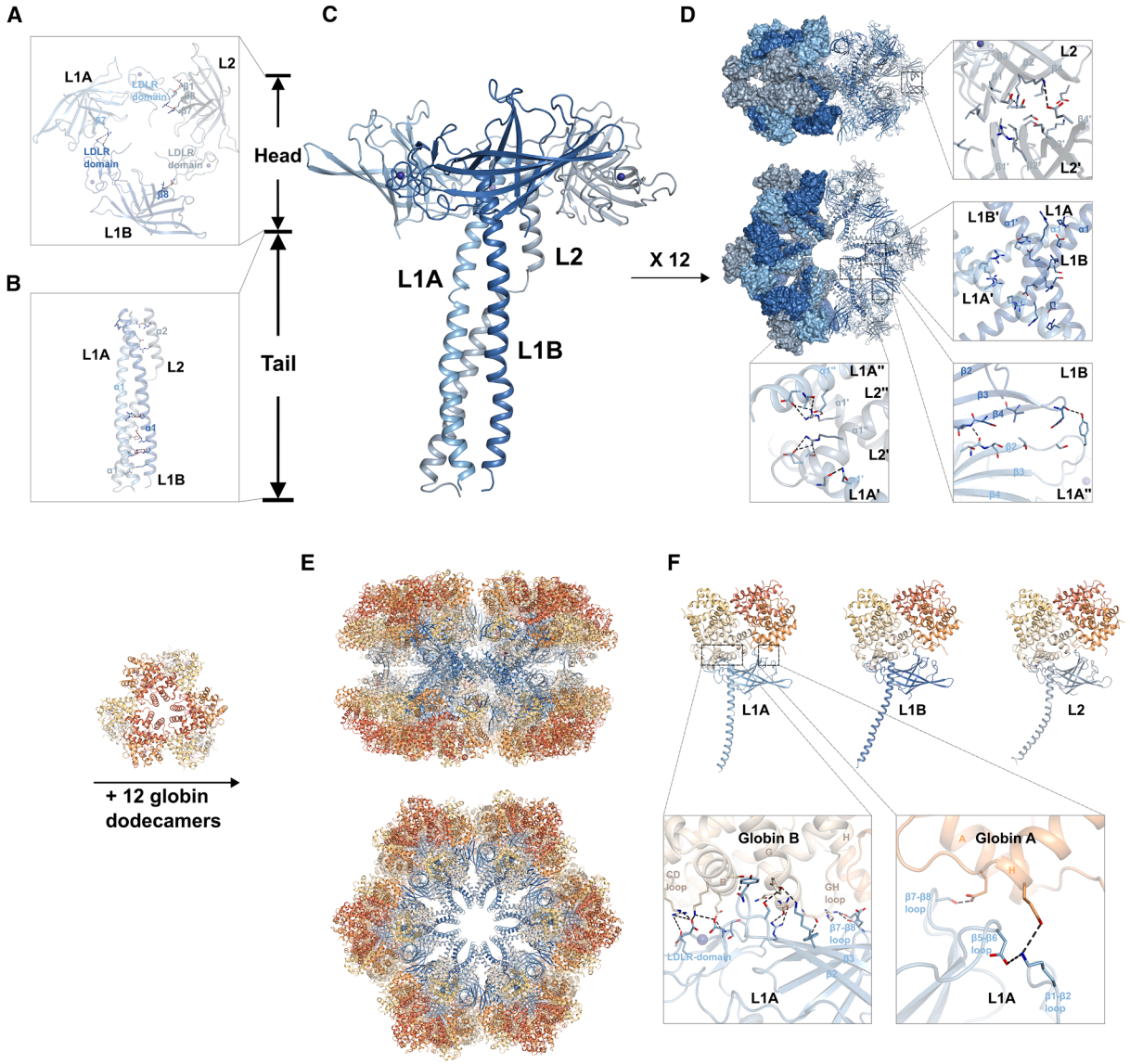


Figure 4. Assembly of linker proteins

(A–C) Through the interfaces of the β-barrel head (A) and helical tail (B), two L1 and one L2 subunit form a heterotrimer (C).

(D) Twelve heterotrimers assemble into the spokes of *Pl*Ec by four interfaces highlighted by rectangles, where residues involved in the interactions are shown as sticks. Residues from the adjacent protomers are labeled with a prime or a double prime.

(E) Upon twelve globin dodecamers binding, the hexagonal bilayer architecture is formed only by interactions between linkers. All subunits in *Pl*Ec are shown as cartoons.

(F) Trimeric linkers interact with globin A and B subunits of dodecamers to form a *Pl*Ec protomer. The helices and loops involved in hydrogen bonds and salt bridges are labeled.

megacomplex adopts an architecture resembling that of *Lt*Ec, exhibiting a rotation of roughly 10° between the two layers. This partially staggered Ec results in a clockwise rotation of the upper layer and a reduced interlayer spacing, yielding a more compact overall structure. Nevertheless, the globin and linker subunits from the opposing layers do not directly interact within Ec assemblies.

In *Arenicola*, the long continuous triple-stranded coiled coils of the linker trimer extend inward from each protomer and interdigitate at the molecular center, forming the connection between the two layers. In contrast, *Lt*Ec contains interrupted coiled coils, producing a more compact bilayer architecture with additional interlayer contacts. However, the *Pl*Ec structure is distinct from either of the two configurations. The two L1 subunits of *Pl*Ec contain continuous N-terminal helices, whereas the L2 subunit possesses a two-segmented helical region (Figure S6). This hybrid arrangement produces an interlayer rotation of approximately 10°, smaller than that observed in *Lumbricus* but greater than in *Arenicola*. These findings suggest that structural variations among the linker proteins might confer the plasticity of Ec assembly.

Please cite this article in press as: Deng et al., Structure of *Perinereis linea* erythrocruorin reveals a compact extracellular globin megacomplex, Structure (2026), https://doi.org/10.1016/j.str.2026.06.004

DISCUSSION

Genomic diversity and evolutionary relationships of *perinereis* globins

A previous study reported that the *Lumbricus* genome contains multiple genes encoding the globin subunits of Ec.[24] In *Perinereis*, we identified 29 genes encoding the M-family globins (PF00042), 14 of which are plausible candidates for Ec components. However, only four of them are modeled in the *Pl*Ec structure. Sequence analyses indicate that *Perinereis* exhibits notable variations of globin genes, resembling those observed in *Lumbricus* (Figures S4A and S7A).

Among these identified genes, two homologous, *FUN_080845* and *FUN_080846*, encode the globin subunit B (Table S1). In addition, four genes (*FUN_046438, FUN_050621, FUN_082744*, and *FUN_137050*) encode identical protein sequences corresponding to globin subunit D (Figure S4A). The genes *FUN_080847* and *FUN_080848* encode globins that contain conserved residues involved in interface I/II formation but possess only two cysteines, preventing the formation of intermolecular disulfide bonds with the neighboring globins. It suggests that they may represent additional variants of the globin subunit D in the megacomplex assembly, potentially giving rise to a minor *Pl*Ec subtype. Unlike *Lumbricus*, which contains multiple genomic copies of each linker subunit, *Perinereis* encodes its linker proteins with only two genes, highlighting the importance of linker composition in *Pl*Ec assembly.

We constructed a mid-rooted maximum-likelihood phylogenetic tree from the 29 globin sequences (Figure S7A), which revealed two primary clusters. One cluster comprises exclusively cellular globins, whereas the other contains both cellular and extracellular globins, with the Ec subunits forming a distinct clade within the latter group. Sequence analysis shows that the extracellular globins exhibit strong conservation, reflecting their shared role in oxygen transport within the hemolymph. In contrast, the cellular globins display notable variations (Figures S7B and S7C), consistent with their functional diversity. For instance, *FUN_030611* and *FUN_129251* may function as neuroglobins rather than classical oxygen carriers.

Comparative and evolutionary insights into the divergence of linker proteins

Previously reported cryo-EM structures of Ecs have identified two different types of Ec assembly.[23] In type I assembly observed in *Lt*Ec,[4] the vertices of the two hexagonal layers are partially staggered, with one layer rotated by about 16° relative to the other layer. In contrast, the type II assembly of *Arenicola marina* Ec exhibits nearly eclipsed vertices between the two layers.[23] Structural analysis shows that *Pl*Ec megacomplex adopts a type I architecture, with an interlayer rotation of approximately 10°, indicating a relatively more compact Ec assembly in polychaete.

Comparative structural analysis of Ecs from *Perinereis*, *Lumbricus*, and *Arenicola* highlights the pivotal role of linker subunits in determining the relative orientation of the two hexagonal layers. Structural and sequence analyses indicate that the calcium binding site within the LDLR domain, as well as the β-barrel stabilized by five disulfide bonds, are highly conserved among linker proteins of annelid Ecs. In contrast, the N-terminal domain, which is unmodeled in the final *Pl*Ec structure, exhibits substantially greater sequence variability, particularly within its helical regions. Sequence alignment of annelid linker proteins reveals a conserved heptad repeat pattern characteristic of coiled-coil motifs (Figure S6B).[25] These residues align along one face of the helix to promote hydrophobic packing and oligomerization. The linkers with uninterrupted heptad repeats favor the formation of continuous coiled coils, and those containing two- or three-residue insertions that disrupt the heptad register, giving rise to fragmented coiled coils.[23] Sequence analyses show that the two linker proteins in *Perinereis* possess continuous heptad repeats; however, L2 adopts a fragmented coiled-coil arrangement. It suggests that the local sequence variations may drive the formation of fragmented coiled-coils, which appear to be necessary for modulating the conformational plasticity of Ec assembly.

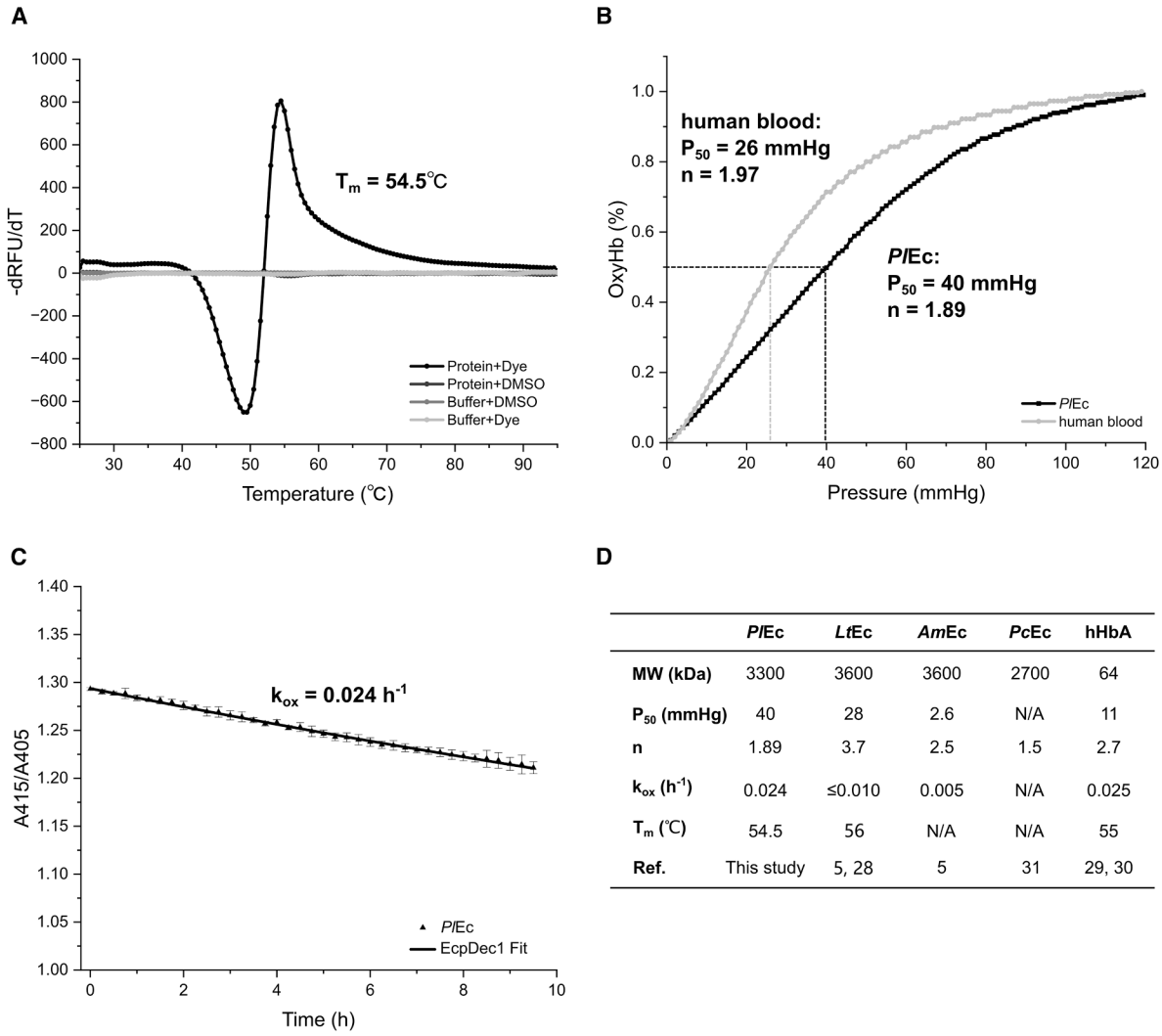
To further investigate the evolution of annelid linker proteins, we conducted a phylogenetic analysis of linker proteins from 184 annelid species (Figure S8). The unrooted maximum-likelihood tree could clearly group the annelid linker proteins into two classes, polychaete and clitellata. The polychaete linkers, including those from *Pl*Ec, contain two additional cysteines in the N-terminal region, which are absent in clitellate linkers. Although the structure of the N-terminal region of the linker has not yet been resolved, the high conservation of these cysteines suggests a potential role in protomer assembly. In the clitellate class, two subgroups can be distinguished, with subgroup 2 characterized by metal-binding DDH repeats at the C-terminus, as previously observed in the *Lumbricus* L2 structure.[7] It implies that the subgroup 2 linker proteins of clitellate possess additional metal-binding sites, such as zinc or calcium, which are proposed to be required for modulating the Ec assembly and functionality.

Previous studies have reported the N-glycosylation of linker proteins on *Glossoscolex paulistus* Ec.[26] Sequence analysis of *Pl*Ec protein components using NetNGlyc 1.0 identified a single N-D-S motif in L2 protein of *Pl*Ec, which is possible for N-linked glycosylation at Asn228. However, it is shown that the presence of a negatively charged residue (Asp) at the X position of N-X-S/T motifs can reduce the efficiency of N-linked glycosylation, likely due to impaired recognition by oligosaccharyltransferase.[27] In fact, our mass spectrometry analysis did not detect any glycosylated peptides, and no obvious glycan density was observed in the cryo-EM map. This observation also agrees with a previous report on *G. paulistus* Ec, in which a potential N-glycosylation site at Asn227 in L2 was identified; however, no corresponding glycan density was detected in the cryo-EM map.[26] Taken together, these results suggest that *Pl*Ec is most likely not glycosylated.

Biomedical application potential of *Pl*Ec

The increasing clinical demand for blood products, combined with the risks of transfusion-transmitted infections and the challenges associated with long-term blood storage, underscores the urgent need for novel blood substitutes. In this context, HBOCs are particularly attractive due to their enhanced stability, high oxygen-carrying capacity, and prolonged circulation time in the bloodstream.[8] *Lt*Ec has been considered as a promising HBOC candidate,[28] as it exhibits several advantages over the adult hHbA, including a lower auto-oxidation rate, a high

Please cite this article in press as: Deng et al., Structure of *Perinereis linea* erythrocruorin reveals a compact extracellular globin megacomplex, Structure (2026), https://doi.org/10.1016/j.str.2026.06.004



	*Pl*Ec	*Lt*Ec	*Am*Ec	*Pc*Ec	hHbA
MW (kDa)	3300	3600	3600	2700	64
P_{50} (mmHg)	40	28	2.6	N/A	11
n	1.89	3.7	2.5	1.5	2.7
k_{ox} (h^{-1})	0.024	≤0.010	0.005	N/A	0.025
T_m (°C)	54.5	56	N/A	N/A	55
Ref.	This study	5, 28	5	31	29, 30

Figure 5. Biochemical characterization of *Pl*Ec

(A) Thermal stability assays of *Pl*Ec and the calculated T_m value of 54.5°C was labeled.

(B) The oxygen equilibrium curve of *Pl*Ec was measured by a Hemox analyzer, and the resulting P_{50} and Hill coefficient (n) were labeled.

(C) Auto-oxidation kinetic curves. The auto-oxidation rate of *Pl*Ec was measured at a protein concentration of 0.3 μM with A415 = 2.6 AU, yielding an auto-oxidation rate constant of 0.024 h^{-1}. Values represent the mean ± SE of three independent measurements.

(D) Comparison of the biochemical parameters among *Pl*Ec, *Lt*Ec, *Am*Ec, *Perinereis cultrifera* erythrocruorin (*Pc*Ec)[31] and hHbA. Experiments on these globins are performed at neutral pH (7.2–7.4), but under various experimental conditions.

molecular weight that restricts extravasation and extends half-life, an elevated thermal stability with the melting temperature T_m of 56°C,[29] and a comparable oxygen affinity with the P_{50} (P_{50} represents the partial pressure of oxygen at which hemoglobin is 50% saturated with oxygen) value of 28 mmHg.[30]

Nevertheless, further improvements in thermal and oxidative stability are essential for developing HBOCs and CO-poisoning antidotes capable of long-term storage, particularly under elevated temperatures. To test the capability of *Pl*Ec as a potential HBOC candidate, we assessed the thermal stability of *Pl*Ec, yielding a T_m value of approximately 54.5°C (Figure 5A). In addition, measurements of oxygen-binding properties show that *Pl*Ec has a moderate oxygen-binding affinity, with a P_{50} of 40 mmHg and a Hill coefficient (n) of 1.89 (Figure 5B). The auto-oxidation rate is approximately 0.024 h^{-1} at a concentration of 0.3 μM *Pl*Ec (Figure 5C), which is comparable to that of hHbA[32] (Figure 5D).

Previous studies on *Perinereis cultrifera* Ec, a polychaete annelid, have reported a lower cooperativity ($n \approx 1.5$) compared to human hemoglobin HbA at neutral pH. It suggests that cooperativity is strongly influenced by environmental factors such as pH, increasing under both acidic and alkaline conditions, with oxygen affinity modulated by ionizable groups.[33] Consistent with this, *Pl*Ec exhibits relatively lower cooperativity at neutral pH. In addition, divalent cations (e.g., Ca^{2+} and Mg^{2+}) are known to modulate both oxygen affinity and cooperativity in Ecs.[33] Structural comparisons have not revealed significant conformational changes between *Lt*Ec and *Pl*Ec. Given their high

structural similarities, the lower apparent cooperativity of *Pl*Ec more likely arises from the different experimental conditions (e.g., absence of regulatory ligands or cations; the oxidative states) rather than structural variations.

Taken together, the large size and covalently stabilized architecture of Ecs confer the physiological and biochemical advantages, including resistance to dissociation into nephrotoxic subunits, prolonged circulatory residence times, and enhanced resilience under variable oxygen and temperature conditions. These properties make Ecs highly promising scaffolds for next-generation HBOCs engineering. Owing to its relatively low oxygen affinity, *Pl*Ec may be less sufficient to carry O_2 over long distances to the hypoxic tissues. On the other hand, this property facilitates more rapid O_2 release from the complex. Therefore, *Pl*Ec may exhibit advantages in applications such as hemoglobin-based sprays. The present *Pl*Ec structure not only advances our understanding of invertebrate oxygen-transport systems but also provides a structural framework for engineering robust, high-capacity oxygen carriers inspired by natural macromolecular assemblies.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Cong-Zhao Zhou (zcz@ustc.edu.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

All study data are included in the article and/or supplemental information. The sequencing data generated in this study have been deposited in the NCBI Sequence Read Archive (SRA) database under Bio Project identifier PRJNA1401452. The coordinates of *Pl*Ec have been deposited to the Protein Data Bank (PDB: 22JW). The EM maps of *Pl*Ec (EMDB: EMD-68387) have been deposited in the Electron Microscopy Data Bank. They are publicly available as of the date of publication. All supporting data in this study are included in the main article and associated files. This paper does not report original code. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

We thank Dr. Yong-Xiang Gao at the Center for Integrative Imaging, University of Science and Technology of China during cryo-EM image acquisition. We thank Associated Professor Shen Li for technical support with the oxygen equilibrium curve measurement at the Institute of Blood Transfusion, Chinese Academy of Medical Sciences, Chengdu. This work was supported by the National Natural Science Foundation of China (grant no. 32241025).

AUTHOR CONTRIBUTIONS

Y.-L.J., C.-Z.Z., and Y.C. conceived, designed, and supervised the project. Y.-L.J., J.-X.D., and C.-Z.Z. analyzed the data and wrote the manuscript. J.-X.D. performed *Pl*Ec purification, cryo-EM sample preparation, and data acquisition. J.-X.D. and W.-B.C. performed the whole genome sequencing and assembly of *Perinereis linea*. J.-X.D., K.X., and P.H. conducted the cryo-EM structure determination and model building. J.-X.D. performed the biochemical assays. All the authors discussed the data and read the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Genome analysis and identification of Ec proteins in *perinereis*
  - *Pl*Ec purification
  - Mass spectrometry analysis
  - Cryo-EM sample preparation and data collection
  - The thermal stability assays
  - The oxygen equilibrium curve
  - Oxidation assays
  - Phylogenetic and conservation analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.str.2026.06.004.

Received: February 2, 2026
Revised: April 28, 2026
Accepted: June 9, 2026

REFERENCES

1. Keppner, A., Maric, D., Correia, M., Koay, T.W., Orlando, I.M.C., Vinogradov, S.N., and Hoogewijs, D. (2020). Lessons from the post-genomic era: Globin diversity beyond oxygen binding and transport. Redox Biol. *37*, 101687. https://doi.org/10.1016/j.redox.2020.101687.
2. Fermi, G., Perutz, M.F., Shaanan, B., and Fourme, R. (1984). The crystal structure of human deoxyhaemoglobin at 1.74 Å resolution. J. Mol. Biol. *175*, 159–174. https://doi.org/10.1016/0022-2836(84)90472-8.
3. Weber, R.E., and Vinogradov, S.N. (2001). Nonvertebrate Hemoglobins: Functions and Molecular Adaptations. Physiol. Rev. *81*, 569–628. https://doi.org/10.1152/physrev.2001.81.2.569.
4. Royer, W.E., Sharma, H., Strand, K., Knapp, J.E., and Bhyravbhatla, B. (2006). *Lumbricus* Erythrocruorin at 3.5 Å Resolution: Architecture of a Megadalton Respiratory Complex. Structure *14*, 1167–1177. https://doi.org/10.1016/j.str.2006.05.011.
5. Elmer, J., and Palmer, A.F. (2012). Biophysical Properties of Lumbricus terrestris Erythrocruorin and Its Potential Use as a Red Blood Cell Substitute. J. Funct. Biomater. *3*, 49–60. https://doi.org/10.3390/jfb3010049.
6. Fass, D., Blacklow, S., Kim, P.S., and Berger, J.M. (1997). Molecular basis of familial hypercholesterolaemia from structure of LDL receptor module. Nature *388*, 691–693. https://doi.org/10.1038/41798.
7. Kao, W., Qin, J., Fushitani, K., Smith, S.S., Gorr, T.A., Riggs, C.K., Knapp, J.E., Chait, B.T., and Riggs, A.F. (2006). Linker chains of the gigantic hemoglobin of the earthworm *Lumbricus terrestris* : Primary structures of linkers L2, L3, and L4 and analysis of the connectivity of the disulfide bonds in linker L1. Proteins *63*, 174–187. https://doi.org/10.1002/prot.20852.
8. Liu, L., Hou, K., Lin, S., Di, Y., Zhuang, Z., Zeng, Z., Sun, Y., Ji, C., Huang, C., and Xiong, R. (2025). Hemoglobin based oxygen carrier and its application in biomedicine. Coord. Chem. Rev. *532*, 216508. https://doi.org/10.1016/j.ccr.2025.216508.
9. Spahn, D.R. (2018). Artificial oxygen carriers: a new future? Crit. Care *22*, 46. https://doi.org/10.1186/s13054-018-1949-5.
10. Xie, P., Jia, S., Tye, R., Xu, W., Zhong, A., Hong, S.J., Galiano, R.D., and Mustoe, T.A. (2016). Topical Administration of Oxygenated Hemoglobin Improved Wound Healing in an Ischemic Rabbit Ear Model. Plast. Reconstr. Surg. *137*, 534–543. https://doi.org/10.1097/01.prs.0000475763.94203.52.

11. Tickle, J., and Bateman, S.D. (2015). Use of a topical haemoglobin spray for oxygenating pressure ulcers: healing outcomes. Br. J. Community Nurs. *20*, S14–S21. https://doi.org/10.12968/bjcn.2015.20.Sup12.S14.

12. Zal, F., Lallier, F.H., Wall, J.S., Vinogradov, S.N., and Toulmond, A. (1996). The multi-hemoglobin system of the hydrothermal vent tube worm Riftia pachyptila. I. Reexamination of the number and masses of its constituents. J. Biol. Chem. *271*, 8869–8874. https://doi.org/10.1074/jbc.271.15.8869.

13. Vinogradov, S.N., and Sharma, P.K. (1994). [8] Preparation and characterization of invertebrate globin complexes. In Methods in Enzymology Hemoglobins Part B: Biochemical and Analytical Methods (Academic Press), pp. 112–124. https://doi.org/10.1016/0076-6879(94)31010-6.

14. Abe, A.S., and Meirelles, N.C. (1985). Oxygen binding properties of erythrocruorin solution and blood PH of the giant earthworm *Glossoscolex paulistus* (oligochaeta, glossoscolecidae). Comp. Biochem. Physiol. Physiol. *80*, 53–55. https://doi.org/10.1016/0300-9629(85)90677-2.

15. Bolognesi, M., Bordo, D., Rizzi, M., Tarricone, C., and Ascenzi, P. (1997). Nonvertebrate hemoglobins: Structural bases for reactivity. Prog. Biophys. Mol. Biol. *68*, 29–68. https://doi.org/10.1016/S0079-6107(97)00017-5.

16. de Haas, F., Zal, F., Lallier, F.H., Toulmond, A., and Lamy, J.N. (1996). Three-dimensional reconstruction of the hexagonal bilayer hemoglobin of the hydrothermal vent tube worm Riftia pachyptila by cryoelectron microscopy. Proteins *26*, 241–256. https://doi.org/10.1002/(SICI)1097-0134(199611)26:3%3C241::AID-PROT1%3E3.0.CO;2-H.

17. Strand, K., Knapp, J.E., Bhyravbhatla, B., and Royer, W.E. (2004). Crystal Structure of the Hemoglobin Dodecamer from *Lumbricus* Erythrocruorin: Allosteric Core of Giant Annelid Respiratory Complexes. J. Mol. Biol. *344*, 119–134. https://doi.org/10.1016/j.jmb.2004.08.094.

18. Royer, W.E., Strand, K., van Heel, M., and Hendrickson, W.A. (2000). Structural hierarchy in erythrocruorin, the giant respiratory assemblage of annelids. Proc. Natl. Acad. Sci. USA *97*, 7107–7111. https://doi.org/10.1073/pnas.97.13.7107.

19. Royer, W.E., Knapp, J.E., Strand, K., and Heaslet, H.A. (2001). Cooperative hemoglobins: conserved fold, diverse quaternary assemblies and allosteric mechanisms. Trends Biochem. Sci. *26*, 297–304. https://doi.org/10.1016/S0968-0004(01)01811-4.

20. Royer, W.E., Jr., Heard, K.S., Harrington, D.J., and Chiancone, E. (1995). The 2.0 Å Crystal Structure ofScapharcaTetrameric Hemoglobin: Cooperative Dimers within an Allosteric Tetramer. J. Mol. Biol. *253*, 168–186. https://doi.org/10.1006/jmbi.1995.0543.

21. Hayashi, T. (2005). Iron: Heme Proteins & Dioxygen Transport & StorageBased in part on the article Iron: Heme Proteins & Dioxygen Transport by Harold M. Goff which appeared in the *Encyclopedia of Inorganic* Chemistry, First Edition. In Encyclopedia of Inorganic and Bioinorganic Chemistry, R.A. Scott, ed. (Wiley). https://doi.org/10.1002/9781119951438.eibc0097.

22. Vinogradov, S.N. (2004). The stoichiometry of the four linker subunits of *Lumbricus terrestris* hemoglobin suggests an asymmetric distribution. Micron *35*, 127–129. https://doi.org/10.1016/j.micron.2003.10.041.

23. Royer, W.E., Omartian, M.N., and Knapp, J.E. (2007). Low Resolution Crystal Structure of *Arenicola* Erythrocruorin: Influence of Coiled Coils on the Architecture of a Megadalton Respiratory Protein. J. Mol. Biol. *365*, 226–236. https://doi.org/10.1016/j.jmb.2006.10.016.

24. Dowd, S., Lagalante, L., Rahlfs, J., Sharo, C., Opulente, D., Lagalante, A., and Elmer, J. (2024). Sequencing of the Lumbricus terrestris genome reveals degeneracy in its erythrocruorin genes. Biochimie *219*, 130–141. https://doi.org/10.1016/j.biochi.2023.11.007.

25. Woolfson, D.N., and Alber, T. (1995). Predicting oligomerization states of coiled coils. Protein Sci. *4*, 1596–1607. https://doi.org/10.1002/pro.5560040818.

26. Ruggiero Bachega, J.F., Vasconcelos Maluf, F., Andi, B., D'Muniz Pereira, H., Falsarella Carazzollea, M., Orville, A.M., Tabak, M., Brandão-Neto, J., Garratt, R.C., and Horjales Reboredo, E. (2015). The structure of the giant haemoglobin from*Glossoscolex paulistus*. Acta Crystallogr. D Biol. Crystallogr. *71*, 1257–1271. https://doi.org/10.1107/S1399004715005453.

27. Shakin-Eshleman, S.H., Spitalnik, S.L., and Kasturi, L. (1996). The Amino Acid at the *X* Position of an Asn-*X*-Ser Sequon Is an Important Determinant of *N*-Linked Core-glycosylation Efficiency (*). J. Biol. Chem. *271*, 6363–6366. https://doi.org/10.1074/jbc.271.11.6363.

28. Dowd, S., Sharo, C., Abdulmalik, O., and Elmer, J. (2024). Optimizing the lyophilization of Lumbricus terrestris erythrocruorin. Artif. Cells, Nanomed. Biotechnol. *52*, 291–299. https://doi.org/10.1080/21691401.2024.2352003.

29. Zimmerman, D., Dilusto, M., Dienes, J., Abdulmalik, O., and Elmer, J.J. (2017). Direct comparison of oligochaete erythrocruorins as potential blood substitutes. Bioeng. Transl. Med. *2*, 212–221. https://doi.org/10.1002/btm2.10067.

30. Hirsch, R.E., Jelicks, L.A., Wittenberg, B.A., Kaul, D.K., Shear, H.L., and Harrington, J.P. (1997). A First Evaluation of the Natural High Molecular Weight Polymeric Lumbricus terrestris Hemoglobin as an Oxygen Carrier. Artificial Cells, Blood Substitutes. Biotechnology *25*, 429–444. https://doi.org/10.3109/10731199709118932.

31. Chiancone, E., Ascoli, F., Giardina, B., Vecchini, P., Antonini, E., Musmeci, M.T., Cinà, R., Zagra, M., D'Amelio, V., and De Leo, G. (1977). Physicochemical and functional properties of Perinereis cultrifera (Grübe) erythrocruorin. Biochim. Biophys. Acta *494*, 1–8. https://doi.org/10.1016/0005-2795(77)90129-5.

32. Talarico, T., Swank, A., and Privalle, C. (1998). Autoxidation of Pyridoxalated Hemoglobin Polyoxyethylene Conjugate. Biochem. Biophys. Res. Commun. *250*, 354–358. https://doi.org/10.1006/bbrc.1998.9312.

33. Ochiai, T., and Weber, R.E. (2002). Effects of Magnesium and Calcium on the Oxygenation Reaction of Erythrocruorin from the Marine Polychaete Arenicola marina and the Terrestrial Oligochaete Lumbricus terrestris. Zool. Sci. *19*, 995–1000. https://doi.org/10.2108/zsj.19.995.

34. Punjani, A., Rubinstein, J.L., Fleet, D.J., and Brubaker, M.A. (2017). cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat. Methods *14*, 290–296. https://doi.org/10.1038/nmeth.4169.

35. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera–a visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605–1612. https://doi.org/10.1002/jcc.20084.

36. Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Cryst D *60*, 2126–2132. https://doi.org/10.1107/S0907444904019158.

37. Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Cryst D *66*, 213–221. https://doi.org/10.1107/S0907444909052925.

38. Palmer, J.M., and Stajich, J. (2020). Funannotate v1.8.1: Eukaryotic genome annotation. Zenodo. https://doi.org/10.5281/zenodo.4054262.

39. Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., et al. (2014). InterProScan 5: genome-scale protein function classification. Bioinformatics *30*, 1236–1240. https://doi.org/10.1093/bioinformatics/btu031.

40. Manni, M., Berkeley, M.R., Seppey, M., Simão, F.A., and Zdobnov, E.M. (2021). BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. Mol. Biol. Evol. *38*, 4647–4654. https://doi.org/10.1093/molbev/msab199.

41. Salvi, T., Khan, M.A., Beyer, G.J., and Palmer, A.F. (2025). Comprehensive biophysical and biochemical characterization of polymerized dopamine coated hemoglobin based oxygen carrier. Int. J. Biol. Macromol. *322*, 146658. https://doi.org/10.1016/j.ijbiomac.2025.146658.

42. Spivack, K., Tucker, M., Zimmerman, D., Nicholas, M., Abdulmalik, O., Comolli, N., and Elmer, J. (2018). Increasing the stability of Lumbricus terrestris erythrocruorin via poly(acrylic acid) conjugation. Artif. Cell Nanomed. Biotechnol. *46*, 1137–1144. https://doi.org/10.1080/21691401.2018.1480491.

43. Katoh, K., and Standley, D.M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol. Biol. Evol. *30*, 772–780. https://doi.org/10.1093/molbev/mst010.

44. Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics *28*, 3150–3152. https://doi.org/10.1093/bioinformatics/bts565.

45. Wong, T.K.F., Ly-Trong, N., Ren, H., Demotte, P., Baños, H., Roger, A.J., Susko, E., Bielow, C., Maio, N.D., Goldman, N., et al. (2026). IQ-TREE 3: phylogenomic inference software using complex evolutionary models. Molecular Biology and Evolution *43*, msag117. https://doi.org/10.1093/molbev/msag117.

46. Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics *25*, 1972–1973. https://doi.org/10.1093/bioinformatics/btp348.

47. Letunic, I., and Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. *49*, W293–W296. https://doi.org/10.1093/nar/gkab301.

Please cite this article in press as: Deng et al., Structure of *Perinereis linea* erythrocruorin reveals a compact extracellular globin megacomplex, Structure (2026), https://doi.org/10.1016/j.str.2026.06.004

STAR★METHODS

KEY RESOURCES TABLE

RREAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
SYPRO™ Orange protein gel stain	Thermo Fisher	Cat. #S5692
Deposited data		
Coordinates of *Perinereis* erythrocruorin	This paper	PDB: 22JW
EM map of *Perinereis* erythrocruorin	This paper	EMD: EMD-68387
Genome of *Perinereis linea*	This paper	BioProject: PRJNA1401452
Software and algorithms		
Origin 2023	OriginLab Corporation, Northampton, MA, USA.	https://www.originlab.com/2023
CryoSPARC	Punjani et al.[34]	https://cryosparc.com/
UCSF Chimera	Pettersen et al.[35]	https://www.cgl.ucsf.edu/chimera
Coot	Emsley et al.[36]	https://www2.mrclmb.cam.ac.uk/personal/pemsley/coot/
PHENIX	Adams et al.[37]	https://phenix-online.org/
PyMOL	PyMOL	http://www.pymol.org/
Funannotate (v1.8.17)	Palmer, and Stajich.[38]	https://github.com/nextgenusfs/funannotate?tab=readme-ov-file
InterProScan (v5.74–105.0)	Jones[39]	ftp://ftp.ebi.ac.uk/pub/software/unix/iprscan/5/
BUSCO (v5.8.3)	Manni[40]	https://gitlab.com/ezlab/busco
Other		
R2/1 200 mesh copper grids	Quantifoil	Q47574
Superose™ 6 Increase 10/300 GL	GE Healthcare	Cat. # 29091596

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

The *Pl*Ec samples were obtained by purifying from the adult worms *Perinereis linea* which were purchased from Taizhou Sandworm Farm of China.

METHOD DETAILS

Genome analysis and identification of Ec proteins in *perinereis*

The genomic DNA was extracted from 3 *Perinereis linea* using magnetic beads, followed with Qubit quantification, and quality control with the Bioanalyzer 2100. Sequencing libraries were constructed and sequenced using the PacBio Template Prep Kit 1.0 (Template Preparation Using BluePippin Size Selection). Furthermore, Hi-C technology was used to assist genome assembly to the chromosomal level. In addition, the Oxford Nanopore Technology (ONT) ultra-long read sequencing libraries were constructed and sequenced according to the manufacturer's protocol (Shanghai Personal Biotechnology Co., Ltd., China). The genome size was estimated according to the K-mers using Illumina sequencing data, which was determined to be approximately 1.09 Gb.

We utilized funannotate (v1.8.17)[38] pipeline to predict protein-coding genes in the genome. Specifically, after masking duplicate sequences, the processed ONT full-length transcriptome sequencing data were used for training, followed by functional annotation of proteins. Motifs and domains of genes were predicted using InterProScan (v5.74–105.0)[39] against protein databases that included ProDom, PRINTS, Pfam, SMART, PANTHER, and PROSITE. The BUSCO (v5.8.3)[40] software was used to assess the completeness of gene prediction.

*Pl*Ec purification

First, rinse the live *Perinereis linea* twice with distilled water to remove visible impurities.[14] Transfer the worms to a small volume of Tris buffer (0.1 M Tris 7.0, 1 mM EDTA) and puncture their dorsal and ventral blood vessels on ice using a syringe needle to obtain crude *Pl*Ec extract.

To remove impurities such as tissue debris, centrifuge the crude extract (12,000 rpm, 4°C, 50 min) and collect the supernatant. Subsequently, use an ultracentrifuge to pellet the large *Pl*Ec protein aggregates (150,000 × g, 4°C, 2 h).[13] Carefully discard the

Please cite this article in press as: Deng et al., Structure of *Perinereis linea* erythrocruorin reveals a compact extracellular globin megacomplex, Structure (2026), https://doi.org/10.1016/j.str.2026.06.004

supernatant and resuspend the protein pellet in a minimal volume of Tris buffer or 0.1 M phosphate buffer (pH 7.4). Approximately 80 mg of *Pl*Ec extract was obtained from 50 g of worms.

After that, the concentrated *Pl*Ec sample was then centrifuged to remove precipitates (12,000 rpm, 4°C, 10 min) and further purified by size-exclusion chromatography using superose 6 Increase column. The elution peak corresponding to the target protein was concentrated for subsequent experiments, which was flash-frozen in liquid nitrogen and stored at −80°C.

Mass spectrometry analysis

*Pl*Ec sample was analyzed on Easy-nLC 1200 liquid chromatography system (Thermo Fisher Scientific) coupled to Q Exactive Plus via a nano-electrospray ion source (Thermo Fisher Scientific, Rockford, IL, USA). Dried peptide samples were dissolved in solvent A (0.1% formic acid in water) and loaded onto a trap column (75 μm × 2 cm, Thermo Fisher Scientific; particle size, 3 μm; pore size, 100 Å) with a maximum pressure of 600 bar using solvent A, then separated on 75 μm × 25 cm Acclaim PepMap RSLC columns (Thermo Fisher Scientific; particle size, 2 μm; pore size, 100 Å) with a gradient of 3–90% mobile phase B (acetonitrile and 0.1% formic acid) at a flow rate of 300 nL/min for 120 min.

The eluted peptides were ionized and using high performance quadrupole precursor selection with high-resolution, accuratemass (HR/AM) Orbitrap detection (Thermo Fisher Scientific). MS analysis was conducted with one full scan (350–1,500 m/z, R = 70,000 at 200 m/z) at an automatic gain control target of 1e6 ions, followed by up to 20 data-dependent MS/MS scans with higher-energy collision dissociation (target 1e5 ions, max injection time 50 ms, isolation window 1.6 m/z, normalized collision energy of 30%). Data were acquired using the Xcalibur software (Thermo Fischer Scientific). The glycosylation peptides were analyzed through the workflow QE_ID_SequestHT_TDA_ecoli_phospho using Thermo Proteome Discover 2.2.0388.

Cryo-EM sample preparation and data collection

The purified protein was concentrated to ∼40 mg/mL. An aliquot of 3.5 μL of the samples was applied to glow-discharged Quantifoil R2/1 200 mesh Cu Holey Carbon Grids. Then the Leica EM GP2 was used to prepare the cryo-EM samples, with a waiting time of 20 s, blotting time of 3.5 s, posttime of 4 s, and humidity level of 95%.

The cryo-EM data were collected under 300 kV FEI Titan Krios microscope equipped with a K3 Summit direct electron camera in counting mode under defocus range of −1 to −2 μm at University of Science and Technology of China. Automated data acquisition was performed with EPU software (Thermo Fisher Scientific), with a nominal magnification of 81,000×, which yielded a final pixel size of 1.0 Å. A total of 6,347 movies (32 frames, each 0.16 s, total dose 60 $e^{-}/Å^{2}$) were collected.

A total of 1,407,115 particles (420 × 420 pixels) were manually picked and extracted from the 6,347 micrographs and background-normalized using cryoSPARC 4.4. After several rounds of 2D and 3D classifications, 1,404,301 good particles were selected for the further 3D refinement, yielding a final resolution of 2.84 Å with D6 symmetry imposed.

The thermal stability assays

The *Pl*Ec protein complex used for thermodynamic parameters measurement were purified in 0.1 M PB (pH 7.4) buffer with 1 mM NADH and GSH as reductive reagents. The thermal stability assays were performed using a modified protocol based on a previous described method for SYPRO orange thermal stability assays.[29] SYPRO Orange Protein Gel Stain (Thermo Fisher, Cat. #S5692) was diluted with DMSO, while the *Pl*Ec were diluted to a final concentration of ∼1 mg/mL. 45 μL *Pl*Ec sample was mixed with 5 μL of the dye in a PCR Microplate (Labselect, Cat. PP-96-NS-0100-W) and analyzed on a CFX96 Touch Real-Time System (BIO-RAD) in triplicate. The fluorescence of each well was monitored while the temperature was increased in 0.5°C every 30 s from 25°C to 95°C. The raw fluorescence data were then analyzed to detect points of inflection which indicate the melting temperature (T_m).

The oxygen equilibrium curve

The oxygen equilibrium curve (OEC) for *Pl*Ec was measured using a Hemox Analyzer (TCS Scientific Corp. New Hope, PA) at 37°C and pH 7.40. 50 μL concentrated *Pl*Ec (20 mg/mL) were diluted in 3 mL Hemox buffer supplemented with 20 μL additive A and 10 μL antifoam agent (TCS Scientific Corp.). The resulting data were fit to the Hill equation to regress the P_{50} and the Hill coefficient (n) to characterize the oxygen binding ability of *Pl*Ec.[41]

Oxidation assays

The *Pl*Ec protein complex used for oxidation assays were prepared after desalting to remove the reductive reagents. Oxidation level measurements were performed in clear 96 well plates with a SpectraMax iD5 Multi-Mode Microplate Reader (Molecular Devices), as described previously.[42] To measure oxidation rates, *Pl*Ec was diluted in purification buffer to 0.3 μM with a volume of 200 μL in a 96-well plate. Each solution was stored in the dark at 37°C and absorbance spectra were measured every 15 min for 10 h. The absorbance measurements at 415 nm (Soret band maxima) and 405 nm (an isosbestic point shared by *Pl*Ec and met*Pl*Ec) with the average of three independent measurements were then used to estimate the oxidation level of the samples at each time point using equations in previous report.[42]

Phylogenetic and conservation analysis

The linker protein sequences (PF16915 family) used for phylogenetic analyses were obtained from InterPro database, which contains 429 linker subunits of the erythrocruorin respiratory complex from 164 annelid worms. The linker sequences were aligned with MAFFT

Please cite this article in press as: Deng et al., Structure of *Perinereis linea* erythrocruorin reveals a compact extracellular globin megacomplex, Structure (2026), https://doi.org/10.1016/j.str.2026.06.004

v7.526 using the G-INS-i algorithm (-maxiterate 1000).[43] Nonredundant linker sequences were clustered at 100% identity using CD-HIT v4.8.1 (-c 1).[44] A maximum likelihood tree was constructed using IQ TREE v3.0.1[45] after alignment with MAFFT and trimming with TrimAL v1.4.rev15.[46] The output tree was visualized using iTOL.[47]

QUANTIFICATION AND STATISTICAL ANALYSIS

Cryo-EM data collection and refinements statistics are summarized in Table S3. Softwares required for quantification and statistical analyses include ChimeraX, PHENIX, CCP4 and COOT.